MODEL BASED CROSS OVER DESIGN FOR POLYTOMOUS RESPONSE

July 14, 2022

Suvendu Samanta

Reg No - 2041411001

Roll No - 01

Dhitashree Paria

Reg No - 20414210002

Roll No - 02

PRESIDENCY UNIVERSITY

KOLKATA

Department of Statistics

Acknowledgement

We would like to express our special thanks of gratitude to our advisor **Dr. Saurav De** for his able guidance and support for the completion of the project on **Model Based Cross-over Design under ordinal catagorical response**. We would like to thank him for his available suggstions and reference books which helped us a lot in finalizing this project within the given time frame. Besides our advisor, we would also like to thank our remaining faculty professors and Head of the Department, Department of Statistics, Presidency University. And also a very special thanks to **Prof. Joideep Basu** for giving his precious time and advice in this regard. We also sincerely thank our parents, fellow classmates and friends who have supported and encouraged me while doing this project.

Contents

Introduction

A **Crossover Design** is a repeated measurements design such that each experimental unit (patient) receives different treatments during the different time periods, i.e., the patients cross over from one treatment to another during the course of the trial. In medicine, a crossover study or crossover trial is a longitudinal study in which subjects receive a sequence of dierent treatments (or exposures). While crossover studies can be observational studies, many important crossover studies are controlled experiments. Crossover designs are common for experiments in many scientic disciplines, for example psychology, pharmaceutical science, and medicine. Randomized, controlled crossover experiments are especially important in health care. In a randomized clinical trial, the subjects are randomly assigned to different arms of the study which receive different treatments. When the trial has a repeated measures design, the same measures are collected multiple times for each subject. A crossover trial has a repeated measures design in which each patient is assigned to a sequence of two or more treatments, of which one may be a standard treatment . Nearly all crossover are designed to have "balance", whereby all subjects receive the same number of treatments and participate for the same number of periods. In most crossover trials each subject receives all treatments, in a random order. Crossover trials are used in repeated measures study where subjects receive the same or dierent treatment during the period of study. This work illustrates the use of a bi-variate polytomous response probability model in the simple two treatment, period, two sequence AB/BA crossover design and presents two assymptotic for comparing treatment eects. Earlier study contains tests for equality of treatment eects depending on various structures(which includes absence or presence of period effect) of response model. The impact of crossover eect can be diluted by inserting a lengthy washout time between two treatment periods.

Regression Models for polytomous responses are available. For example, McCullagh(1980) and Anderson (1984). Majority of works has been done in Clinical Trials for dichotomous or binary response. For example, Jones and Kenward (1989). Previously, the cross-over design has been applied in many elds, but mainly for dichotomous response. Model based polytomous response in cross-over study is really very rare in the past literature. Cros-over design is a major part in Clinical Trials. We apply cross-over design mainly in case of Chronic or recurring disease. Here,in this cases, treatments are interchanged or rotated. Eects of Treament A may depend on eects of Treatment B or viceversa. So, it draws our attention to perform cross-over study in case of model based Polytomous Response.

Here,we consider a group of patients and devide this group of patients into a number of ordinal stages. For example, 0 being the best health stage of the respective patient and G being the worst stage. In this context, G being the worst stage, G is taken as 2 and 3. When $G = 2$,then 0 is the most favourable stage and 2 being the worst stage of the respective patient, that is, dichotomous response. When $G = 3$, then 0 is the best stage, 1 and 2 being the intermediate stage and 3 is the worst stage of the respective patient. The objectives in the cross-over design are as follows:

- To create a rational model for the two-period three treatment Crossover Design under polytomous response.
- The performance of the model can be judged in the light of the few testing procedures. The two treatments involved in the cross-over design are A and B, say. Through cross-over design, we will introduce the test procedures and assess their performance through the power of the test procedures.
- Also, we will focus on the carry-over effect ρ and discuss about its inferential procedures, which is an important factor in cross-over design.

Model Based on Cross-over Treatment under Ordinal categorical response

Suppose response of a **Cronic** disease can be categorized into $(G + 1)$ categories, $G \geq 1$, lowest category represents most favourable condition and highest category indicating very severe condition of the disease. Let, there only 2 options are available A and B for treating the disease, our treatment appliedto a patient for a particular period. Suppose, we restrict the trials to only two periods. A patient who receives k treatment in the 1st period , will receive treatment $k'(\neq k)$ in the 2nd period, $k = A, B$;$k' = A, B$. Therefore in our scenario a patient on trial will receive either AB or BA sequence of treatments in two periods.

Suppose in the first period, N_1patients are allocated randomly to treatment A and N_2patients are allocated to treatment B; where $N_1 + N_2 = N$. Those N_1patients who are treated with treatment A in the 1st period will be treated with treatment B in the 2nd period. Simillarly the other N_2patients will be given treatment A in the 2nd period.

Such design is called **Two-treatment Two-period Cross-over** design.

Let, $X_{ks} = 1$; if in 1st period a patient is treated with k treatment , and shows response in sth category;
$= 0$ otherwise ; where, $k = A$ or B ; $s = 0, 1, ..., G$.

Where, 0 and G are representing the score of lowest and highest category respectively.

$$\Rightarrow \sum_{s=0}^{G} X_{ks} = 1$$

$$\Rightarrow P\left[X_{k0} = x_{k0}, X_{k1} = x_{k1}, ..., X_{kG} = x_{kG}\right] = \pi_{k0}^{x_{k0}}.\pi_{k1}^{x_{k1}}......\pi_{kG}^{x_{kG}}$$

; where π_{ks} is the probability that the randomly choosen patient from 1st period, treated by treatment k will come from sth category such that $0 < \pi_{ks} < 1$ and $\sum_{s=0}^{G} \pi_{ks} = 1$ such that $x_{ks} = 0, 1$ where, $\sum_s x_{ks} = 1$.

i.e,

$$(X_{k0}, X_{k1}, ..., X_{kG}) \sim MN(1, \pi_{k0}, \pi_{k1}, ..., \pi_{kG}), where \sum_{s=0}^{G} \pi_{ks} = 1$$

Now for **Second period** ;

$X_{kk's} = 1(0)$; if a patient receiving treatment kk' will (will not) respond in sth category in the 2nd period, $kk' = AB$ or BA, $s = 0, 1, ..., G$.

Define, The conditional probability,

$$P\left[X_{kk's} = 1 \mid X_{k0} = x_{k0}, X_{k1} = x_{k1}, ..., X_{kG} = x_{kG}\right]$$

$$= \frac{1}{(1+\alpha)}\left[\pi_{kk's} + \sum_{i=0}^{G} \alpha_i \left\{(\pi_{kk's} - \pi_{ki}) + x_{ki}\right\}\right]; \alpha = \sum_{i=0}^{G} \alpha_i$$

$$= \pi_{kk's}(\underset{\sim}{x}), (say) \ \ ; \underset{\sim}{x} = (x_{k0}, x_{k1}, ..., x_{kG}) \ \exists 0 < \pi_{kk's} < 1 \ \forall s$$

Now the marginal probability will be,

$p[X_{kk's} = 1] =$

$$\sum_{x_{k0}, x_{k1}, ..., x_{kG} \ni \sum_{s=0}^{G} \pi_{ks} = 1} p[X_{kk's} = 1 | X_{k0} = x_{k0}, X_{k1} = x_{k1}, ..., X_{kG} = x_{kG}].p[X_{k0} = x_{k0}, X_{k1} = x_{k1}, ..., X_{kG} = x_{kG}]$$

$$= \frac{1}{(1+\alpha)} \sum_{x_{k0}, x_{k1}, ..., x_{kG} \ni \sum_{s=0}^{G} \pi_{ks} = 1} \left[\pi_{kk's} + \sum_{i=0}^{G} \alpha_i \left\{(\pi_{kk's} - \pi_{ki}) + x_{ki}\right\}\right].\pi_{k0}^{x_{k0}}.\pi_{k1}^{x_{k1}}......\pi_{kG}^{x_{kG}}$$

$$= \frac{1}{(1+\alpha)}(\pi_{k0}[\pi_{kk's} + \alpha_0 (\pi_{kk's} - \pi_{k0} + 1) + \alpha_1 (\pi_{kk's} - \pi_{k1} + 0) + \cdots + \alpha_G (\pi_{kk's} - \pi_{kG} + 0)]$$

$$+\pi_{k1}[\pi_{kk's} + \alpha_0 (\pi_{kk's} - \pi_{k0} + 0) + \alpha_1 (\pi_{kk's} - \pi_{k1} + 1) + \cdots + \alpha_G (\pi_{kk's} - \pi_{kG} + 0)]$$

$$+ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots +$$

$$\pi_{kG}[\pi_{kk's} + \alpha_0\left(\pi_{kk's} - \pi_{k0} + 0\right) + \alpha_1\left(\pi_{kk's} - \pi_{k1} + 1\right) + \cdots + \alpha_G\left(\pi_{kk's} - \pi_{kG} + 0\right)])$$

$$= \frac{1}{(1+\alpha)}[(\pi_{kk's}(\pi_{k0} + \pi_{k1} + \cdots + \pi_{kG}) + \alpha_0\left(\pi_{kk's} - \pi_{k0}\right).(\pi_{k0} + \pi_{k1} + \cdots + \pi_{kG}) + \alpha_0\pi_{k0})$$

$$+\alpha_1\left(\pi_{kk's} - \pi_{k1}\right).(\pi_{k0} + \pi_{k1} + \cdots + \pi_{kG} + \alpha_1\pi_{k1} + \cdots\cdots + \alpha_G\left(\pi_{kk's} - \pi_{kG}\right).(\pi_{k0} + \pi_{k1} + \cdots + \pi_{kG}) + \alpha_G\pi_{kG})]$$

$$= \frac{1}{(1+\alpha)}\left[\pi_{kk's} + \alpha_0\pi_{kk's} + \alpha_1\pi_{kk's} + \cdots\cdots + \alpha_G\pi_{kk's}\right]$$

$$= \frac{\pi_{kk's}}{(1+\alpha)}(1 + \alpha_0 + \alpha_1 + \cdots\cdots + \alpha_G) = \frac{\pi_{kk's}}{(1+\alpha)}.(1+\alpha) = \pi_{kk's} \in (0,1); \sum_{s=0}^{G}\pi_{kk's} = 1$$

Hence,

$$p\left[X_{kk'0} = x_{kk'0}, X_{kk'1} = x_{kk'1}, \ldots\ldots\ldots, X_{kk'G} = x_{kk'G} | X_{k0} = x_{k0}, X_{k1} = x_{k1}, \ldots\ldots\ldots, X_{kG} = x_{kG}\right]$$

$$= \pi_{kk'0}^{x_{kk'0}}(\underset{\sim}{x}^{(k)}).\pi_{kk'1}^{x_{kk'1}}(\underset{\sim}{x}^{(k)})\ldots\ldots\pi_{kk'G}^{x_{kk'G}}(\underset{\sim}{x}^{(k)});$$

where ; $0 < \pi_{kk's}(\underset{\sim}{x}^{(k)}) < 1$, $\sum_s \pi_{kk's}(\underset{\sim}{x}^{(k)}) = 1$, $x_{kk's} = 0, 1 \ni \sum_s x_{kk's} = 1$.

$$\implies (X_{kk'0}, X_{kk'1}, \ldots\ldots\ldots, X_{kk'G})|_{\underset{\sim}{x}} \sim MN(1, \pi_{kk'0}(\underset{\sim}{x}^{(k)}).\pi_{kk'1}(\underset{\sim}{x}^{(k)})\ldots\ldots\pi_{kk'G}(\underset{\sim}{x}^{(k)})); \sum_s \pi_{kk's}(\underset{\sim}{x}^{(k)}) = 1$$

Also marginally, $(X_{kk'0}, X_{kk'1}, \ldots\ldots\ldots, X_{kk'G}) \sim MN(1, \pi_{kk'0}, \pi_{kk'1}, \ldots\ldots, \pi_{kk'G}); \sum_s \pi_{kk's} = 1$.
Note that,
$\pi_{Ai} = \pi_{Bi} \forall i$ and $\pi_{ABi}(\underset{\sim}{x}^{(A)}) = \pi_{BAi}(\underset{\sim}{x}^{(B)}) \forall i$

$$\implies \frac{1}{(1+\alpha)}\left[\pi_{ABi} + \sum_{s=0}^{G}\alpha_s\left\{(\pi_{ABs} - \pi_{As}) + x_{As}\right\}\right] = \frac{1}{(1+\alpha)}\left[\pi_{BAi} + \sum_{s=0}^{G}\alpha_s\left\{(\pi_{BAs} - \pi_{Bs}) + x_{Bs}\right\}\right]\forall i$$

$$\implies \pi_{ABi} + (\sum_s \alpha_s)\pi_{ABi} - \sum_s \alpha_s\pi_{As} + \sum_s \alpha_s x_{As} = \pi_{BAi} + (\sum_s \alpha_s)\pi_{BBi} - \sum_s \alpha_s\pi_{Bs} + \sum_s \alpha_s x_{Bs} \forall i$$

$$\implies (1+\alpha)\pi_{ABi} + \sum_s \alpha_s x_{As} = (1+\alpha)\pi_{BAi} + \sum_s \alpha_s x_{Bs} \forall i$$

Now , if we take a look on cumulative probabilities of less than type-

$$\gamma_{As} = p[X_{A0} = 1 or X_{A1} = 1 or \cdots\cdots or X_{As} = 1]$$

$$= \pi_{A0} + \pi_{A1} + \cdots\cdots + \pi_{As}$$

And consider a function $g(.)$ such that;

$$g(x) = log\frac{x}{1-x}$$

Now let,

$$log\frac{\gamma_{As}}{1-\gamma_{As}} = \alpha_A + \psi(s);$$

$$log\frac{\gamma_{Bs}}{1-\gamma_{Bs}} = \alpha_B + \psi(s)$$

where , $\psi(s)$ is a monotonic increasing function of s.

