Choosing the Best Test Method for Heteroscedasticity in a Multiple Regression Model

(With Practical Application)

اختيار افضل طريقة اختبار لمشكلة عدم تجانس التباين في الانحدار الخطي المتعدد مع تطبيق متعدد

أ.م.د عدنان كريم نجم الدين رائد اسمر عبد الله

قسم الاحصاء / كلية الادارة والاقتصاد/ جامعة كربلاء

بحث مستل من رسالة ماجستير في الاحصاء

المستخلص

في هذا البحث تم تسليط الضوء على احدى مشاكل تحليل الانحدار و التي تعتبر واحدة من الافتراضات الاساسية التي تقوم عليها تم تسلسطريقة المربعات الصغرى الاعتيادية (OLS) ، كذلك تعتبر واحدة من الشروط الاساسية لتحليل التباين. وهي مشكلة عدم تجانس التباين (Heteroscedasticity) ، ففي الكثير من التطبيقات العملية لا يمكن تحقيق هذا الشرط وبالتالي تكون طريقة (OLS) لتقدير معالم نموذج الانحدار الخطي غير مجدية في اعطاء نتائج صحيحة ودقيقة وهذا بسبب وجود هذه المشكلة .

ويهدف البحث الى اختيار اوتحديد افضل اختبار للكشف عن وجود هذه المشكلة ضمن مجموعة من الاختبارات المعلمية ، حيث تم كتابة برنامج بلغة (R) لاجل عمل مقارنة يكون فيها المعيار هو نسبة الكشف الصحيح عن وجود او عدم وجود مشكلة عدم تجانس التباين من خلال دراسة محاكاة بالاعتماد على بيانات افتراضية لنموذجين الاول يحتوي على مشكلة عدم تجانس التباين والاخر لايحتوي على المشكلة ، حيث تم اعتماد خمس اختبارات هي (كولد فيلد-كوانت ، بروش- بيجين – جود فري ، وايت، NCV ، هاريسون مكابي) .

كما تعتمد الرسالة جانبا تطبيقيا تم الأعتماد به على البيانات الحقيقية والتي تمثل الاطفال المصابين بمرض اليرقان، واستعمال الاختبار الافضل اللذي تم تحديده للكشف عن وجود المشكلة ثم تقدير نموذج الانحدار استنادا الى وجود المشكلة من عدمها.

Abstract

This research is considered on attempt to shed light on one of the problems of regression, which is one of the basic assumptions underlying ordinary least squares method (OLS). It is considered one of the fundamental conditions of variance analysis, and in many practical applications this hypothesis cannot achieve the (OLS) to estimate the model of linear regression which make it is less (not benefited) to give true and accurate, results and this leads to which make it is the problem of (the heteroscedasticity).

The aim of this thesis is to choose the best parametric test in order to detect this problem from a group of parametric tests. The researchers wrote an program in (R) language for a comparison in the best of the ratio of the detection true for the presence or absence of a problem through simulation within chapter III has a virtual reality two of the models, the first has a problem of heteroscedasticity the other not contain the problem depending on five tests (Gold field Quant, Prush bejen, White, NCV, Harrison-McCabe).

Chapter four contains the Application of the study and depending on real data, in this chapter the researchers choose the best test to detect the problem or not, then we estimate the regression model.

1-المقدمة

يعتبر تحليل الانحدار من اهم المواضيع الاحصائية ذات التطبيقات الواسعة اذ يقوم على عدة افتراضات اساسية وفي حالة عدم توفر احد هذه الافتراضات فأن ذلك يؤدي الى نتائج غير صحيحة ومن اهم هذه الافتراضات هو ثبات حد التباين اذ ان عدم تحقق هذا الشرط يسبب مشكلة تدعى بمشكلة عدم تجانس التباين (عدم ثبات تباين حد الخطأ) والتي قد تأتي من اختلاف العينات والتي تتبع مجتمعات ذات تباينات مختلفة ، وهذا بدوره يؤدي الى قرارات غير دقيقة او مضللة ففي حالة اختبار الفرضيات اذ يرتفع مستوى المعنوية بشكل تلقائي لذلك يجب ان تكون الاختلافات العشوائية ضمن المعاملات على درجة كبيرة من التجانس وهذا بدوره يجعل الاختلافات العشوائية متساوية بالنسبة للعينات المختلفة مما يجعل معه الحصول على تباين مشترك للخطأ العشوائي لجميع العينات ممكنا، وتعتبر مشكلة عدم تجانس التباين من مشاكل القياس الاقتصادي التي تواجه الباحثين عند وصفهم وبنائهم لنموذج انحدار خطي لمجموعة من الظواهر وعلى هذا الاساس اصبح لزاماً على الباحث ان يختبر ويفحص البيانات والمعلومات لاجل معرفة وتوضيح العلاقة بين العوامل المختلفة وكذلك تحليل النتائج بشكل ادق واشمل . وبذلك تتجلى اهمية لجوء الباحث الى استعمال الطرق الاحصائية التي تمهد له الوصف الموضوعي الدقيق لاجراء تجربته على اسس خالية من الاجتهادات والعوامل الشخصية التي بدورها قد يكون لها تأثيراً مباشراً على نتائج البحث ، ولاجل الحيلولة دون حصول هذه الامور فقد تم اقتراح عدد من الاختبارات تستعمل للكشف عن وجود هذه المشكلة او عدم وجودها . حيث سيتم التركيز في هذه الرسالة على المقارنة بين مجموعة من الاختبارات المعلمية (Parametric Tests). ولتوضيح منهجية هذه الرسالة فقد تم تقسيمها الى اربعة فصول، حيث يشمل الفصل الاول المقدمة وهدف ومشكلة البحث وكذلك الاستعراض المرجعي ، اما الفصل الثاني فقد تضمن الجانب النظري والذي تم عرض فيه توضيحا لموضوع الانحدار وايضا الاختبارات الخاصة بتجانس التباينات المعلمية في هذا المجال ، اما الفصل الثالث فقد تضمن مفهوم المحاكاة وفائدة استعمال هذا الأسلوب واساليب توليد البيانات واجراء تجربة محاكاة (لغة Rللبرمجة الاحصائية) وتحليل النتائج الخاصة بكل اختبار والمقارنة بين الاختبارات على اساس معيار نسبة نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين (Non-homogeneity detection ratio) ، وتضمن الفصل الرابع التطبيق العملي للاختبار الافضل (اللذي يمتلك اعلى نسبة كشف) على بيانات حقيقية ، ليتم بعد ذلك عرض الاستنتاجات التي تم التوصل اليها من خلال تجربة المحاكاة والجانب التطبيقي وبعد ذلك ذكر التوصيات على ضوء ما تم عرضه من استنتاجات .

2- اهمية البحث

عندما نعتمد على انموذج انحدار خطي عام لتمثيل البيانات فأننا نستعمل طريقة المربعات الصغرى الاعتيادية (OLS) لتقدير معالم هذا النموذج وتعتمد هذه الطريقة عند تطبيقها على تحقق فرضيات وشروط اساسية معينة حتى تكون هنالك دقة في تقدير معالم انموذج الانحدار، وواحدة من الفرضيات الاساسية التي نعتمدها عند عملية التقدير بطريقة (OLS) هي فرضية ثبات تجانس التباين (Homoscedasticity) لحدود الاخطاء والتي يعبر عنها رياضيا بالصيغة الاتية :

$$Var(\underline{U}) = E(UU') = \sigma^2 \ln \qquad (1)$$

وفي كثير من التطبيقات العملية لا يمكن تحقيق هذه الفرضية وبالتالي تكون طريقة (OLS) لتقدير معالم نموذج الانحدار الخطي غير مجدية في اعطاء نتائج صحيحة ودقيقة وهذا يؤدي الى الوقوع في مشكلة عدم تجانس التباين (Heteroscedasticity) والتي يعبر عنها رياضيا بالصيغة التالية:

$$Var(\underline{U}) = E(UU') = \sigma^2\Omega \qquad (2)$$

مضافا لذلك ان المعالم المقدرة بوجود مشكلة عدم التجانس سوف لا تعطينا خاصية افضل تقدير خطي غير متحيز (Blue) ومعنى هذا ان المعلمة المقدرة سوف لا تمتلك خاصية اقل تباين ممكن ومن هنا جاءت اهمية هذا البحث لاجل اختيار الاختبار الأحصائي الذي يكون ملائما للكشف عن المشكلة في حالة عدم التحقق لاحد الشروط الأساسية لأجراء الأختبارات الاحصائية المعلمية وهو شرط التجانس للتباين لاجل ذلك استعملت بعض الطرائق المعلمية وأجراء المقارنة للمفاضلة بين الاختبارات للطرائق المعلمية في حالة انتهاك هذا الفرض .

3- هدف البحث

الهدف من هذا البحث هو اختيار افضل اختبارللكشف عن مشكلة عدم تجانس التباين من خلال المقارنة بين بعض الطرائق الأحصائية المعلمية في ظل انتهاك احدى فرضيات تحليل الانحدار المتعدد وهي ثبات التباين لحد الخطأ العشوائي ويتم ذلك بالاستعانة باسلوب المحاكاة (لغة R) يتبعه تطبيق عملي للبيانات .

4-اهمية البحث

العديد من الدراسات تناولت كيفية اختبار ومعالجة مشاكل الانحدار الناتجة عن خرق احد فروض الانحدار، وأحدى هذه المشاكل المهمة هي مشكلة عدم تجانس التباين وتأتي اهمية هذا البحث في اختيار نموذجين النموذج الاول يحتوي على مشكلة عدم تجانس التباين والنموذج الثاني لايحتوي على مشكلة عدم تجانس التباين ليتم بعد ذلك اختيار الاختبار الافضل واللذي يستعمل للكشف عن وجود او عدم وجود مشكلة عدم تجانس التباين

5-مفهوم الانحدار

يقال ان الانحدار هو دراسة الاعتمادية وهذا يعني اعتماد متغير على متغير اخر كتأثير تغيير حجم الصف على نجاح الطلاب ، ويعتبر الانحدار من المواضيع الاساسية ويعد جزءا مهم من النظرية الاحصائية ، ويصنف الانحدار ضمن المواضيع المتميزة بسبب استخداماته الواسعة في مختلف العلوم الادارية منها والاقتصادية وكذلك الطبيعية ، فعلا سبيل المثال لا حصرا ، في المجال الاقتصادي ، يعتبر الاداة العلمية التحليلية في القياس الاقتصادي ، والاقتصاد الكلي التحليلي ، حيث من الممكمن استخدامه للتعبير عن العلاقات المستعملة في ربط المتغيرات الاقتصادية فيما بينها ، وذلك على شكل نماذج رياضية تسمى (نماذج الانحدار) ، حيث تم تقديم الانحدار عام (1886) من قبل كالتون (Galton) وذلك في مقالته التي التي تم فيها دراسة استقرارية توزيع الاطوال في المجتمع حيث استخدم عينة تزيد على الف عائلة. حيث اكدت النتائج بالرغم من وجود ميل لجميع الاباء طويلي القامة للحصول على اطفال طويلي القامة، بينما الاباء قصار القامة لهم ميل ان يحصلوا على اباء قصار القامة ، وان متوسط طول الاطفال الذين ولدوا لاباء من طول معين يكون تحركه (انحداره) باتجاه المتوسط لطول اطفال المجتمع ككل . وقد تم استعمال مصطلح الانحدار من قبل كالتون للاشارة الى اتجاه الاطوال لاطفال المجتمع ككل نحو المتوسط العام ، وقد توالت وتعددت استعمالات هذا النوع من التحليل لتشمل مختلف جوانب الحياة [1] [12] .