$$\gamma_{As} = \gamma_{Bs} \forall s \implies \alpha_A = \alpha_B$$

Also,

$$\pi_{kk's}(\underset{\sim}{x}) = \frac{1}{(1+\alpha)}\left[\pi_{kk's} + \sum_{i=0}^{G}\alpha_i\left\{(\pi_{kk's} - \pi_{ki}) + x_{ki}\right\}\right]; \alpha = \sum_{i=0}^{G}\alpha_i$$

$$\implies \gamma_{kk's}(\underset{\sim}{x}) = \sum_{j=0}\pi_{kk'j}(\underset{\sim}{x}) = \frac{1}{(1+\alpha)}\left[(1+\alpha)\sum_{j=0}^{G}\pi_{kk'j} + (s+1)\sum_{i=0}^{G}\alpha_i\left\{(x_{ki} - \pi_{ki})\right\}\right]$$

$$= \gamma_{kk's} - \frac{s+1}{1+\alpha}(\sum_{i=0}^{G}\alpha_i\pi_{ki}) + \frac{s+1}{1+\alpha}\underset{\sim}{\alpha}'\underset{\sim}{x}_{\mu}$$

where,

$$log\frac{\gamma_{kk's}}{1-\gamma_{kk's}} = \alpha'_k + \psi(s); \psi(s) \uparrow s.$$

Now consider a new variable Z such that;

$$Z_{ks} = \sum_{i=0}^{s} X_{ki} \sim Ber(\gamma_{ks}) \forall s = 0, 1, \cdots, G$$

$$Z_{kk's} = \sum_{i=0}^{s} X_{kk'i} \sim Ber(\gamma_{kk's}) \forall s = 0, 1, \cdots, G$$

So,

$$g\left(\bar{Z}_{As}\right) = log\left(\frac{\bar{Z}_{As}}{1-\bar{Z}_{As}}\right)$$

$$g'\left(\bar{Z}_{As}\right) = \frac{1}{\left(\frac{\bar{Z}_{As}}{1-\bar{Z}_{As}}\right)}\left(\frac{\left(1-\bar{Z}_{As}\right)+\bar{Z}_{As}}{\left(1-\bar{Z}_{As}\right)^2}\right) = \frac{1}{\bar{Z}_{As}\left(1-\bar{Z}_{As}\right)}$$

Using Delta Method one get,

$$g\left(\bar{Z}_{As}\right) \longrightarrow^{\infty} N\left(log\frac{\gamma_{As}}{1-\gamma_{As}}, \frac{1}{\gamma_{As}^2(1-\gamma_{As})}\right)$$

$$\equiv N\left(\alpha_A + \psi_1(s), \frac{1}{N_1\gamma_{As}(1-\gamma_{As})}\right)$$

Simillarly,

$$g\left(\bar{Z}_{Bs}\right) \longrightarrow^{\infty} N\left(\alpha_B + \psi_1(s), \frac{1}{N_2\gamma_{Bs}(1-\gamma_{Bs})}\right)$$

Therefore as, $Z_{As} \sim Ber(\gamma_{As}) \forall s = 0, 1, \cdots, G$;

$$\implies (Z_{Ar}, Z_{As} - Z_{Ar}) \sim MN(1, \gamma_{Ar}, \gamma_{As} - \gamma_{Ar}), r < s$$

$$\implies cov(Z_{Ar}, Z_{As} - Z_{Ar}) = -\gamma_{Ar}(\gamma_{As} - \gamma_{Ar}); r < s$$

$$\implies cov(Z_{Ar}, Z_{As}) - var(Z_{Ar}) = -\gamma_{Ar}\gamma_{As} + \gamma_{Ar}^2, r < s$$

$$\implies cov(Z_{Ar}, Z_{As}) = -\gamma_{Ar}\gamma_{As} + \cancel{\gamma_{Ar}^2} + \gamma_{Ar} - \cancel{\gamma_{Ar}^2} = \gamma_{Ar}(1 - \gamma_{As}), r < s$$

$$\implies cov(\overline{Z}_{Ar}, \bar{Z}_{As}) = cov(\frac{1}{N_1}\sum_{i=1}^{N_1} Z_{Ari}, \frac{1}{N_1}\sum_{i=1}^{N_1} Z_{Asi}) = \frac{N_1}{N_1^2} cov(Z_{Ar}, Z_{As}) = \frac{\gamma_{Ar}(1-\gamma_{As})}{N_1}$$

And

$$E\begin{pmatrix} \bar{Z}_{A0} \\ \bar{Z}_{A1} \\ . \\ . \\ . \\ \bar{Z}_{AG} \end{pmatrix} = \begin{pmatrix} \gamma_{A0} \\ \gamma_{A1} \\ . \\ . \\ . \\ \gamma_{AG} \end{pmatrix}$$

$$Disp\begin{pmatrix} \bar{Z}_{A0} \\ \bar{Z}_{A1} \\ . \\ . \\ . \\ \bar{Z}_{A\overline{G-1}} \end{pmatrix} = \frac{1}{N_1}\begin{pmatrix} \gamma_{A0}(1-\gamma_{A0}) & \gamma_{A0}(1-\gamma_{A1}) & . \; . \; . & \gamma_{A0}(1-\gamma_{A\overline{G-1}}) \\ \gamma_{A0}(1-\gamma_{A1}) & \gamma_{A1}(1-\gamma_{A1}) & . \; . \; . & \gamma_{A1}(1-\gamma_{A\overline{G-1}}) \\ . & . & . \; . \; . & . \\ . & . & . \; . \; . & . \\ . & . & . \; . \; . & . \\ \gamma_{A0}(1-\gamma_{A\overline{G-1}}) & \gamma_{A1}(1-\gamma_{A\overline{G-1}}) & . \; . \; . & \gamma_{A\overline{G-1}}(1-\gamma_{A\overline{G-1}}) \end{pmatrix}$$

Now we have to calculate $cov(g(x), g(y)$;
Using Taylor's expansion ,

$$g(x) = g(\mu_x) + \frac{(x-\mu_x)}{1!}g'(\mu_x) + \frac{(x-\mu_x)^2}{2!}g''(\mu_x) + \cdots\cdots\cdots$$

So, $E(g(x)) = g(\mu_x) \implies g(x) - g(\mu_x) = \frac{(x-\mu_x)}{1!}g'(\mu_x) + \frac{(x-\mu_x)^2}{2!}g''(\mu_x) + \cdots\cdots\cdots$; and

$$E(g(x) - g(\mu_x))^2 = \sigma_y^2(g(\mu_y))^2;$$

$$E((g(x) - g(\mu_x))(g(y) - g(\mu_y))) = E[(x-\mu_x)(y-\mu_y)]g'(\mu_x)g'(\mu_y)$$

$$\implies cov(g(x), g(y)) = \sigma_{xy}g'(\mu_x)g'(\mu_y)$$

Therefore using this we can say;

$$cov\left(log\frac{\bar{Z}_{Ar}}{1-\bar{Z}_{Ar}}, log\frac{\bar{Z}_{As}}{1-\bar{Z}_{As}}\right) = \frac{\cancel{\gamma_{Ar}(1-\gamma_{As})}}{N_1} . \frac{1}{\cancel{\gamma_{Ar}}(1-\gamma_{Ar})} . \frac{1}{\gamma_{As}\cancel{(1-\gamma_{As})}} = \frac{1}{N_1(1-\gamma_{Ar})\gamma_{As}}; r < s$$

So we can see that;

$$\begin{pmatrix} g(\bar{Z}_{A0}) \\ g(\bar{Z}_{A1}) \\ . \\ . \\ . \\ g(\bar{Z}_{A\overline{G-1}}) \end{pmatrix} \longrightarrow^{n\to\infty} N_G\left[\begin{pmatrix} log\frac{\gamma_{A0}}{1-\gamma_{A0}} \\ log\frac{1}{1-\gamma_{A1}} \\ . \\ . \\ . \\ log\frac{\gamma_{A\overline{G-1}}}{1-\gamma_{A\overline{G-1}}} \end{pmatrix}, \frac{1}{N_1}\underbrace{\begin{bmatrix} \frac{1}{(1-\gamma_{A0})\gamma_{A0}} & \frac{1}{(1-\gamma_{A0})\gamma_{A1}} & . \; . \; . & \frac{1}{(1-\gamma_{A0})\gamma_{A\overline{G-1}}} \\ \frac{1}{(1-\gamma_{A0})\gamma_{A1}} & \frac{1}{(1-\gamma_{A1})\gamma_{A1}} & . \; . \; . & \frac{1}{(1-\gamma_{A\overline{G-1}})\gamma_{A1}} \\ . & . & . \; . \; . & . \\ . & . & . \; . \; . & . \\ . & . & . \; . \; . & . \\ \frac{1}{(1-\gamma_{A0})\gamma_{A\overline{G-1}}} & \frac{1}{(1-\gamma_{A1})\gamma_{A\overline{G-1}}} & . \; . \; . & \frac{1}{(1-\gamma_{A\overline{G-1}})\gamma_{A\overline{G-1}}} \end{bmatrix}}_{\Sigma_A(say)}\right]$$

Simillarly, we can say the same thing for $\left(g(\bar{Z}_{B0}), g(\bar{Z}_{B1}), ..., g(\bar{Z}_{B\overline{G-1}})\right)'$.

Now, from this two ——————————

$$\begin{pmatrix} g(\bar{Z}_{A0}) - g(\bar{Z}_{B0}) \\ g(\bar{Z}_{A1}) - g(\bar{Z}_{B1}) \\ . \\ . \\ . \\ g(\bar{Z}_{A\overline{G-1}}) - g(\bar{Z}_{B\overline{G-1}}) \end{pmatrix} \longrightarrow^{\infty} N_G \left[\begin{pmatrix} \alpha_A - \alpha_B \\ \alpha_A - \alpha_B \\ . \\ . \\ . \\ \alpha_A - \alpha_B \end{pmatrix}, \left(\frac{1}{N_1} \Sigma_A + \frac{1}{N_2} \Sigma_B \right) \right]$$

•Testing Procedure:

Consider the null hypothesis;

$$H_0 : \alpha_A = \alpha_B = \alpha; (say)$$

$$\because log \frac{\gamma_{As}}{1 - \gamma_{As}} = \alpha + \psi(s)$$

$$\implies \frac{\gamma_{As}}{1 - \gamma_{As}} = e^{\alpha + \psi(s)}$$

$$\implies \gamma_{As} = \frac{e^{\alpha + \psi(s)}}{1 + e^{\alpha + \psi(s)}} = \gamma_{Bs};$$

Where α be the common value of $\alpha_A$and α_B.

$$\Rightarrow \gamma_{As}(1 - \gamma_{Ar}) = \frac{e^{\alpha + \psi(s)}}{1 + e^{\alpha + \psi(s)}} \frac{1}{1 + e^{\alpha + \psi(r)}} \forall 0 \leq r \leq s \leq G - 1$$

Then, as under $H_0, \alpha_A = \alpha_B$ then , $\Sigma_A = \Sigma_B = \Sigma^{(1)}_{H_0}$

$$\implies \begin{pmatrix} g(\bar{Z}_{A0}) - g(\bar{Z}_{B0}) \\ g(\bar{Z}_{A1}) - g(\bar{Z}_{B1}) \\ . \\ . \\ . \\ g(\bar{Z}_{A\overline{G-1}}) - g(\bar{Z}_{B\overline{G-1}}) \end{pmatrix} \longrightarrow^{H_0}_{n \to \infty} N_G \left[\underset{\sim}{0}, \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \Sigma^{(1)}_{H_0} \right]$$

$$\implies \sum_{s=0}^{G-1} \left(g(\bar{Z}_{As}) - g(\bar{Z}_{Bs}) \right) \longrightarrow^{\infty}_{H_0} N_1 \left(0, \left(\frac{1}{N_1} + \frac{1}{N_2} \right) \underset{\sim}{1}' \Sigma^{(1)}_{H_0} \underset{\sim}{1} \right)$$

$$\tau^{(1)}_{H_0} = \frac{\sum_{s=0}^{G-1} \left(g(\bar{Z}_{As}) - g(\bar{Z}_{Bs}) \right)}{\sqrt{\left(\frac{1}{N_1} + \frac{1}{N_2} \right) \underset{\sim}{1}' \Sigma^{(1)}_{H_0} \underset{\sim}{1}}} \sim_{H_0} N(0, 1)$$

Reject $H_0 : \alpha_A = \alpha_B$ in favour of $H_1 : \alpha_A > \alpha_B$ if $\tau^{(1)}_{H_0} > \tau_k$ at $100k\%$ level of significance. In simillar way a large sample test can be formed for 2nd period where,

$$\tau^{(2)}_{H_0} = \frac{\sum_{s=0}^{G-1} \left(g(\bar{Z}_{BAs}) - g(\bar{Z}_{ABs}) \right)}{\sqrt{\left(\frac{1}{N_1} + \frac{1}{N_2} \right) \underset{\sim}{1}' \Sigma^{(2)}_{H_0} \underset{\sim}{1}}} \sim_{H_0} N(0, 1)$$

⊙Formation of combined Test Statistic:

(Choice:1) ▷ Chatterjee & De Test Statistic

Let,

$$\underset{\sim}{\tau}_{H_0} = \left(\tau_{H_0}^{(1)}, \tau_{H_0}^{(2)}\right) \longrightarrow \left(\hat{\underset{\sim}{\mu}}, \hat{\sum}\right)$$

based on all sample values.
Now,

$$\mu_i = \left(\underset{\sim}{\tau}_{H_{0i}} - \hat{\underset{\sim}{\mu}}\right)' \hat{\sum}^{-1} \left(\underset{\sim}{\tau}_{H_{0i}} - \hat{\underset{\sim}{\mu}}\right) \forall i = 1(1)n$$