6-انواع تحليل الانحدار [6]

يبنى التقسيم او التفرقة بين انواع الانحدار على اساس عدد المتغيرات المستقلة فالانحدار اللذي يشتمل على متغير مستقل واحد ومتغير تابع واحد يطلق عليه الانحدار الخطي البسيط ويعتبر من ابسط انواع التحليل، ويسمى الانحدار اللذي يتضمن اكثر من متغيرين مستقلين بالانحدار المتعدد .

7-الانحدار الخطي المتعدد [7]

في الانحدار الخطي البسيط تقدر العلاقة بين اثنين من المتغيرات، اما في الانحدار المتعدد تكون المتغيرات مضاعفة، اي ان العلاقة ان وجدت فهي علاقة تتضمن اكثر من متغيرين وكذلك العلاقة ان وجدت فقد تكون غير خطية و يعتبر أنموذج الانحدار المتعدد الامتداد المنطقي والطبيعي للنموذج الخطي بمتغيرين ، فعندما يتم استعمال (K) من المتغيرات المستقلة ($X_1. X_2. \dots . X_k$) لاجل تفسير التباين للمتغير المعتمد (Y) في معادلة الانحدار. فأن تعدد المتغيرات التوضيحية تجعل من التعامل مع طرق الجبر الخطي اي (جبر المصفوفات) هي المستعملة لاجل تقدير واختبار ثم تحليل نماذج الانحدار المتعدد . ليتم بعد ذلك تعميمها ثم تطبيقها في حالة المتغيرين او ثلاث متغيرات او اي عدد ممكن من المتغيرات ولكن بشرط ان لايتجاوز عدد المتغيرات عدد المشاهدات المستعملة للتقدير.

8- مشكلة عدم تجانس التباين :

من الافتراضات المهمة هو مايحدده الفرض الثالث لمقدر (OLS) اي ان يكون تباين حد الخطأ ثابت ويكون متساويا لجميع مشاهدات العينة بمعنى اخر:

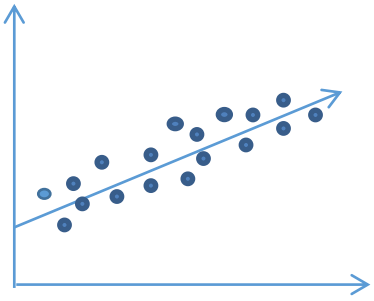
$$\mathrm{Var}\,(U_i) = E[u_i - E(u_i)]^2 = \sigma_u^2 \qquad (3)\ E(u_i)^2$$

وهذا الفرض يسمى بفرض تجانس تباين الخطأ (Homoscedasticity).وتشير كلمة (homo) الى معنى التساوي وكلمة (scedasticity) الى معنى الانتشار ومعنى هذا الفرض ان احتمالية توزيع حدود الخطأ العشوائي (u_i) تبقى نفسها حول جميع مشاهدات قيم (X_i) اي ان تباين حد الخطأ (u_i) هو نفسه لكل قيم المتغير المستقل (التوضيحي) . وعندما لا تتحقق هذه الفرضية في اي حالة معينة فأن هذا يؤدي الى مخالفة فرض التجانس اي ان تباين العينة اصبح يختلف باختلاف مشاهدات العينة وهذا ما يدعى بعدم ثبات تجانس التباين (heteroscedasticity) . كذلك اذا كانت فرضية التجانس للتباين غير متحققة ، فأن ذلك يؤدي الى ان تكون مصفوفة التباين - التباين المشترك للأخطاء معرفة كالتالي[2] [3] [4] [8] :

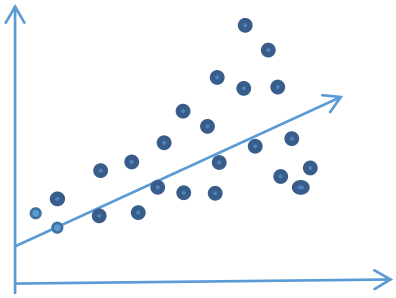
$$\Omega_\varepsilon = E(\varepsilon\varepsilon') = \begin{pmatrix} \sigma_{\varepsilon.1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\varepsilon.2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\varepsilon.n}^2 \end{pmatrix} \neq \sigma_{£}^2 I_n \qquad (4)$$

نلاحظ ان تباينات الاخطاء غير ثابتة على القطر الرئيسي ويؤدي هذا الى ان يكون تباين الاخطاء مرتبط بقيم المتغير المستقل . ويمكن توضيح شكل العلاقة المتوقعة بين المتغير التابع والمتغير التوضيحي في حالة ثبات التباين وفي حالة عدم ثبات التباين كالاتي :

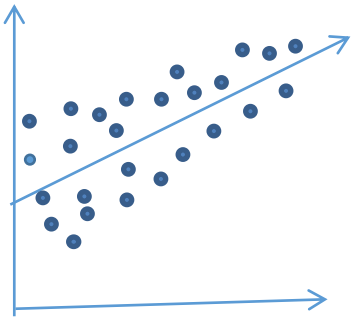
شكل (1) يوضح حالات ثبات وعدم ثبات التباين



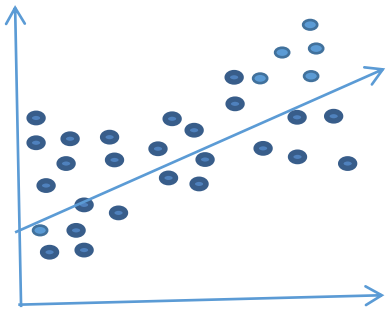
شكل (1-A) يوضح حالة ثبات التباين



شكل (1-B) يوضح حالة عدم ثبات التباين



شكل (1-C) يوضح حالة عدم ثبات التباين مع زيادة في المتغير التوضيحي(x_i)



شكل (1-D) يوضح حالة عدم ثبات التباين مع نقصان في المتغير التوضيحي (x_i)

المصدر : [9]

9- اسباب مخالفة فرض التباين

اولا : نماذج خطأ التعلم (Learning Error models) ونقصد به سلوكية وتصرف الافراد التي تقل الاخطاء فيها بمرور الزمن وبالتالي فأن تباين (σ_i^2) يتناقص ايضا خلال المدة الزمنية مثال ذلك عدد الاخطاء على ماكنة معينة فأنها تقل بزيادة الخبرة لدى العامل خلال الزمن [4] [2] [11] .

ثانيا : يزداد التباين لحد الخطأ (σ_i^2) وذلك بزيادة مستوى الدخل لدى الافراد وذلك عائد الى اختلاف وتعدد الناس في اختياراتهم مثال ذلك تباين الانفاق على المواد الغذائية يمكن ان يزيد بزيادة دخل الاسرة .

ثالثا : مع تحسن اساليب جمع البيانات يودي الى انخفاض تباين حد الخطأ (σ_i^2)، لان جمع البيانات الدقيقة والتي تكون قريبة من الواقع تقلل الاخطاء مثال ذلك حساب واردات البلد بالاعتماد على الحاسب الالي تكون ادق من التي يتم حسابها بطريقة يدوية.

رابعا : تحدث المشكلة احيانا بسبب وجود القيم (الشاذة)، فالقيمة الشاذة تعتبر مشاهدة مختلفة بشكل ممكن يمكن ملاحظته من بين بقية المشاهدات (سواء كانت كبيرة جدا او كانت صغيرة جدا)، بعبارة ادق المشاهدة الشاذة هي مشاهدة مأخوذة من مجتمع يختلف عن المجتمع الذي تم توليد منه بقية المشاهدات، وان وجود او حذف مثل هذا النوع من البيانات لاسيما عنما يكون حجم العينة صغيرا من الممكن ان يكون هذا مؤثرا على نتائج تحليل الانحدار .

خامسا : قد يكون سبب الاختلاف هو التوصيف الخاطئ لنموذج الانحدار، فالعديد من اختلاف التباين يكون بسبب حذف متغيرات مهمة من النموذج، وبالتالي فأن البواقي التي يتم الخصول عليها من نموذج الانحدار قد توحي ان تباين الاخطاء غير ثابت .

سادسا : من مصادر اختلاف التباين قد يكون التواء توزيع متغير واحد او اكثر من المتغيرات المنحدرة والتي توجد ضمن النموذج .

سابعا : من الاسباب الاخرى لاختلاف التباين هو التحويل الخاطئ للبيانات او بسبب عدم صياغة الدالة بشكل سليم.

10- الاختبارات المستعملة في الكشف عن المشكلة:

تعد الاختبارات الاحصائية المعلمية واحدة من اهم التطبيقات التي قدمها علم الاحصاء،حيث على معلمات المجتمع اعتماداً على ما توافر من بيانات لدى الباحث خاصة بالعينة المأخوذة من هذا المجتمع وتستعمل الأختبارات المعلمية في حالة العينات الكبيرة التي يشترط فيها توافر المعلومات من مجتمعاتها (معلومات الاصل) مثل ان يكون توزيع البيانات توزيعاً طبيعياً ، والتباينات تكون متجانسة والعينات مأخوذة بصورة عشوائية والعلاقة تكون خطية واستقلال العينات غيرها ويستعمل مع البيانات التي تتوفر عنها معلومات بشكل كمي قابل للقياس .

11- اختبار اختبار بروش- بيجين – جود فري [5] [4] .

يعتبر هذا الاختبارمن الاختبارات المهمة المستعملة للكشف عن وجود مشكلة عدم تجانس تباين الخطأ في الانموذج الخطي المتعدد حيث تم اقتراح هذا الاختبار عام (1979) :

$$Y=\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_m X_m + u \qquad (5)$$

حيث يفترض ان :

$$\sigma_i^2 = f(\alpha_0 + \alpha_1 z + \cdots + \alpha_k z_k) \qquad (6)$$

اذ ان (z) تمثل بعض او كل المتغيرات التوضيحية (X_i).

وفي حال ان : $\alpha_1 = \alpha_2 = \cdots = \sigma_k = 0$ فإن ($\sigma_i^2 = \alpha_0$) وهذا معناه ان التباين ثابت (constant) اما خطوات الاختبار فهي كالتالي :

أ- تم اجراء انحدار(Y) على كل (X'_S) للحصول على الاخطاء :($e_1 . e_2 . \ldots . e_n$) .

ب- يتم تقدير ($\hat{\sigma}_u^2 = \frac{\sum e_i^2}{n}$) والتي تمثل تقدير للإمكان الاعظم لتباين الخطأ.

ت- يتم تعريف : ($P_i = \frac{e_i^2}{\hat{\sigma}^2}$) .

ث- يتم تقدير P_i على (Z'_S) :

$$P_i = \alpha_0 + \alpha_1 z_1 + \alpha_2 z_2 + \cdots + \alpha_m z_m + v_i \qquad (7)$$

ج- الحصول على ESS (مجموع المربعات المشروحة) وذلك مع فرض تحقق التوزيع الطبيعي للمتغير العشوائي (u) .

ح- يتم تعريف الاحصاءة (BPG) كالتالي : $BPG = \frac{1}{2} ESS$ وان الاحصاءة (BPG) تتوزع حسب توزيع مربع كاي (X^2_{n-1} ~ BPG) .في حالة $BPG > x^2_c$ فهذا يؤدي الى رفض فرضية العدم (H_0) وهذا بدوره يشير الى وجود مشكلة عدم تجانس التباين، علما ان هذا الاختبار يشترط التوزيع الطبيعي للبيانات.

12-اختبار وايت [4]:

يعد من الاختبارات المهمة للكشف عن وجود مشكلة عدم تجانس التباين من عدمها ، حيث تم اقتراح هذا الاختبار عام (1980) وهو اختبار عام بينما كان اختبار كولد فيلد – كوانت) يحتاج الى معرفة اي متغير توضيحي هو الذي يسبب المشكلة ، كذلك نجد اختبار(BPG) يفترض ان يكون التوزيع للمتغير العشوائي توزيعا طبيعيا .