Again if

$$\underset{\sim}{\tau}_{H_0} \sim BVN\left(\underset{\sim}{\mu}, \sum\right)$$

$$\Longrightarrow \left(\underset{\sim}{\tau}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\underset{\sim}{\tau}_{H_0} - \underset{\sim}{\mu}\right) \sim \chi_2^2$$

Then, say, $\widehat{\rho} = corr(\tau_{H_0}^{(1)}, \tau_{H_0}^{(2)})$ based on n paired data; taken in two stages;

$$Q_n = \sqrt{\frac{\tau_{H_{0n}}^{(1)^2} + \tau_{H_{0n}}^{(2)^2} - 2\widehat{\rho}_n \tau_{H_{0n}}^{(1)}, \tau_{H_{0n}}^{(2)}}{(1 - \widehat{\rho}_n^2)}}; if \tau_{H_0}^{(1)}, \tau_{H_0}^{(2)} > 0$$

$$= \frac{\tau_{H_0}^{(2)} - \widehat{\rho}\tau_{H_0}^{(1)}}{\sqrt{(1 - \widehat{\rho}^2)}}; if \tau_{H_0}^{(1)} < \tau_{H_0}^{(2)}, \tau_{H_0}^{(1)} < 0$$

$$= \frac{\tau_{H_0}^{(1)} - \widehat{\rho}\tau_{H_0}^{(2)}}{\sqrt{(1 - \widehat{\rho}^2)}}; if \tau_{H_0}^{(1)} > \tau_{H_0}^{(2)}, \tau_{H_0}^{(2)} < 0$$

$$= 0; otherwise$$

(Choice:2) ▷ Mean of two Statistic(D_n)

Based on these 2 stage test statistics $\tau_{H_0}^{(1)}, \tau_{H_0}^{(2)}$ another choice of combined mode test statistics is ;

$$D_n = \tau_{mean} = \frac{\left(\tau_{H_0}^{(1)} + \tau_{H_0}^{(2)}\right)}{2}$$

(Choice:2) ▷ Maximum of those two Statistics (W_n)

Based on these 2 stage test statistics $\tau_{H_0}^{(1)}, \tau_{H_0}^{(2)}$ another choice of combined mode test statistics is ;

$$W_n = \tau_{max} = max\left\{\tau_{H_0}^{(1)}, \tau_{H_0}^{(2)}\right\}$$

therefore these 3 statistics Q_n, $D_n$and W_n will be used for further purposes.

⊙ Choice of $\psi(s)$:

Here $\psi(s)$ is a monotic increasing function s. Here after taking several choices i.e; e^s,s^2, $\frac{s}{s+1}$ etc. we can emperically say that $log(s)$, $log(s^2+1)$ etc $log(polynomial_s)$ will be arguabaly best choices of $\psi(s)$. Therefore here 3of log polynomials have been choosen:

- $$log(s)$$
- $$log(s^2+1)$$
- $$log(s^3+s^2-1)$$

List of All Outcome Table:

Sr.	Plot Name	Figure/Table	$\psi(s)$	δ	n
1	Power vs Size with certain α	Figure1	$log\,(s)$	$0.1-0.5$	50-1000
2	Power vs Size with certain α	Figure 7	$log(s^2+1)$	$0.1-0.5$	50-1000
3	Power vs Size with certain α	Figure 10	$log(s^3+s^2-1)$	$0.1-0.5$	50-1000
4	Normality Check of test statistics of both stage	Figure 2	$log\,(s)$	0.3	500
5	Normality Check of test statistics of both stage	Figure 4	$log\,(s)$	0.3	50
6	Cdf plot of Chaterjee & De test-stat	Figure 3	$log\,(s)$	0.3	500
7	Normality Check of test statistics of both stage	Figure5	$log(s^2+1)$	0.3	500
8	Normality Check of test statistics of both stage	Figure8	$log(s^3+s^2-1)$	0.3	500
9	Cdf plot of Chaterjee & De test-stat	Figure6	$log(s^2+1)$	0.3	500
10	Cdf plot of Chaterjee & De test-stat	Figure9	$log(s^3+s^2-1)$	0.3	500
11	Power Table with different Size & choices of α	Table1	$log\,(s)$	0.1	50-1000
12	Power Table with different Size & choices of α	Table 2	$log\,(s)$	0.2	50-1000
13	Power Table with different Size & choices of α	Table 3	$log\,(s)$	0.3	50-1000
14	Power Table with different Size & choices of α	Table 4	$log\,(s)$	0.4	50-1000
15	Power Table with different Size & choices of α	Table 5	$log\,(s)$	0.5	50-1000
16	Power Table with different Size & choices of α	Table 6	$log(s^2+1)$	0.1	50-1000
17	Power Table with different Size & choices of α	Table 7	$log(s^2+1)$	0.2	50-1000
18	Power Table with different Size & choices of α	Table 8	$log(s^2+1)$	0.4	50-1000
19	Power Table with different Size & choices of α	Table 9	$log(s^2+1)$	0.5	50-1000
20	Power Table with different Size & choices of α	Table 10	$log(s^3+s^2-1)$	0.1	50-1000
21	Power Table with different Size & choices of α	Table 11	$log(s^3+s^2-1)$	0.2	50-1000
22	Power Table with different Size & choices of α	Table 12	$log(s^3+s^2-1)$	0.4	50-1000
23	Power Table with different Size & choices of α	Table 13	$log(s^3+s^2-1)$	0.5	50-1000

Table 1: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\delta = 0.1$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.7437	0.0608	0.0591
100	0.754	0.0622	0.0674
250	0.799	0.0844	0.0928
500	0.8414	0.0946	0.1341
1000	0.8949	0.559	0.2237

Table 2: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\delta = 0.2$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.799	0.083	0.0854
100	0.8372	0.1032	0.1051
250	0.9047	0.1643	0.1961
500	0.954	0.2638	0.3406
1000	0.9883	0.415	0.5925

*[The table is **hyperlinked properly** with all the plots and tables; so to check that figure just click the number in the Figure/Table coloumn]*

⊙ Computation with $\psi(s) = log(s)$:

⊘Sample Size = 500, $\alpha_A = 0.5$***,*** $\alpha_B = 0.2$ ***:***

1st stage test-statistic and 2nd stage test-statistic both considerably follows $N(0,1)$ assymptotically when $n \longrightarrow \infty$. Here is some plots which will assure us about that.

•Forming of Chatterjee & De Test statistic:

Checking if

$$\left(\mathcal{I}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\mathcal{I}_{H_0} - \underset{\sim}{\mu}\right) \sim \chi^2_2$$

or not; Hence we make a cdf plot, qq-plot (3) and make a **Kolmogorov Smrinov test;**

As of Figure 3 the **CDF** of our used test statistic and chi-square cdf looks almost simillar in this plot. It also becomes another proof of test statistic's distribution is **chi-square** assymptotically.

And the result of Kolmogorov-Smrinov test is;

```
Two-sample Kolmogorov-Smirnov test

data:  u and chi_sq_2
D = 0.0101, p-value = 0.6875
```

Table 3: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\delta = 0.3$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.8273	0.0957	0.109
100	0.8871	0.1296	0.1762
250	0.9628	0.2268	0.396
500	0.9923	0.3443	0.6614
1000	0.9997	0.554	0.923