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u \qquad (8)$$

ويمكن تلخيص خطوات هذا الاختبار كالتالي :

أ- يتم تقدير الانموذج (معادلة 5) للحصول على البواقي (e_i) .

ب- اجراء انحدار(e^2) على $X_1 X_2 . X_2^2 . X_1^2 . X_2 . X_1$ والثابت.

$$e_i^2 = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_1 X_2 + \alpha_4 X_1^2 + \alpha_5 X_2^2 + v_i \qquad (9)$$

ت- حساب قيمة (R^2) من معادلة الانحدار ، بعد ذلك نجد (nR^2) حيث تمثل (nR^2) احصاءة الاختبار اذ تتوزع حسب توزيع مربع كاي : $nR^2 \sim X_5^2$ وبشكل عام فأن $nR^2 \sim X^2 \frac{k(k+3)}{2}$ ، حيث ان (k) تمثل عدد المتغيرات التوضيحية في النظام .

ث- صناعة القرار : اذا كانت $nR^2 > X_c^2$ عند ذلك يتم رفض فرضية العدم اي توجد مشكلة عدم تجانس التباين ، اما فرضيات هذا الاختبار هي :

$$H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_5 = 0$$

$$H_1: \alpha_1 \neq \alpha_2 \neq \cdots \neq \alpha_5 \neq 0$$

13- اختبار جولد فيلد- كواندت

يحتسب تباين الخطأ للعينة الجزئية الاولى (S_i^2 1) ، وكذلك للعينة الجزئية الثانية (S_i^2 2) وحسب الصيغ التالية يعتبر من الاختبارات المهمة والتي تستعمل للكشف عن مشكلة عدم التجانس لتباين الخطأ ويستعمل في حالة العينات الكبيرة الحجم والخطوات هي كالاتي[3] :

أ- ترتب البيانات الخاصة بالمتغير التوضيحي (X_i) ترتيبا تصاعديا .

ب- تحذف المشاهدات الوسطية من البيانات للعينة ، ويفضل عادة حذف ($\frac{1}{5}$) المشاهدات .

ت- يتم قسمة المشاهدات الباقية بشكل عينتين جزئيتين متساويتين تضم الاولى على قيم (X_i) الاولى (الصغيرة) مع قيم المتغير المعتمد (Y_i) التي تقابلها اما العينة الاخرى فتظم قيم (X_i) التالية (الاكبر) مع قيم المتغير المعتمد (Y_i) التالية .

ث- يتم تقدير معلمات العلاقة الخطية ($\hat{b}_2 . \hat{b}_1$) بين المتغير المستقل والمتغير التوضيحي لكل عينة جزئية بشكل منفرد .

$$S_i^2\ 1 = \frac{\sum e_i^2}{T_1 - 2} \qquad (10)$$

$$S_i^2\ 2 = \frac{\sum e_i^2}{T_2 - 2} \qquad (11)$$

ج- تحسب إحصاءة (F^*) اعتمادا على الصيغة التالية :

$$F = \frac{S_i^2 2}{S_i^2 1} \qquad (12)$$

بعد ايجاد (F^*) المحتسبة نقارنها مع (F) الجدولية عند مستوى معين من المعنوية وبدرجة حرية مقدارها ($T_2 - 2$) للبسط و($T_1 - 2$) للمقام فاذا كانت القيمة المحتسبة اصغر من القيمة الجدولية نأخذ بفرضية العدم والتي تؤكد على عدم وجود المشكلة حيث $H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$ ، اما في حالة كون قيمة (F^*) المحتسبة اكبر من قيمة (F) الجدولية عند نفس المستوى من المعنوية ودرجة الحرية فعند ذلك نأخذ الفرضية البديلة : $H_1 \neq \sigma_1^2 \neq \sigma_2^2 \neq \cdots \neq \sigma_n^2$ القائلة بوجود المشكلة اي عدم تجانس التباين للخطأ .

14-اختبار هارسون مكابي [10]:

ان الاثار المترتبة من مشكلة عدم تجانس التباين على تقدير المربعات الصغرى الاعتيادية لنماذج الانحدار الخطي تكون معروفة ، ومن المهم ان تكون هنالك مقدرة على كشف مشكلة عدم تجانس التباين على نطاق واسع ، وخلال السنوات الاخير تم اجراء عدة اختبارات متقدمة في هذا المجال ،ان معظم الدراسات التجريبية تستند على الاوقات الزمنية، وقد اشار كل من هاريسون ومكابي عام (1975) وكذلك ايبس (1977) ان صفة عدم التجانس قد تكون في كثير من الاحيان سمة في السلاسل الزمنية وكذلك في نماذج المقاطع العرضية، حيث تستعمل لهذا الاختبار احصاءة

معينة والتي هي عبارة عن نسبة معينة لمجاميع بواقي الانحدار، وهذه النسبة هي (مجموع المربعات من مجموعة فرعية اقل الى اجمالي المجموع المتبقي من المربعات وصيغة هذه النسبة

$$b = \frac{e'Ae}{e'e} \qquad (13)$$

حيث تقارن قيمة (**Sig.**) لاحصاءة اختبار هارسون مع مستوى المعنوية (0.05) فأذا كانت فأذا كانت قيمة (**Sig.**) اقل فهذا دليل على وجود مشكلة عدم تجانس التباين ، اما اذا كانت هذه القيمة اكبر من (0.05) فهذا يدل على عدم وجود مشكلة عدم تجانس التباين .

15-اختبار (NCV) [10] :

يعد اختبار (NCV) من الاختبارات المهمة والمستعملة في الكشف عن مشكلة عدم تجانس التباين وغالبا مايسمى هذا الاختبار بأختبار(Breusch-Pagan) اذ تم اقتراح هذا الاختبار مع بعض التعديلات على اختبار (Breusch-Pagan) عام 1983 من قبل (Weisberg & Cook)، حيث تتم المقارنة بالاعتماد على قيمة (Sig.) فاذا كانت قيمة احصاءة الاختبار اقل من (0.05) فهذا يدعونا الى رفض فرضية العدم والتي تفترض ان التباينات ثابتة اي وجود مشكلة عدم تجانس التباين.

16- تجارب المحاكاة:

تم استعمال برنامج في تنفيذ تجارب المحاكاة وهذا البرنامج يستند على لغة(R) وهي لغة متطورة عن لغة (S-Pluss) ومن مميزات هذا البرنامج هو الامكانية العالية في العلوم الإحصائية والهندسية والرياضية حيث يعمل على توظيف الايعازات والادوات بصيغة مباشرة لاجل الحصول على برمجة متقدمة . تعتبر لغة R من بين اللغات التي تألق نجمها حديثا وبشكل سريع في مجال البرمجيات العلمية في مجالي الإحصاء والمعلوماتية حيث اصبحت لغة معتمدة على نطاق كبير في كثير من مراكز البحث العلمية والجامعات، وصار استعمال هذه اللغة والاشارة إليها في المجلات العلمية المحكّمة والبحوث المنشورة يزداد بشكل طردي بوتيرة متسارعة وبشكل طردي ، مضافا الى هذا تعد لغة البرمجة(R) لغة حرة ذات مصدر مفتوح و توزيعها يخضع لترخيص GPL االمشهور. كل هذا ساعد في تزايد ما هو متوفر وموجود على شبكة (الإنترنت) من مراجع لها ، فهناك ماهو موجود على شكل الكتب الإلكترونية والدروس التعليمية بل حتى المناهج الأكاديمية وكذلك الدورات التدريبية مضافا لذلك البرمجيات الجاهزة , والتي تكون مكتوبة بلغة (R) لاجل تنفيذ المهمة هذه أو تلك، حتى اصبحت هذه اللغة تحظى ببعض الامتياز عند مقارنتها مع العديد من العمالقة في قطاع البرامج الإحصائية والرياضية العلمية مثل (SASو SPSS) ولاسيما في مجال توفر الطرق الجديدة والخوارزميات الحديثة ، حيث يلاحظ هذا توفر الطرق الجديدة والخوارزميات الحديثة .

17 - الحصول على المعلمات:

أ : المعلمات(b_i) :

تم تقدير المعلمات $(b_1; b_2; b_3; b_4)$ للبيانات الحقيقية وذلك بالاعتماد على البرنامج الاحصائي (SPSS) ثم اخذت قيم مفترضة اعلى وقيم مفترضة ادنى والقيم الاصلية للمعلمات الحقيقية .

ب: المتوسطات(u_i) :

تم تقدير المتوسطات $(u_1; u_2; u_3; u_4)$ للبيانات الحقيقية وذلك بالاعتماد على البرنامج الاحصائي (SPSS) ثم اخذت قيم مفترضة اعلى وقيم مفترضة ادنى والقيم الاصلية للمتوسطات الحقيقية .

ت: التباينات (σ_i^2) :

تم تقدير التباينات $(\sigma_1^2; \sigma_2^2; \sigma_3^2; \sigma_4^2)$ للبيانات الحقيقية وذلك بالاعتماد على البرنامج الاحصائي (SPSS) ثم اخذت قيم مفترضة اعلى وقيم مفترضة ادنى والقيم الاصلية للتباينات الحقيقية .

علما ان المعيار اللذي سيتم المقارنة (المفاضلة) على اساسه من بين الاختبارات الخمسة هو [نسبة الكشف (Detection ratio) وهي عبارة عن قسمة عدد مرات الكشف الصحيح على عدد المحاولات مضروب * 100].

تجربة رقم (1): تم محاكاة نموذج انحدار بالمعلمات الافتراضية التالية
$B_i = 0.732, -0.112, 0.034, -0.123$ والجدول التالي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين حيث ان $u_1 = u_2 = u_3 = u_4 = 0$ ، $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$

الجدول (1)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على مشكلة عدم تجانس التباين

Size of sample / Tests	20	30	50	100	150	200
R.BP	2	3	13	3	4	12
R.NCV	20	21	34	25	28	39
R.white	38	44	36	27	41	51
R.GQ	87	99	100	100	100	100
R.HM	82	98	100	100	100	100

المصدر : الجدول (1) من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

18- تحليل نتائج محاكاة نموذج الانحدار

أ - من خلال النتائج اعلاه (الجدول رقم 1) يتبين ان نسبة الكشف لاختبار (كولد فيل-كوانت) بلغت 87% عند حجم العينة 20 و99% عند حجم العينة اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون حيث تساويا عند نسبة الكشف 100% عند احجام العينات الاخرى ، بينما كانت نسب كشف اختبار (بروش- بيجين – جود فري) 2% عند حجم العينة 20 و3% عند حجم العينة 30 و13% عند حجم العينة 50 و 3% عند حجم العينة 100 و4% عند حجم العينة 150 و12% عند حجم العينة 200.

تجربة رقم (2): تم محاكاة نموذج انحدار بالمعلمات الافتراضية التالية

$B_i = 0.512, -0.221, 0.011 - 0.381$ والجدول التالي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين حيث ان $u_1 = u_2 = u_3 = u_4 = 0$ ، $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$

الجدول (2)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين						
Size of sample / Tests	20	30	50	100	150	200
R.BP	10	3	5	2	10	1
R.NCV	23	23	24	23	39	26
R.white	62	40	33	41	40	13
R.GQ	82	99	100	100	100	100
R.HM	68	97	100	100	100	100

المصدر : الجدول (2) من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

ب- من خلال النتائج (الجدول رقم 2) يتبين ان نسبة الكشف لاختبار (كولد فيلد كوانت) بلغت 82% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون حيث تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 10% عند حجم العينة 20 و3% عند حجم العينة 30 و5% عند حجم العينة 50 و 2% عند حجم العين 100 و 10% عند حجم العينة حجم العينة 150 و 1% عند حجم العينة 200.