Figure 1: Power against Size with different δ for $\psi(s) = log(s)$

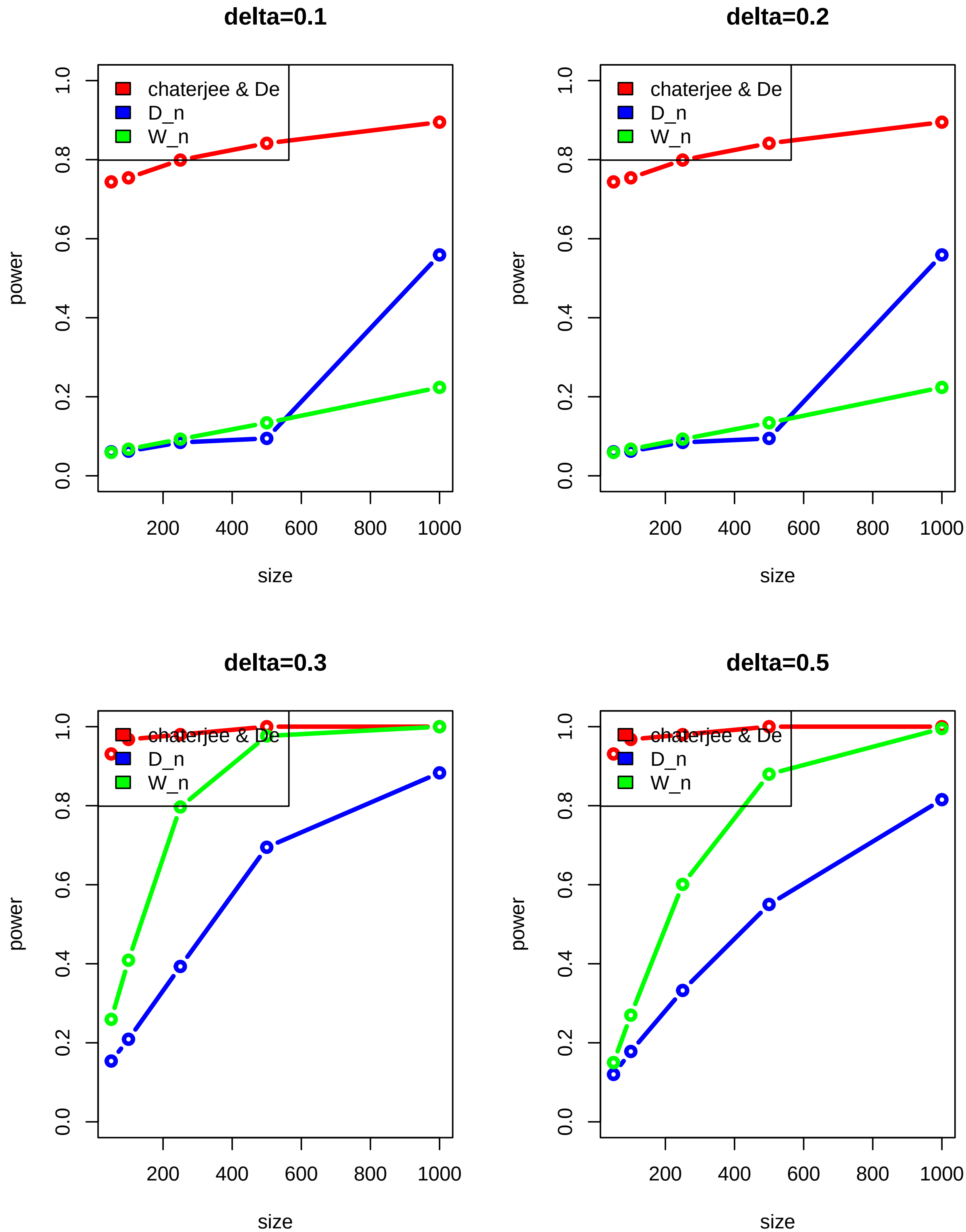


Figure 2: Checking Normality of $\tau^{(1)}_{H_0}$ and $\tau^{(1)}_{H_0}$

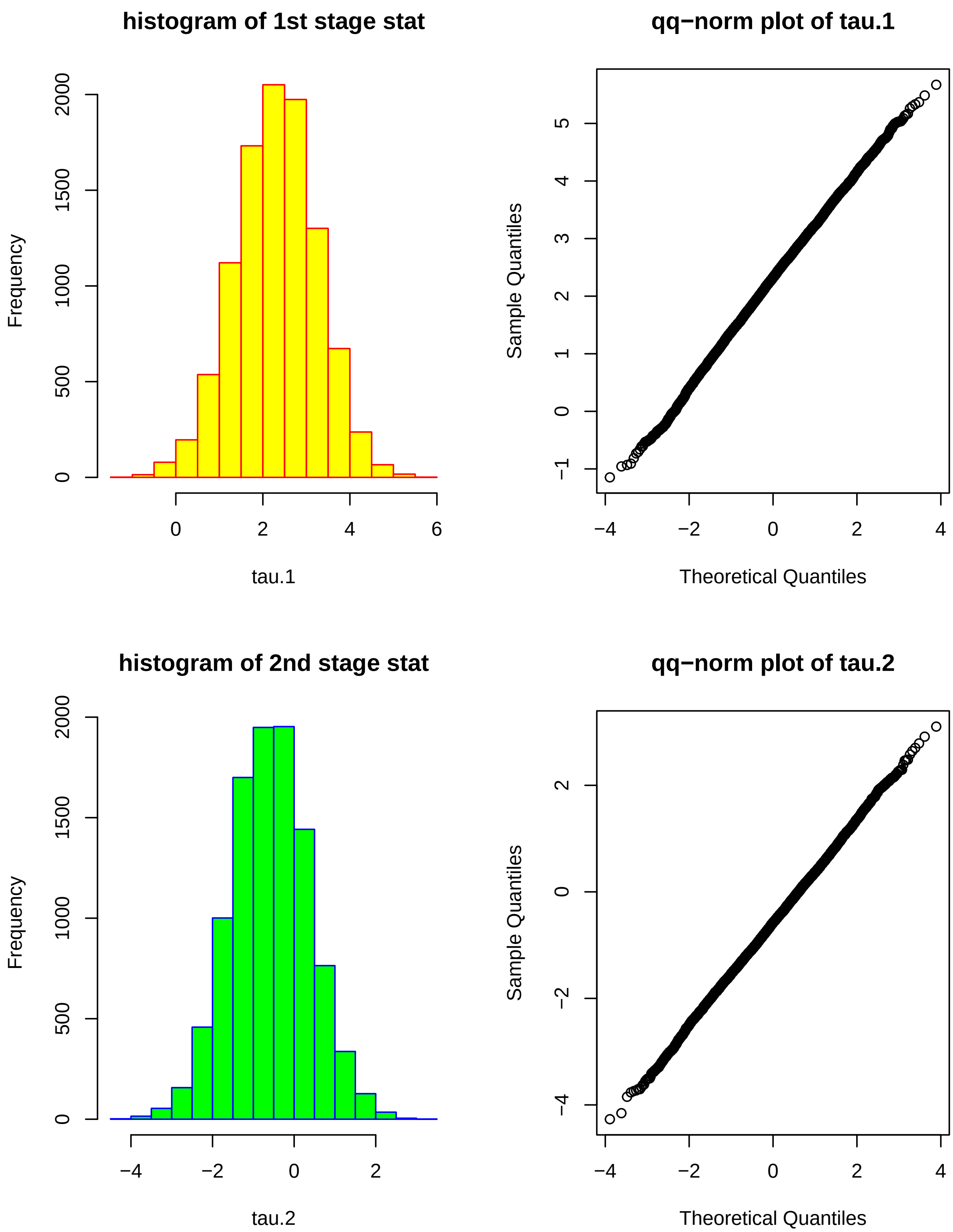


Table 4: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\delta = 0.4$

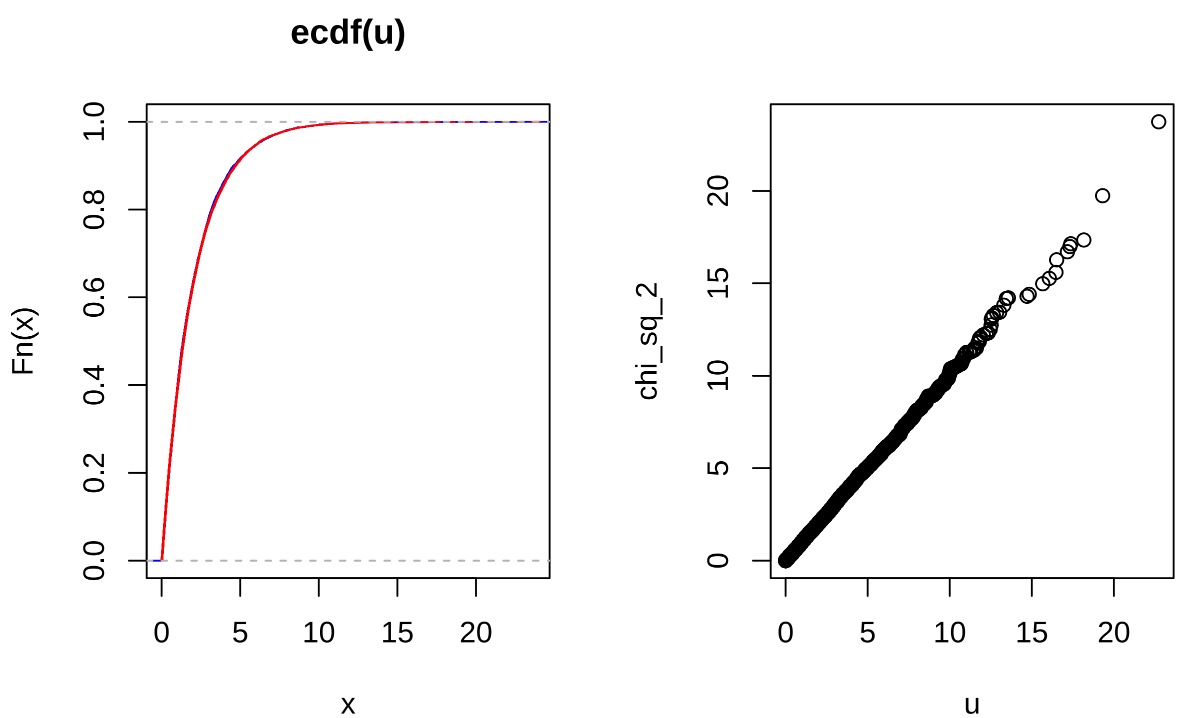
Sample Size	chaterjee & De	Mean test statistic	Maximum of two
50	0.883	0.12	0.15
100	0.938	0.178	0.27
250	0.9879	0.3325	0.6009
500	0.9995	0.5501	0.8798
1000	1	0.8153	0.9952

Table 5: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\delta = 0.5$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.9309	0.1538	0.2593
100	0.9676	0.209	0.409
250	0.9799	0.3932	0.7967
500	1	0.6948	0.9764
1000	1	0.8831	0.9999

Figure 3: Cdf plot of $\left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right) and \chi_2^2$

```
Attaching package:  'dgof'
The following object is masked from 'package:stats':
   ks.test
```



```
alternative hypothesis: two-sided
```

So, we can use **Chaterjee & De test statistic** in this scenario.

⊘***Sample Size*** $=$ ***100,*** $\alpha_A = 0.5$***,*** $\alpha_B = 0.2$ ***:***

Here also the same result breaks out. Therefore all the predefined distributions can be used further in these context.

Simillarly same results will continue for these two choices of $\psi(s)$.

⊙ Remarks:

- The more gap increases the tests become more powerful in case of all 3 test statistics. [Figure 1]
- Number of Sample when are increasing the power gets more improved. [Figure 1]
- **Chatterjee & De** test statistic has a lower rate of increase with respect to the other two. [Figure 1]
- For small δ the rate of increase of Power is higher in less sample points than in high sample points. [Figure 1]
- For big δ the rate of increase of Power is higher in more sample points than in less sample points. [Figure 1]

⊙ Computation with $\psi(s) = log(s^2 + 1)$:

⊘***Sample Size*** $=$ ***500,*** $\alpha_A = 0.5$***,*** $\alpha_B = 0.2$ ***:***

1st stage test-statistic and 2nd stage test-statistic both considerably follows $N(0,1)$ assymptotically when $n \longrightarrow \infty$. Here is some plots which will assure us about that.

•Forming of Chatterjee & De Test statistic:

Checking if

$$\left(\mathcal{I}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\mathcal{I}_{H_0} - \underset{\sim}{\mu}\right) \sim \chi^2_2$$

or not; Hence we make a cdf plot, qq-plot and make a **Kolmogorov Smrinov test;**

The **CDF** of our used test statistic and chi-square cdf looks almost simillar in this plot. It also becomes another proof of test statistic's distribution is **chi-square** assymptotically.

And the result of Kolmogorov-Smrinov test is;

```
Two-sample Kolmogorov-Smirnov test

data:  u and chi_sq_2
D = 0.0112, p-value = 0.5573
alternative hypothesis: two-sided
```

So, we can use **Chaterjee & De test statistic** in this scenario.

Figure 4: Checking Normality of $\tau^{(1)}_{H_0}$ and $\tau^{(1)}_{H_0}$

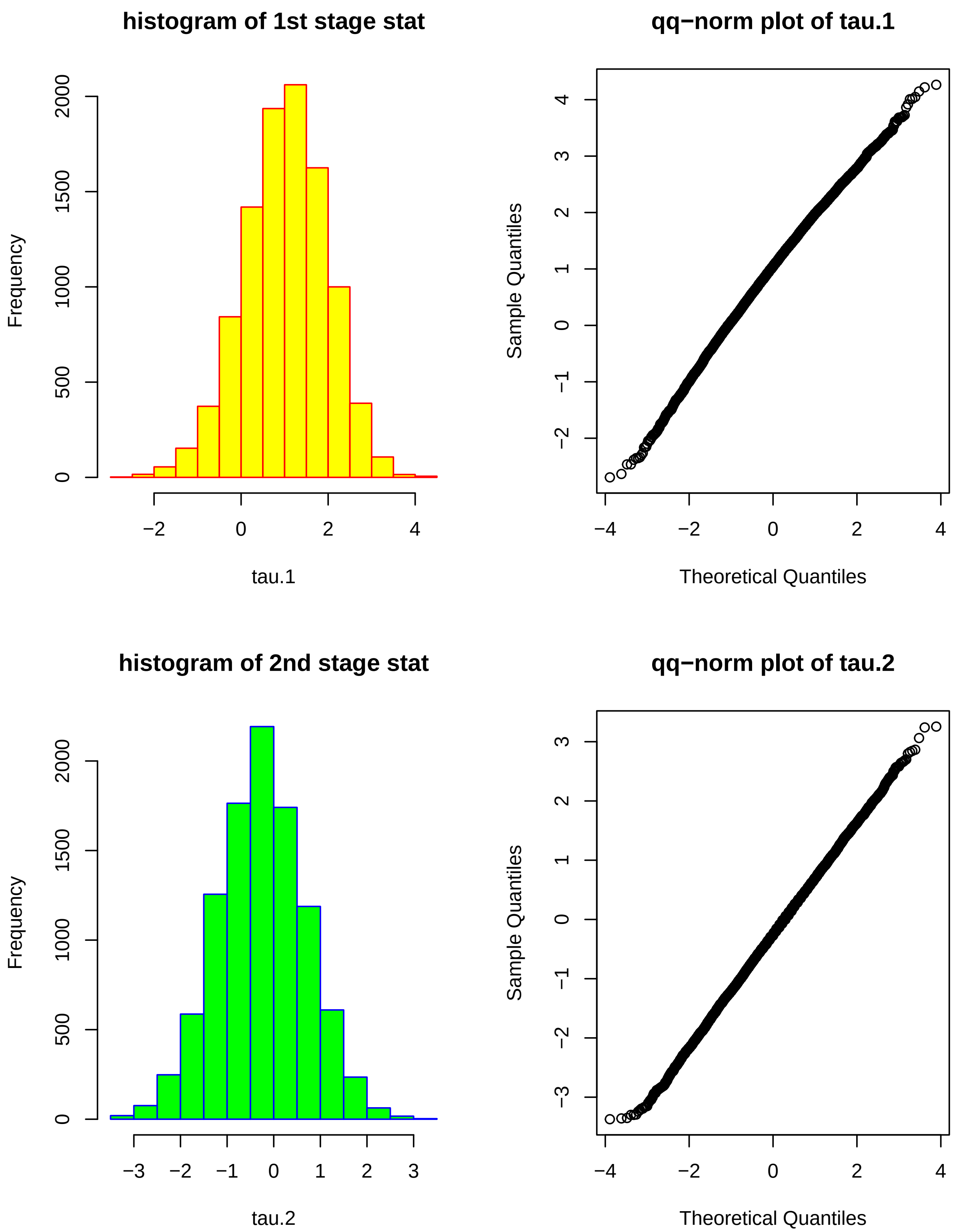


Figure 5: Checking Normality of $\tau_{H_0}^{(1)}$ and $\tau_{H_0}^{(1)}$

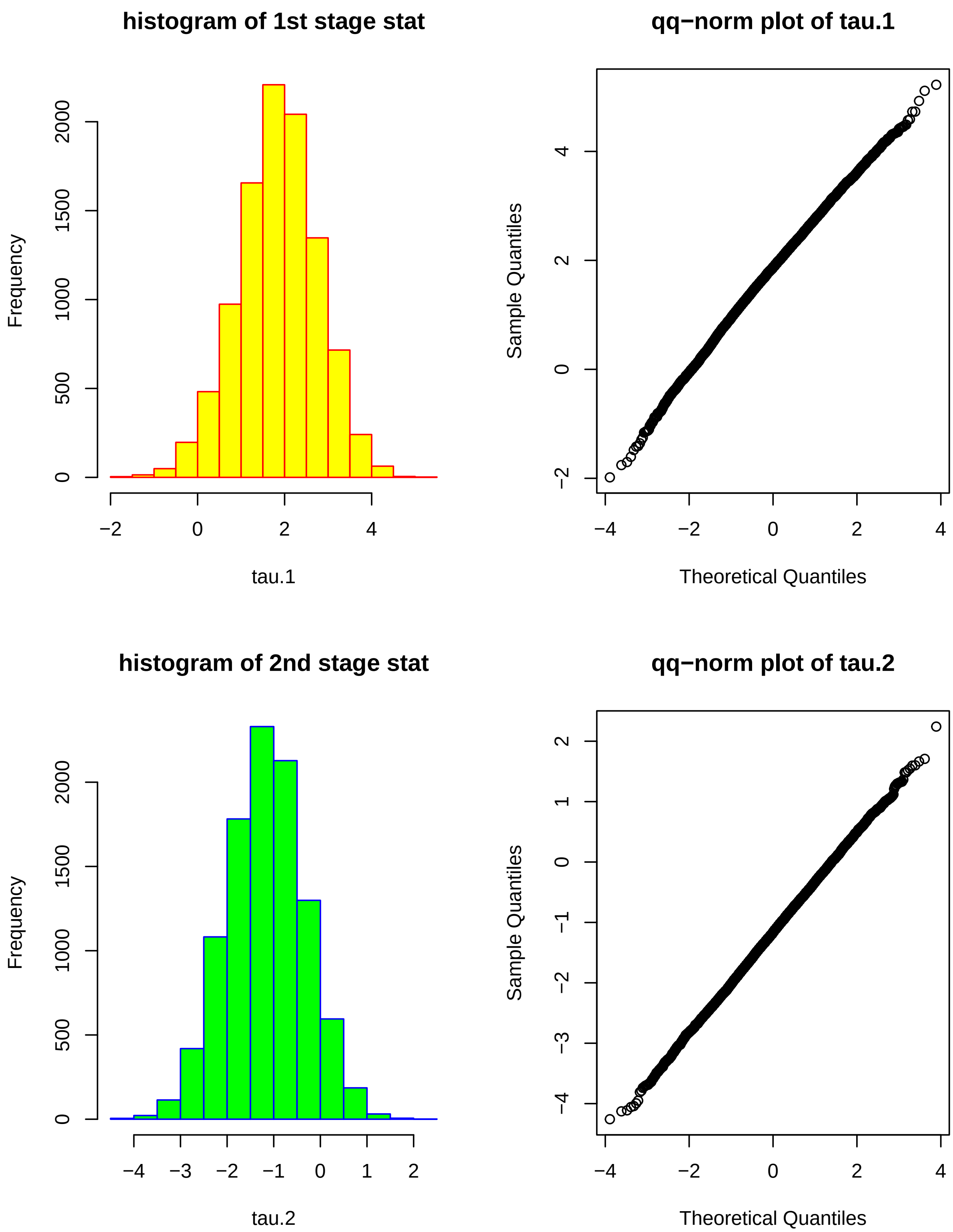


Figure 6: Cdf plot of $\left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right) and \chi^2_2$

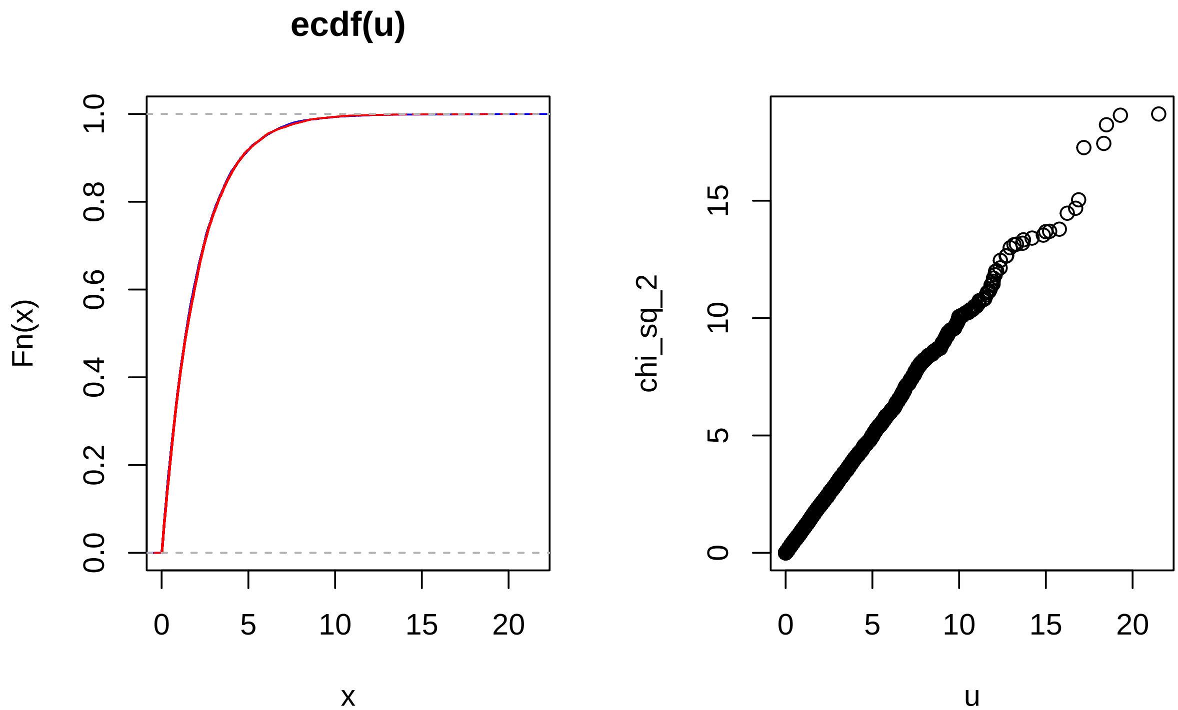


Table 6: Table of **Power** against different choices of n for $\psi(s) = log(s^2 + 1)$ and $\alpha_A = 0.5, \alpha_B = 0.1$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.9001	0.0508	0.16
100	0.9524	0.0585	0.35
250	0.9804	0.0658	0.515
500	0.997	0.0846	0.7434
1000	1	0.1595	0.9608

Table 7: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\alpha_A = 0.5, \alpha_B = 0.2$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.898	0.0536	0.1954
100	0.9027	0.0732	0.2051
250	0.9247	0.0843	0.2461
500	0.9741	0.1638	0.4806
1000	0.9983	0.19415	0.7925

Figure 7: Power against Size with different δ for $\psi(s) = log(s^2 + 1)$

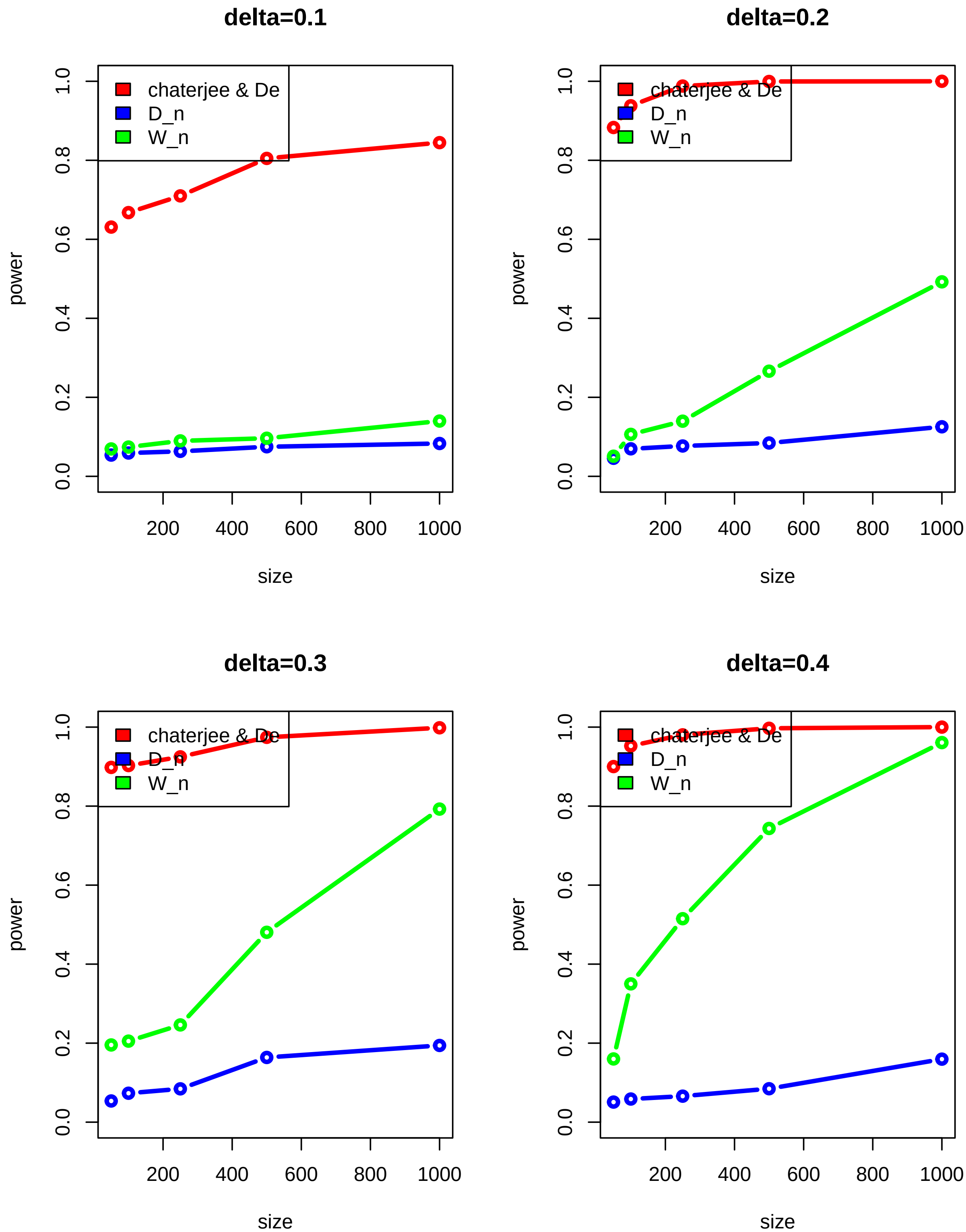


Table 8: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\alpha_A = 0.5, \alpha_B = 0.3$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.7873	0.0457	0.0509
100	0.8371	0.0696	0.1062
250	0.8628	0.0768	0.1396
500	0.9023	0.0843	0.26614
1000	0.9697	0.1254	0.4923

Table 9: Table of **Power** against different choices of n for $\psi(s) = log(s)$ and $\alpha_A = 0.5, \alpha_B = 0.4$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.6309	0.0538	0.0693
100	0.6676	0.059	0.07409
250	0.7099	0.0632	0.08967
500	0.8048	0.0748	0.0964
1000	0.8447	0.0831	0.1399

⊙ Remarks:

From [Figure 7] we see;

- The more gap increases the tests become more powerful in case of all 3 test statistics.
- Number of Sample when are increasing the power gets more improved as power gets more higher.
- **Chatterjee & De** test statistic has a lower rate of increase with respect to the other two.
- For all δ the rate of increase of Power is higher in more sample points than in less sample points.

⊙ Computation with $\psi(s) = log(s^3 + s^2 - 1)$:

⊘***Sample Size* = *500,*** $\alpha_A = 0.5$***,*** $\alpha_B = 0.2$ ***:***

1st stage test-statistic and 2nd stage test-statistic both considerably follows $N(0, 1)$ assymptotically when $n \longrightarrow \infty$. Here is some plots which will assure us about that.

•Forming of Chatterjee & De Test statistic:

Checking if

$$\left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right) \sim \chi^2_2$$

or not; Hence we make a cdf plot, qq-plot and make a **Kolmogorov Smrinov test;**

The **CDF** of our used test statistic and chi-square cdf looks almost simillar in this plot. It also becomes another proof of test statistic's distribution is **chi-square** assymptotically.

And the result of Kolmogorov-Smrinov test is;

```
Two-sample Kolmogorov-Smirnov test

data:  u and chi_sq_2
D = 0.0084, p-value = 0.8722
alternative hypothesis: two-sided
```

So, we can use **Chaterjee & De test statistic** in this scenario.

Figure 8: Checking Normality of $\tau_{H_0}^{(1)}$ and $\tau_{H_0}^{(1)}$

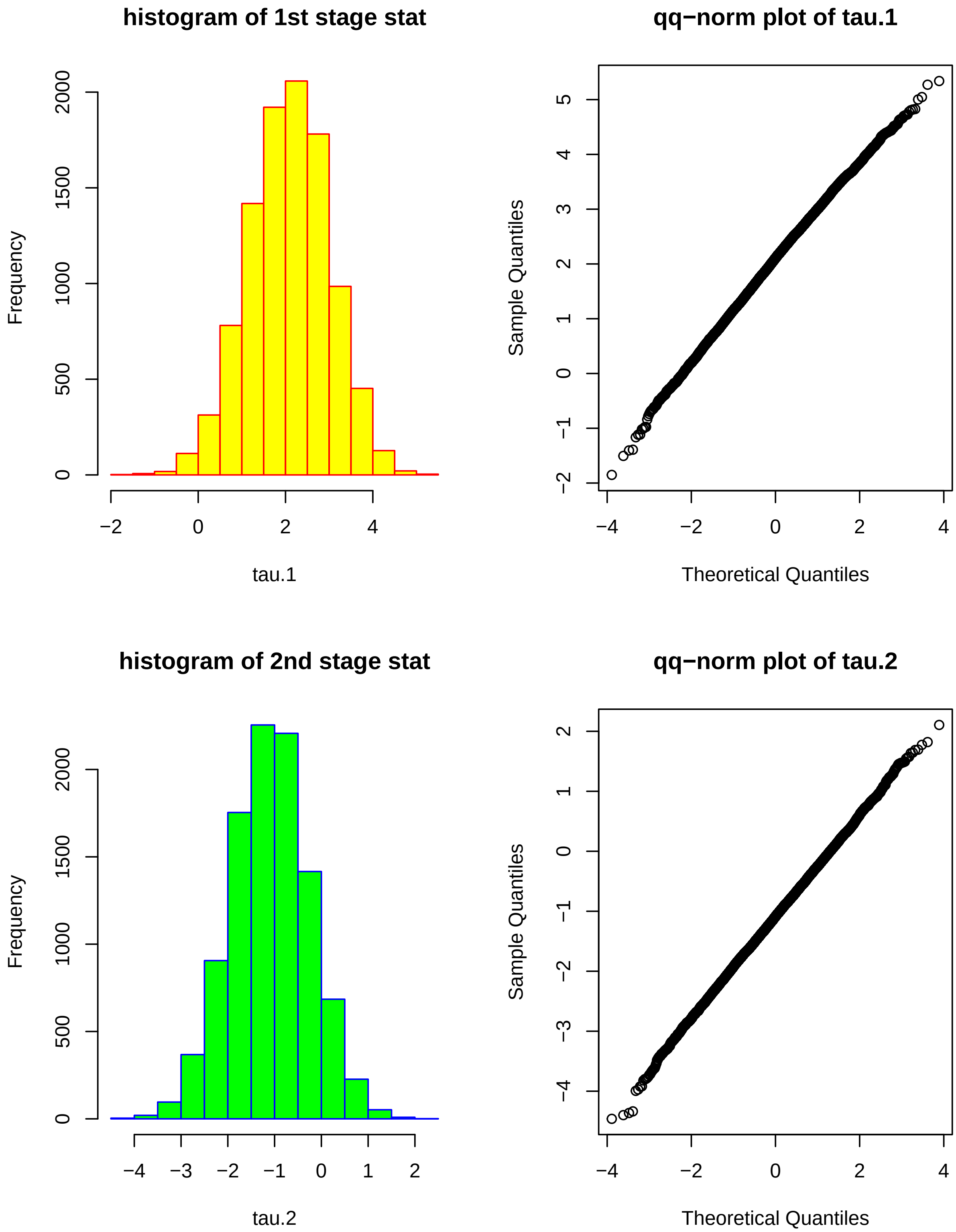


Figure 9: Cdf plot of $\left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\underset{\sim}{T}_{H_0} - \underset{\sim}{\mu}\right) and \chi^2_2$