تجربة رقم (3): تم محاكاة نموذج انحدار بالمعلمات الافتراضية التالية

$B_i = 0.641, -0.144, 0.023. -0.268$ والجدول التالي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين حيث ان $u_1 = u_2 = u_3 = u_4 = 0$ ،

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = 1$$

الجدول (3)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في حالة احتواء النموذج على عدم تجانس التباين

Size of sample / Tests	20	30	50	100	150	200
R.BP	2	3	10	8	7	4
R.NCV	21	20	32	32	34	29
R.white	70	32	54	11	45	39
R.GQ	82	99	100	100	100	100
R.HM	72	98	100	100	100	100

المصدر : الجدول (3) من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

ت- يتبين من خلال النتائج اعلاه (الجدول رقم 3) ان نسبة الكشف لاختبار (كولد فيلد كوانت) قد بلغت 82% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون حيث تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 2% عند حجم العينة 20 و3% عند حجم العينة 30 و10% عند حجم العينة 50 و8% عند حجم العين 100 و 7% عند حجم العينة حجم العينة 150 و4% عند حجم العينة 200.

تجربة رقم (4): تم محاكاة نموذج انحدار بالمعلمات الافتراضية التالية :

$B_i = 0.641, -0.144, 0.023. -0.268$ والجدول التالي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين حيث ان:

$$,u_1.u_2.u_3.u_4 = 1.92.5.43.25.41.1.94$$

$$\sigma_1^2.\sigma_2^2.\sigma_3^2.\sigma_4^2 = 0.132 .11.234 . 18.123 . 0.94$$

الجدول (4)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين

Size of sample / Tests	20	30	50	100	150	200
R.BP	3	3	4	10	14	2
R.NCV	16	23	28	34	42	34
R.white	29	46	32	46	46	46
R.GQ	86	99	100	100	100	100
R.HM	75	97	100	100	100	100

المصدر : الجدول (4) من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

ج- من خلال النتائج اعلاه (الجدول رقم 4) يتبين ان نسبة الكشف لاختبار (كولد فيلد كوانت) بلغت 86% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون حيث تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 3% عند حجم العينة 20 و3% عند حجم العينة 30 و4% عند حجم العينة 50 و10% عند حجم العين 100 و 14% عند حجم العينة حجم العينة 150 و2% عند حجم العينة 200.

تجربة رقم (5): تم محاكاة نموذج انحدار بالمعلمات الافتراضية التالية

$B_i = 0.641, -0.144, 0.023. -0.268$ والجدول التالي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين حيث ان:

$$u_1 . u_2 . u_3 . u_4 = 3 \cdot 24 . 7 \cdot 29 . 27 \cdot 32 . 3 \cdot 46$$

$$\sigma_1^2 . \sigma_2^2 . \sigma_3^2 . \sigma_4^2 = 1 . 14.561 . 20.621 . 2.35$$

الجدول (5)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين						
Size of sample / Tests	20	30	50	100	150	200
R.BP	5	2	10	5	5	2
R.NCV	18	28	33	30	31	25
R.white	49	26	50	33	46	31
R.GQ	78	100	100	100	100	100
R.HM	60	99	100	100	100	100

المصدر : الجدول (5) من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

ج- من خلال النتائج اعلاه (الجدول رقم 5) يتضح ان نسبة الكشف لاختبار (كولد فيلد كوانت) قد بلغت 78% عند حجم العينة 20 و100% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون حيث تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 5% عند حجم العينة 20 و2% عند حجم العينة 30 و10% عند حجم العينة 50 و5% عند حجم العينة 100 و 5% عند حجم العينة حجم العينة 150 و2% عند حجم العينة 200 .

تجربة رقم (6): تم محاكاة نموذج انحدار بالمعلمات الافتراضية التالية

$B_i = 0.641, -0.144, 0.023. -0.268$ **والجدول التالي يبين نسبة الكشف الصحيح عن مشكلة عدم تجانس التباين حيث ان:**

$$u_1 . u_2 . u_3 . u_4 = 2.745 . 6.14 . 26.63 . 2.18$$

$$\sigma_1^2 . \sigma_2^2 . \sigma_3^2 . \sigma_4^2 = 0.195 . 13.758 . 19.104 . 1.462$$

الجدول (6)

نسبة الكشف عن وجود مشكلة عدم تجانس التباين في نموذج يحتوي على عدم تجانس التباين

Size of sample / Tests	20	30	50	100	150	200
R.BP	2	3	1	2	4	2
R.NCV	19	22	20	21	30	26
R.white	54	47	23	20	60	19
R.GQ	85	99	100	100	100	100
R.HM	70	98	100	100	100	100

المصدر : الجدول (6) من عمل الباحث واعتمادا على نتائج المحاكاة (لغة R)

ح- من خلال النتائج اعلاه (الجدول رقم 6) تبين ان نسبة الكشف لاختبار (كولد فيلد كوانت) قد بلغت 85% عند حجم العينة 20 و99% عند حجم العينة 30 اما بقية احجام العينة فقد كانت نسبة كشفه اكثر من جميع الاختبارات ماعدا اختبار هارسون حيث تساويا عند نسبة الكشف 100% لاحجام العينات الاخرى ، اما اختبار (بروش- بيجين – جود فري) فسجل 2% عند حجم العينة 20 و3% عند حجم العينة 30 و1% عند حجم العينة 50 و2% عند حجم العينة 100 و 4% عند حجم العينة حجم العينة 150 و2% عند حجم العينة 200.

19- الجانب التطبيقي:

في الاسلوب التجريبي تم محاكاة نموذج انحدار بمعلمات مختلفة حيث تم تطبيق عدة اختبارات على بيانات، وبعد ان تم التأكد من افضلية طريقة (كولد فيلد –كوانت) للكشف عن مشكلة عدم تجانس التباين ، سيتم تطبيق هذا الاختبار بشكل عملي على بيانات فعلية نظراً لأهمية الجانب الصحي وتأثيره على حياة الانسان .