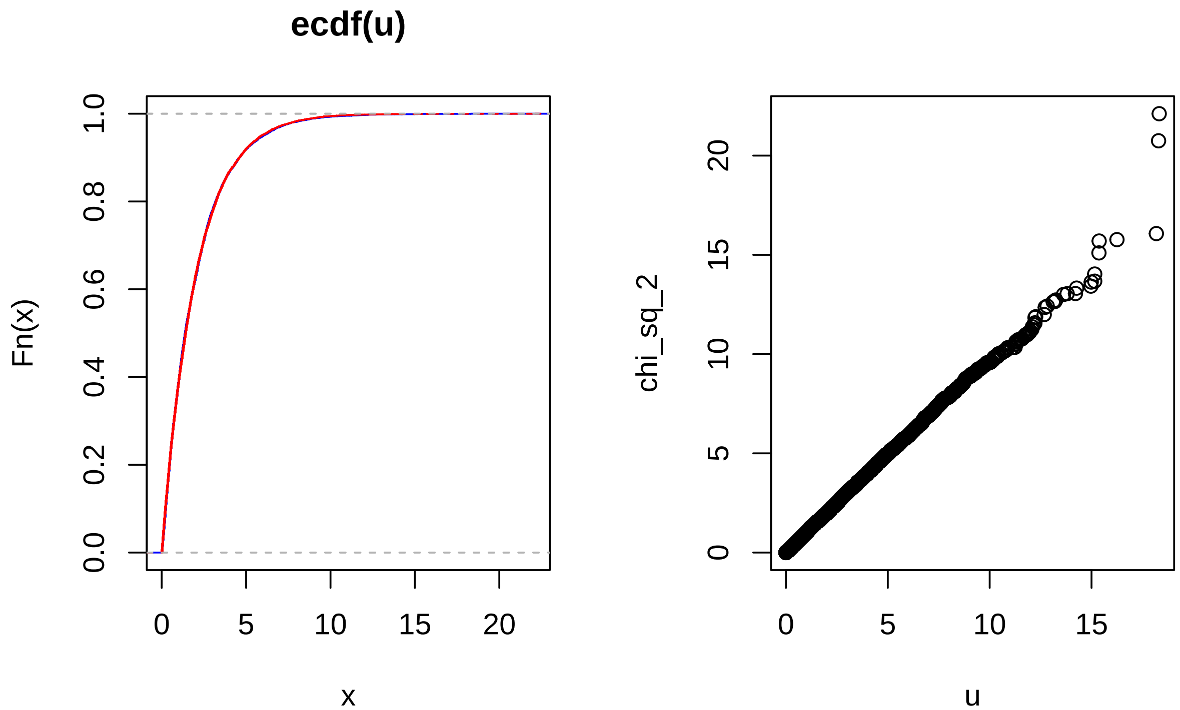


Table 10: Table of **Power** against different choices of n for $\psi(s) = log(s^3 + s^2 - 1)$ and $\alpha_A = 0.6, \alpha_B = 0.4$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.7655	0.0501	0.0606
100	0.8072	0.0578	0.075
250	0.8686	0.0958	0.1515
500	0.9341	0.1446	0.27434
1000	0.9761	0.2295	0.49608

Table 11: Table of **Power** against different choices of n for $\psi(s) = log(s^3 + s^2 - 1)$ and $\alpha_A = 0.5, \alpha_B = 0.15$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.8898	0.0636	0.1754
100	0.9427	0.0832	0.3351
250	0.9747	0.1143	0.4861
500	0.9941	0.1638	0.7206
1000	0.9993	0.2394	0.9425

Figure 10: Power against Size with different δ for $\psi(s) = log(s^3 + s^2 - 1)$

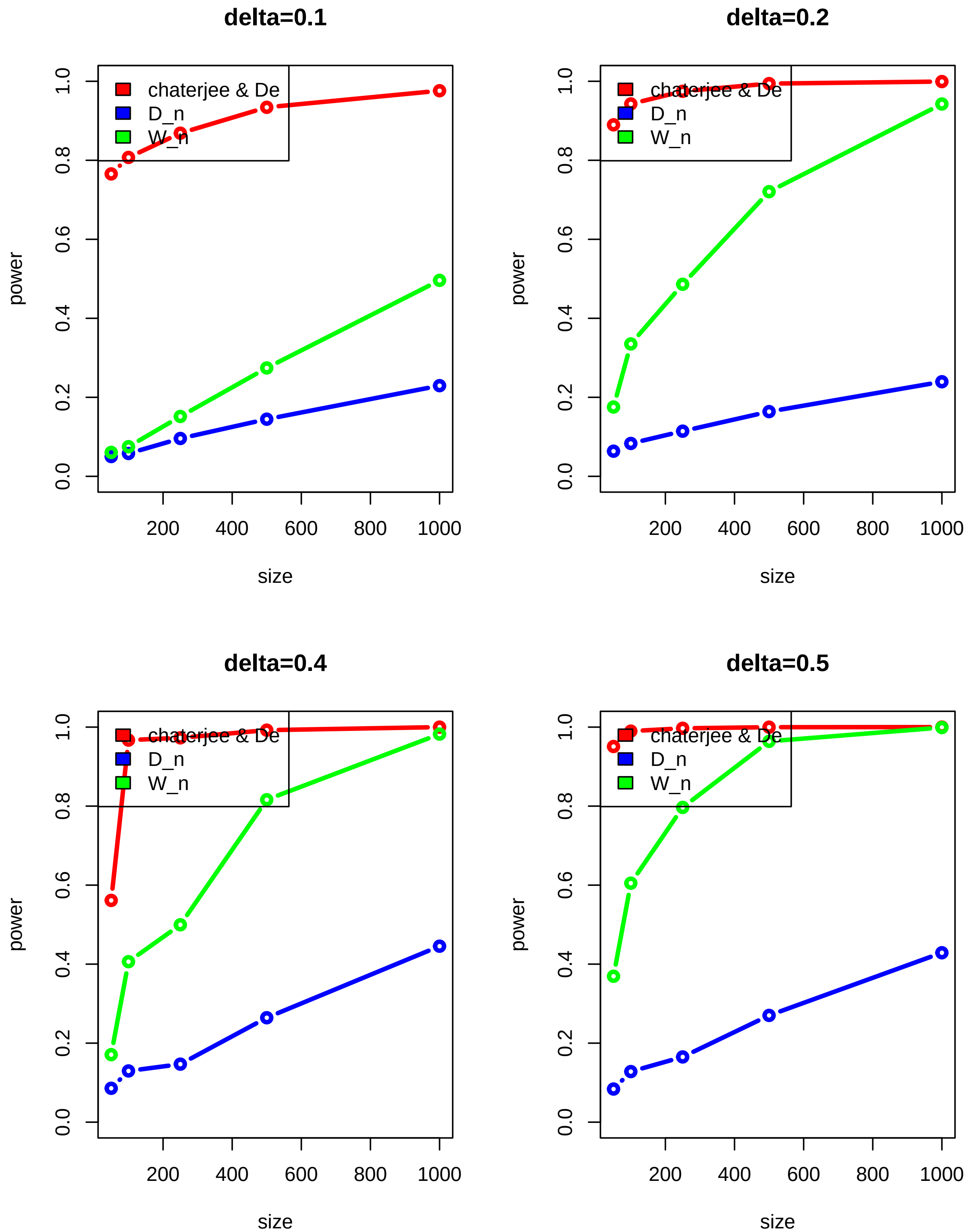


Table 12: Table of **Power** against different choices of n for $\psi(s) = log(s^3 + s^2 - 1)$ and $\alpha_A = 0.6, \alpha_B = 0.2$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.5613	0.0857	0.1709
100	0.9671	0.1296	0.4062
250	0.9728	0.1468	0.4996
500	0.9923	0.2643	0.81614
1000	1	0.4454	0.9823

Table 13: Table of **Power** against different choices of n for $\psi(s) = log(s^3 + s^2 - 1)$ and $\alpha_A = 0.6, \alpha_B = 0.1$

Sample Size	Chaterjee & De (Q_n)	D_n	W_n
50	0.9509	0.0838	0.3693
100	0.9898	0.128	0.6049
250	0.997	0.1651	0.7967
500	0.9998	0.2702	0.964
1000	1	0.4288	0.9989

⊙ **Remarks:**

From [Figure 7] we see;

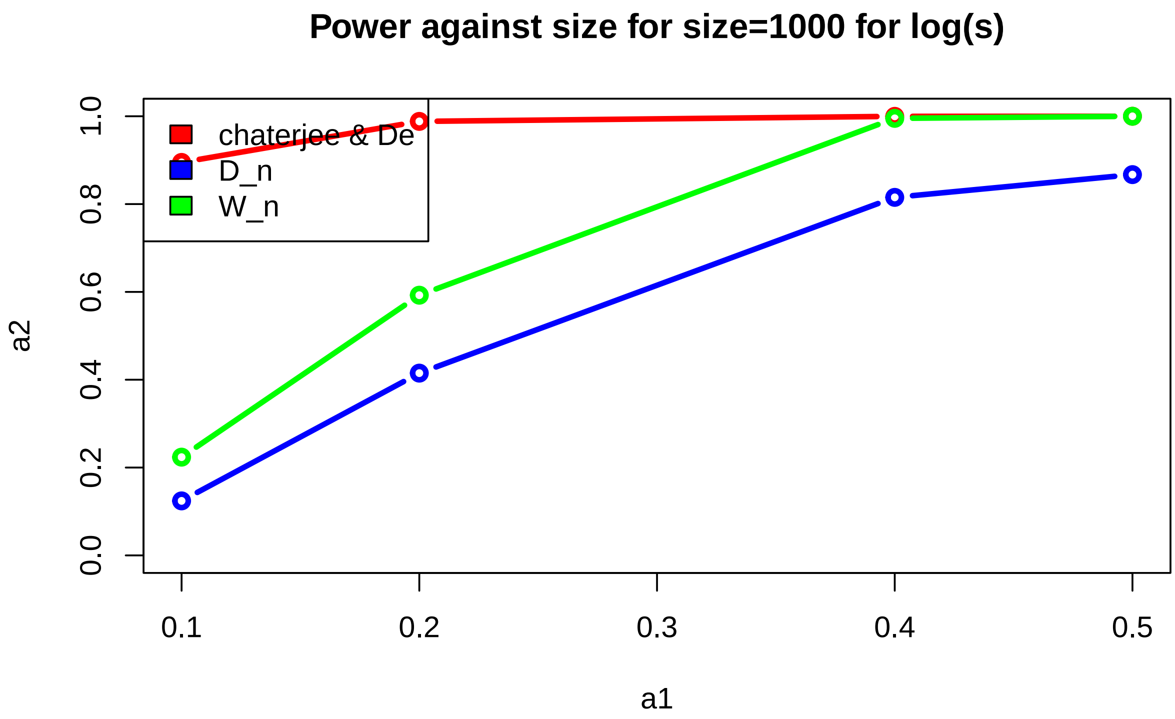
- The more gap increases the tests become more powerful in case of all 3 test statistics.
- Number of Sample when are increasing the power gets more improved.
- **Chatterjee & De** test statistic has a lower rate of increase with respect to the other two.
- For $\delta = 0.1$ the rate of increase of Power is almost same.
- For big $\delta = 0.2$ the rate of increase of Power is almost same for Chatterjee & De and mean stat , wheather the max. statistic power has increased expotentially. And the same story continues wheather the δ increases.

⊙ Comparison in Q_n, D_n and W_n:

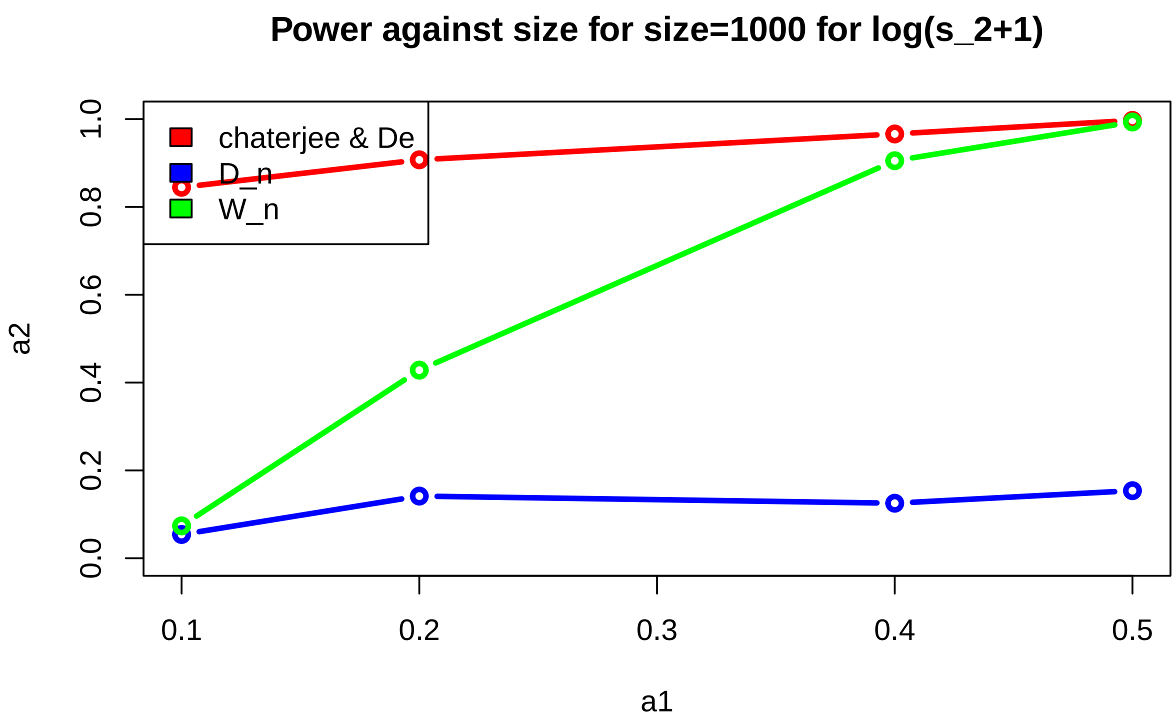
To make a decision on this topic we have choose the way to plot power against δ for two choices of Size ;

Choice 1: $n = 1000$

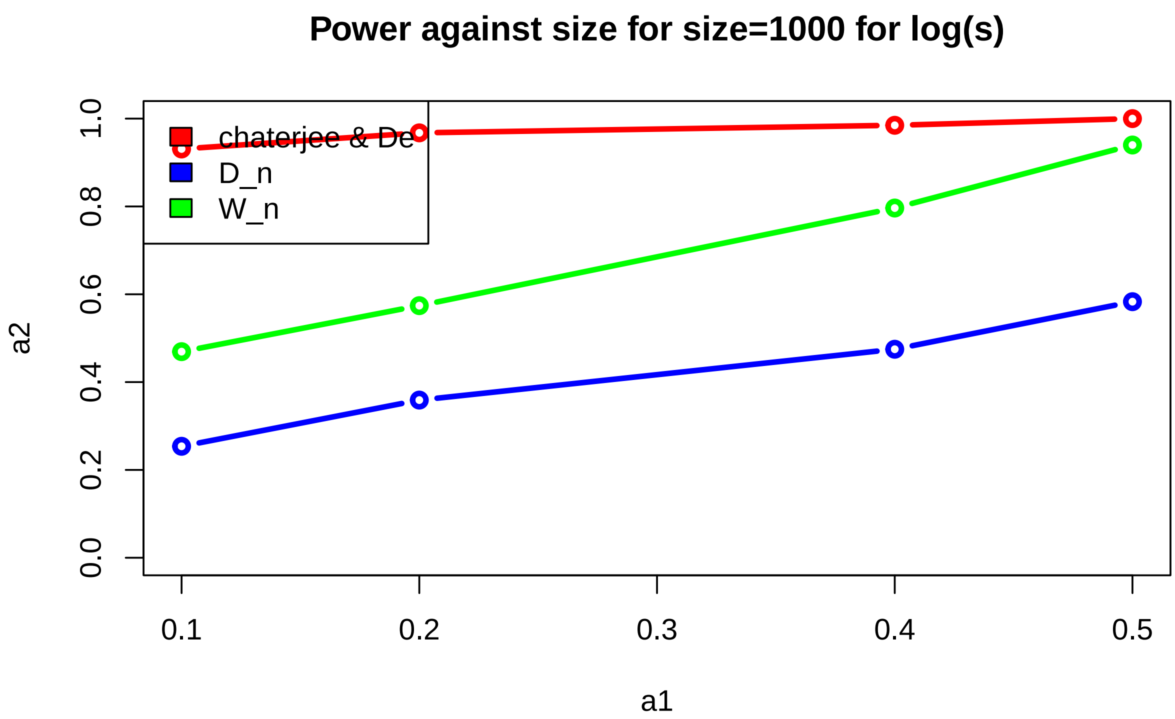
$\psi(s) = log(s)$:


Power against size for size=1000 for log(s)
chaterjee & De
D_n
W_n
a2
a1
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5

$\psi(s) = log(s^2 + 1)$:

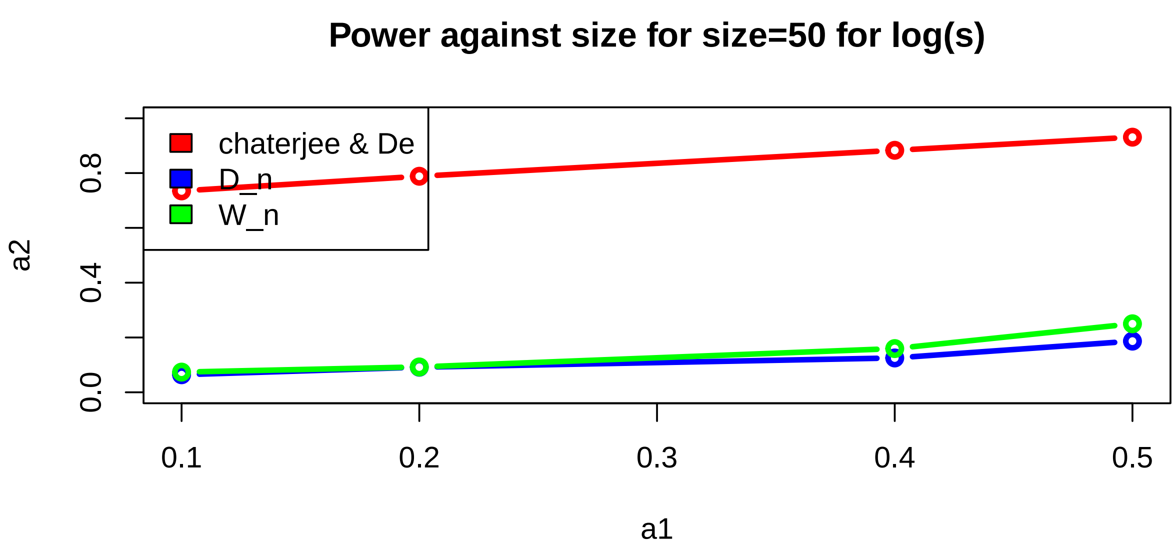

Power against size for size=1000 for log(s_2+1)
chaterjee & De
D_n
W_n
a2
a1
0.0
0.2
0.4
0.6
0.8
1.0
0.1
0.2
0.3
0.4
0.5

$\psi(s) = log(s^3 + s^2 - 1)$:

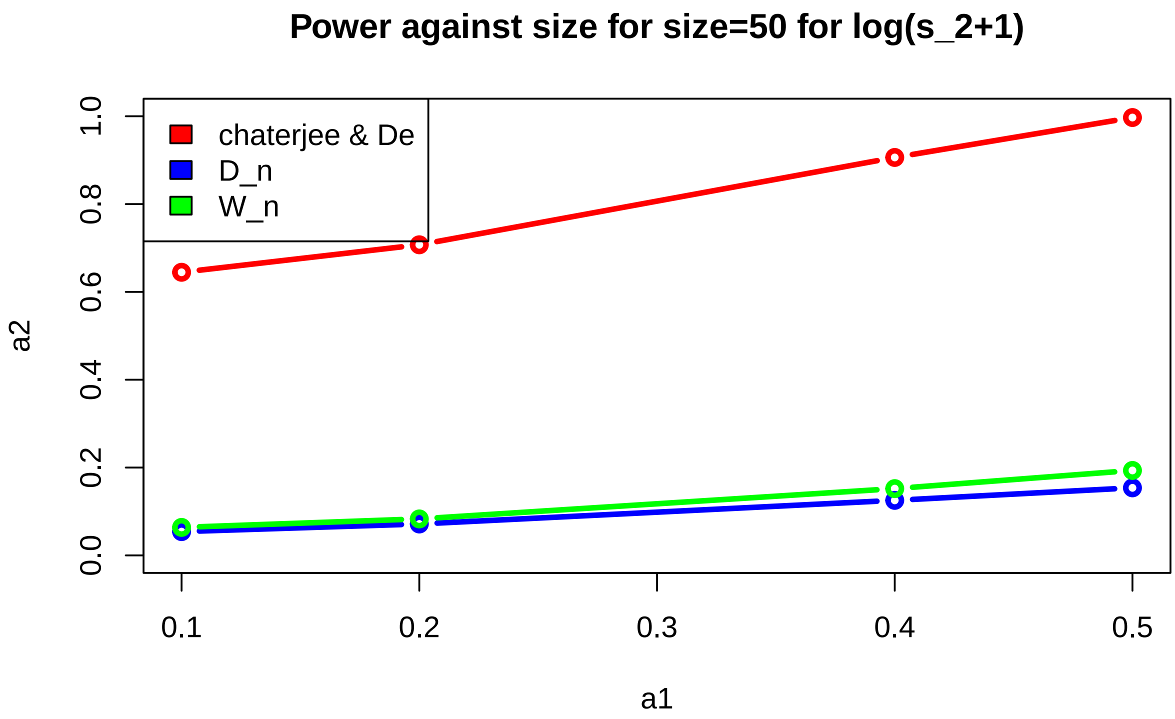


Choice 2: $n = 50$

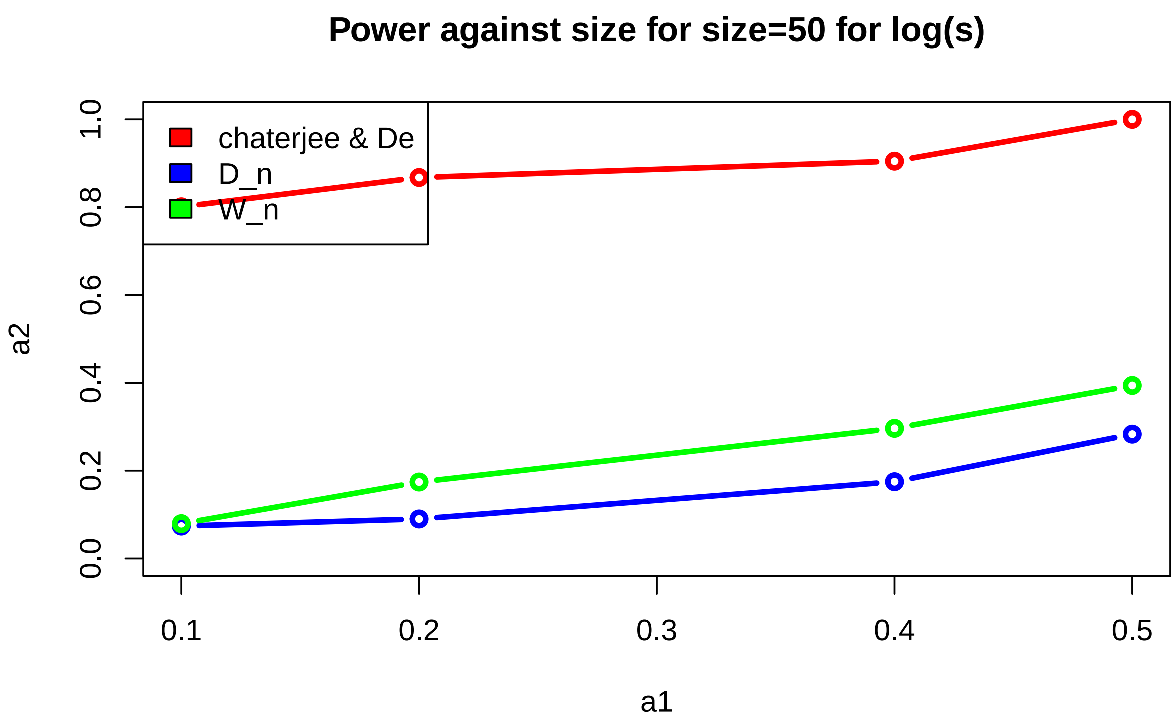
$\psi(s) = log(s)$:



$\psi(s) = log(s^2 + 1)$:



$\psi(s) = log(s^3 + s^2 - 1)$:



Remarks:

This diagram also tells us the same story that Chatterje & De statistic is far more better than other two in all aspect. That is why I have been used it will preferably used in more advance role.

⊙Conclusion: [Future Scope]

In Statistical field of study Clinical trial has grown as one of the major topics . A most important part of this kind of work is to maitain the code of ethics. Here in our project we are in a search for better cross-over treatment which will perform better in **case of cronic diseases.**

But to acompolish this goal we consider some limitations;

- The choice of $\psi(s)$ is taken limitedly; which can be explored in a more vast way .
- Here no covariates regarding patient has been considered; such as **age, gender, height, weight etc.** If it can be inserted in the model the model will be surely more complicated but will become more accurate too; and will help us to do our best in these fields.
- Here i have taken only upto 2nd period ; but in future it can be extended to more catagory to get more best treatment with better result.
- We have worked here with only **two** treatments; to get more better cross-over treatments regarding this.
- We have observed that when two drugs are equally effective (i.e their distributions are identical) then the mean proportion of allocation remains around 0.5 for both the treatments which is what we expect. Even though test yield almost constant power (at some low value) for varying choices of n and hence again here clinical precision is maintained while compromising statistical precision as it still does not reach our desired value.

References:

1. Atanu Biswasa;∗, Jing-Shiang Hwangb (2002). A new bivariate binomial distribution ;tatistics & Probability Letters 60 (2002) 231 – 240
2. Snourm Krsnonu Chatterjee, Nanda Kisor de (1972). Calcutta Statistics Association
3. The Use and Reporting of the Cross-Over Study Design in Clinical Trials and Systematic Reviews: A Systematic Assessment Sarah Jane Nolan ,Ian Hambleton,Kerry Dwan; 2016; `https://doi.org/10.1371/journal.pone.0159014`
4. Design, analysis, and presentation of crossover trials; DOI:10.1186/1745-6215-10-27, 2009; Edward J Mills, An-Wen Chan , Ping Wu , Andy Vail .
5. Power and Sample Size Calculations in Clinical Trials with Patient-Reported Outcomes under Equal and Unequal Group Sizes Based on Graded Response Model: A Simulation Study; (2016) Marziyeh Doostfatemeh, Seyyed Mohammad Taghi Ayatollah*, Peyman Jafar; journal homepage: `www.elsevier.com/locate/jval`

⊙Appendix:

Lab Work:

Basically we have used R for all the calculations and simulation of dataset. Here the whole R_code for 3 category ordinal response has been added. In a simillar way one can just extend the category to 4.

```
rm(list=ls())
n = 10 #ideally 1000
t = 100 #ideally 10000
alpha_a = 0.5
alpha_b = 0.2
tau.1 = NULL
tau.2 = NULL
```

The function to get γ_{ks} from known value of α :

```
gamma.log<- function(x,a){
  g = exp(a + log(x))/(1 + exp(a + log(x)))
  return(g)
}
```

$g(.)$ and σ_{xy} also defined below;

```
g.func<- function(x){
  func_value = log(x/(1-x))
  return(func_value)
}

sigma_xy<- function(x,y){
  sig_xy = 1/((1-x)*y)
  return(sig_xy)
}
```

The main random variable i.e the indicator variable which will describe in what category the response belongs is ;

```
indicator.func <- function(z,i,s){
  z = c(rep(0,n))
  if(z[i] == s){
    x = 1
  }
  else
    x = 0
  return(x)
}
```

Now with help of these above predefined functions here all the marginal and conditional probabilities are calculated with given α_A, α_B values:

```
pie.A.1 = gamma.log(1,alpha_a)
pie.A.2 = gamma.log(2,alpha_a) - gamma.log(1,alpha_a)
pie.A.3 = 1 - pie.A.1 - pie.A.2

pie.B.1 = gamma.log(1,alpha_b)
pie.B.2 = gamma.log(2,alpha_b) - gamma.log(1,alpha_b)
pie.B.3 = 1 - pie.B.1 - pie.B.2


#2nd stage
```

```
#consider differnet alpha for getting different gamma
#for log(s)
pie.AB.1 = gamma.log(1, alpha_b)
pie.AB.2 = gamma.log(2, alpha_b) - gamma.log(1, alpha_b)
pie.AB.3 = 1 - pie.AB.1 - pie.AB.2

pie.BA.1 = gamma.log(1, alpha_a)
pie.BA.2 = gamma.log(2, alpha_a) - gamma.log(1, alpha_a)
pie.BA.3 = 1 - pie.BA.1 - pie.BA.2

#conditional probability
cond.prob.AB.1 = (1/(1+alpha_b))*(pie.AB.1 + (alpha_b/3)*((pie.AB.1 - pie.A.1)+ indicator.func(ZA,1,1))
cond.prob.AB.2 = (1/(1+alpha_b))*(pie.AB.2 + (alpha_b/3)*((pie.AB.2 - pie.A.1)+ indicator.func(ZA,1,1))
cond.prob.AB.3 = 1 - cond.prob.AB.1 - cond.prob.AB.2

cond.prob.BA.1 = (1/(1+alpha_a))*(pie.BA.1 + (alpha_a/3)*((pie.BA.1 - pie.B.1)+ indicator.func(ZB,1,1))
cond.prob.BA.2 = (1/(1+alpha_a))*(pie.BA.2 + (alpha_a/3)*((pie.BA.2 - pie.B.1)+ indicator.func(ZB,1,1))
cond.prob.BA.3 = 1 - cond.prob.BA.1 - cond.prob.BA.2
```

Now the Sample genration and estimating the γ_{ks} and $\gamma_{kk's}$:

```
for(i in 1:t){
  ZA=as.vector(c(1,2,3)%*%rmultinom(n, size = 1, prob = c(pie.A.1, pie.A.2, pie.A.3)))
  ZB=as.vector(c(1,2,3)%*%rmultinom(n, size = 1, prob = c(pie.B.1, pie.B.2, pie.B.3)))

  #calculation stage 1

  fa1=sum(ZA==1)
  fa2=sum(ZA==2)
  fa3=sum(ZA==3)


  fb1=sum(ZB==1)
  fb2=sum(ZB==2)
  fb3=sum(ZB==3)

  #calculating Z-bar:
  ca1=(fa1)/n
  ca2=(fa1 + fa2)/n
  ca3=(fa1 + fa2 + fa3)/n

  cb1=(fb1)/n
  cb2=(fb1 + fb2)/n
  cb3=(fb1 + fb2 + fb3)/n

  gamma.est.A.1 = fa1/n
  gamma.est.A.2 = fa1/n + fa2/n
  gamma.est.A.3 = fa1/n + fa2/n +fa3/n

  gamma.est.B.1 = fb1/n
  gamma.est.B.2 = fb1/n + fb2/n
  gamma.est.B.3 = fb1/n + fb2/n +fb3/n

  sum.1 = sum((g.func(ca1)-(g.func(cb1))), (g.func(ca2)-(g.func(cb2))))
```

```
sig.11.A = sigma_xy(gamma.est.A.1, gamma.est.A.1)
sig.12.A = sigma_xy(gamma.est.A.1, gamma.est.A.2)
sig.21.A = sigma_xy(gamma.est.A.1, gamma.est.A.2)
sig.22.A = sigma_xy(gamma.est.A.2, gamma.est.A.2)

SIGMA_1.A = matrix(c(sig.11.A, sig.12.A,  sig.21.A, sig.22.A) , nrow = 2)

sig.11.B = sigma_xy(gamma.est.B.1, gamma.est.B.1)
sig.12.B = sigma_xy(gamma.est.B.1, gamma.est.B.2)
sig.21.B = sigma_xy(gamma.est.B.1, gamma.est.B.2)
sig.22.B = sigma_xy(gamma.est.B.2, gamma.est.B.2)

SIGMA_1.B = matrix(c(sig.11.B, sig.12.B,  sig.21.B, sig.22.B) , nrow = 2)
SIGMA_1.hat = matrix(c(mean(sig.11.A,sig.11.B), mean(sig.12.A,sig.12.B), mean(sig.21.A,sig.21.B), mean
tau_1 = sum.1/sqrt((2*sum(SIGMA_1.hat))/n)  #test_statistic

tau.1 = c(tau.1,tau_1)


#for stage 2
alpha_a_hat.1 = log(gamma.est.A.1/(1-gamma.est.A.1)) - log(1)
alpha_a_hat.2 = log(gamma.est.A.2/(1-gamma.est.A.2)) - log(2)
alpha_a_hat = (alpha_a_hat.1+alpha_a_hat.2)/2

alpha_b_hat.1 = log(gamma.est.B.1/(1-gamma.est.B.1)) - log(1)
alpha_b_hat.2 = log(gamma.est.B.2/(1-gamma.est.B.2)) - log(2)
alpha_b_hat = (alpha_b_hat.1 + alpha_b_hat.2)/2

alpha_hat = (alpha_a_hat + alpha_b_hat)/2


for(j in 1:n)
{
  if(ZA[j]==1)
  {
    ZAB[j]=c(1,2,3)%*%rmultinom(1,1,prob=c(cond.prob.AB.1, cond.prob.AB.2, cond.prob.AB.3))
  }
  else if(ZA[j]==2)
  {
    ZAB[j]=c(1,2,3)%*%rmultinom(1,1,prob=c(cond.prob.AB.1, cond.prob.AB.2, cond.prob.AB.3))
  }
  else if(ZA[j]==3)
  {
    ZAB[j]=c(1,2,3)%*%rmultinom(1,1,prob=c(cond.prob.AB.1, cond.prob.AB.2, cond.prob.AB.3))
  }

  if(ZB[j]==1)
  {
    ZBA[j]=c(1,2,3)%*%rmultinom(1,1,prob=c(cond.prob.BA.1, cond.prob.BA.2, cond.prob.BA.3))
  }
  else if(ZB[j]==2)
  {
    ZBA[j]=c(1,2,3)%*%rmultinom(1,1,prob=c(cond.prob.BA.1, cond.prob.BA.2, cond.prob.BA.3))
```

```
    }
    else if(ZB[j]==3)
    {
      ZBA[j]=c(1,2,3)%*%rmultinom(1,1,prob=c(cond.prob.BA.1, cond.prob.BA.2, cond.prob.BA.3))
    }
  }


fab1=sum(ZAB==1)
fab2=sum(ZAB==2)
fab3=sum(ZAB==3)


fba1=sum(ZBA==1)
fba2=sum(ZBA==2)
fba3=sum(ZBA==3)


fab=c(fab1,fab2,fab3)
fba=c(fba1,fba2,fba3)

#calculating Z-barfor 2nd stage :
cab1=(fab1)/n
cab2=(fab1 + fab2)/n
cab3=(fab1 + fab2 + fab3)/n

cba1=(fba1)/n
cba2=(fba1 + fba2)/n
cba3=(fba1 + fba2 + fba3)/n

gamma.est.AB.1 = fa1/n
gamma.est.AB.2 = fa1/n + fa2/n
gamma.est.AB.3 = fa1/n + fa2/n +fa3/n

gamma.est.BA.1 = fba1/n
gamma.est.BA.2 = fba1/n + fba2/n
gamma.est.BA.3 = fba1/n + fba2/n +fba3/n

sum.2 = sum((g.func(cba1)-(g.func(cab1))), (g.func(cba2)-(g.func(cab2))))

sig.11.AB = sigma_xy(gamma.est.AB.1, gamma.est.AB.1)
sig.12.AB = sigma_xy(gamma.est.AB.1, gamma.est.AB.2)
sig.21.AB = sigma_xy(gamma.est.AB.1, gamma.est.AB.2)
sig.22.AB = sigma_xy(gamma.est.AB.2, gamma.est.AB.2)

SIGMA_1.AB = matrix(c(sig.11.AB, sig.12.AB,  sig.21.AB, sig.22.AB) , nrow = 2)

sig.11.BA = sigma_xy(gamma.est.BA.1, gamma.est.BA.1)
sig.12.BA = sigma_xy(gamma.est.BA.1, gamma.est.BA.2)
sig.21.BA = sigma_xy(gamma.est.BA.1, gamma.est.BA.2)
sig.22.BA = sigma_xy(gamma.est.BA.2, gamma.est.BA.2)

SIGMA_1.BA = matrix(c(sig.11.BA, sig.12.BA,  sig.21.BA, sig.22.BA) , nrow = 2)
```

```
  SIGMA_2.hat = matrix(c(mean(sig.11.AB,sig.11.BA), mean(sig.12.AB,sig.12.BA), mean(sig.21.AB,sig.21.BA)
  tau_2 = sum.2/sqrt((2*sum(SIGMA_2.hat))/n)  #test_statistic

  tau.2 = c(tau.2,tau_2)
}
```

Now to test the data if applicable for using **Chatterjee & De** test statistic :
Statement Check:

$$\left(\underset{\sim}{\tau}_{H_0} - \underset{\sim}{\mu}\right)' \sum^{-1} \left(\underset{\sim}{\tau}_{H_0} - \underset{\sim}{\mu}\right) \sim \chi^2_2$$

```
u = NULL
tau_0 = cbind(tau.1,tau.2)
mean_tau = apply(tau_0,2,mean)

SIGMA_hat_0 = cov(tau_0)
for(i in 1:t){
  u[i] = (t(tau_0[i,] - mean_tau))%*%solve(SIGMA_hat_0)%*%((tau_0[i,] - mean_tau))
}
```

CDF plot and test of Kolmogorov-Smrinov and Q-Q plot:

```
chi_sq_2 = rchisq(t,2)
#KS test:
#install.packages("dgof")
library("dgof")
plot(ecdf(u),
     xlim = range(c(u, chi_sq_2)),
     col = "blue")
plot(ecdf(chi_sq_2),
     add = TRUE,
     lty = "dashed",
     col = "red")
ks.test(u, chi_sq_2)
#qq plot
qqplot(u,chi_sq_2)
```

Defining Corelation coefficient among the two test statistics;

```
Q = NULL
cor_coef = NULL

cor_coef.1 = cor(tau.1[c(1:10)], tau.2[c(1:10)])
for(i in 1:t){
  if(i<=10){
    cor_coef[i] = cor_coef.1
  }
  else{
    cor_coef[i] = cor(tau.1[c(1:i)], tau.2[c(1:i)])
  }
}
```

Defining the Chatterjee & Das test statistic:

```
for(i in 1:t){
  if(tau.1[i] > 0 && tau.2[i] > 0){
    Q[i] = sqrt((((tau.1[i])^2) + ((tau.2[i])^2) - (2*cor_coef[i]*tau.1[i]*tau.2[i]))/(1 - (cor_coef[i]
  }
  else if(tau.2[i] > tau.1[i] && tau.1[i] < 0){
    Q[i] = (tau.2[i] - (cor_coef[i]*tau.1[i]))/(1 - (cor_coef[i])^2)
  }
  else if(tau.2[i] < tau.1[i] && tau.2[i] < 0){
    Q[i] = (tau.1[i] - (cor_coef[i]*tau.2[i]))/(1 - (cor_coef[i])^2)
  }
  else{
    Q[i] = 0
  }
}
```

To get all the requires emperical size and powers;

```
#chaterjee stat statistic:
chi.value = qchisq(0.05,2,lower.tail = TRUE)
emperical_power_Q = sum((Q > chi.value))/t #chi_square(0.05,2)


#mean test stat
mean_tau = (tau.1 + tau.2)/2
mean_tau_order = sort(mean_tau, decreasing = FALSE)

emperical_power_mean = sum((mean_tau > 1.13343))/t #chi_square(0.05,2)

mean_tau_star = mean_tau/(sqrt((1+cor_coef)/2))
#maximum tau
max_tau = NULL
max_tau = pmax(tau.1, tau.2)
max_tau_order = sort(max_tau, decreasing = FALSE)

emperical_power_max = sum((max_tau > 1.917603))/t #chi_square(0.05,2)

emperical_power_Q
emperical_power_mean
emperical_power_max
```

Plot code:

```
par(mfrow = c(1,1))
a1 = c(0.1,0.2,0.4,0.5)




a2 = c(0.9309,
       0.9676,

       0.9848,
       1
```

```
)
a3 = c(0.2538,
       0.359,

       0.4748,
       0.5831

)
a4 = c(0.4693,
       0.57409,

       0.7964,
       0.9399

)
plot(a1, a2, main = "Power against size for size=1000 for log(s)", ylim = c(0,1), col = "red", lty = 1,
lines(a1, a3, ylim = c(0,1), col = "blue", lty = 1, lwd = 3, type = "b")
lines(a1, a4, ylim = c(0,1), col = "green", lty = 1, lwd = 3, type = "b")
legend(x="topleft",legend = c("chaterjee & De","D_n","W_n"), fill = c("red","blue","green"))
```