أ- تهيئة البيانات:

تم الحصول على البيانات من مستشفى الرميثة العام وهي بيانات طبية اخذت كعينة من شعبة الاحصاء ومختبر التحليلات المرضية للمستشفى عن طريق عينة عشوائية تشمل الاشهر(كانون الثاني– شباط- اذار من عام 2018) لنسبة اليرقان في الدم والعوامل االمؤثرة على زيادة او نقصان هذه النسبة، وتم فرز البيانات الاصلية عن طريق استعمال برنامج (Excel) كل قسم على حده ،حيث تم اخذ عينة مكونة من (100) مشاهدة ، وتم افتراض المتغيرات وهي :

- المتغير المعتمد(Y) ويمثل نسبة الاصابة بمرض اليرقان والذي يرمز له دوليا بالرمز (P59).
- المتغير التوضيحي (X_1) ويمثل وزن الطفل(الخديج) بالكغم لعينة مكونة من (100) طفل خديج.
- المتغير التوضيحي (X_2) ويمثل عمر الطفل(الخديج) بالايام .
- المتغير التوضيحي(X_3) ويمثل عمر الام (بالسنين) .
- المتغير التوضيحي (X_4) ويمثل فترة رقود الطفل(الخديج) في المستشفى بالايام .

20- الكشف عن وجود او عدم وجود مشكلة عدم تجانس التباين:

بالاعتماد على برنامج لغة (R) تم حساب احصاءة الاختبار لاختبار (كولدفيلد-كوانت) اذ تم تطبيق هذا الاختبار على البيانات الحقيقية بعد ان كان هذا الاختبار هو الافضل في اسلوب المحاكاة في حالة احتواء النموذج على مشكلة عدم تجانس التباين .

اختبار (كولدفيلد-كوانت)

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \neq \cdots \neq \sigma_n^2$$

جدول رقم (7) يوضح احصاءة اختبار(Gold field-Quant)

Test	Statistic	Df	Df	ig.
GQ	0.59971	45	45	0.9551

على البيانات الحقيقية، مستوى المعنوية (0.05)

المصدر : الجدول من عمل الباحث واعتمادا على نتائج برنامج (لغة R)

من خلال النظر في قيم (.Sig) والخاصة لاختبار كولد فيلد- كوانت في الجدول (7) نجد ان قيمة (.Sig) والتي تساوي (0.9551) هي اكبر من مستوى المعنوية (0.05) .

21- الاستنتاجات والتوصيات

اولا : الاستنتاجات

أ- من خلال النتائج المعروضة في الجدول رقم (7) لاختبار كولد فيلد- كوانت نلاحظ ان قيمة (.Sig) والتي تساوي (0.9551) هي اكبر من مستوى المعنوية (0.05) ومعنى هذا ان البيانات لاتعاني من مشكلة عدم تجانس التباين اي الاخذ بالفرضية الصفرية للبحث ($H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$).

ب- تفوق اختبار (كولد فيلد –كوانت) على بقية الاختبارات في النموذج اللذي يحتوي على مشكلة عدم تجانس التباين وذلك لامتلاكه اكبر نسبة كشف صحيح.

ت- هنالك تقارب بين اختبار (كولد فيلد- كوانت) واختبار (بروش- بيجين – جود فري) في حالة احتواء النموذج على مشكلة عدم تجانس التباين .

ث- في حالة النموذج اللذي يحتوي على مشكلة عدم تجانس التباين نجد ان اختبار (كولد فيلد- كوانت) يتفوق على جميع الاختبارات الا انه يتساوى مع اختبار(Harrison-McCabe) عند احجام العينات (50،100،150،200).

ثانيا : التوصيات

أ- استعمال اسلوب لغة البرمجة (R) في المقارنة بين اختبارات معلمية (اختبارات معلمية اخرى).

ب- اعتماد اختبار(كولد فيلد-كوانت) في الكشف الصحيح كونه اعطى افضل نسبة كشف صحيح عن وجلغة البرمجة (R) ود مشكلة عدم تجانس التباين.

ت- يمكن اعتماد اختباري (Harrison-McCabe اوGold field-Quant) للعينات ذات الحجم اكبر من (50) في الكشف عن المشكلة كونهما متكافئان.

ث- البحث عن اختبارات معلمية او اختبارات معلمية حديثة .

المصادر

اولا: المصادر العربية

1- النعيمي ،محمد عبد العال.طعمة،حسن ياسين. (2015)، " الاحصاء التطبيقي" ، ،ط2،الاردن، عمان، دار وائل للنشر.
2- عبد، حميد عبيد. (2016)، "الاقتصاد القياسي"، الطبعة الاولى، مطبعة دار الكتب، العراق، كربلاء.
3- شيخي ، محمد ، (2011) ، " طرق الاقتصاد القياسي " ، الطبعة الاولى، دار الحامد ،عمان ، الاردن .
4- التميمي، زهرة حسن عباس. واخرون.(2014)، "تحليل الانحدار"، الطبعة الاولى، مديرية دار الكتب للطباعة والنشر، جامعة البصرة، العراق، البصرة.
5- بخيت، حسين علي. فتح الله، سحر. (2012)،" الاقتصاد القياسي "، مطبعة دار الكتب ، بغداد.

ثانيا : المصادر الاجنبية

6. Anderson, David R., Sweeney, Dennis J., Williams, Thomas A. (2008),"Statistics For business And Economics", Eleventh Edition, South-Western Cengage Learning, USA.
7. Leeckley, Robers M. (2010)," Applied Statistics in Business and Economics ", CRC Press Taylor & Francis Group, Boca Raton London.
8. Gujarati, Damodar N. (1995)," Basic econometrics", fourth edition, McGraw-Hill, Inc.
9. Salvatore, Dominick. And Reagie, Derrick (2002) "Statistics and Econometrics", Second Edition, London.
10. Harrison, M. J., McCabe, B. P. M. (1979)," A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals.
11. Doane, David P., Seward Lori E. (2007)," Applied statistics in business and economics", Third Edition, McGraw-Hill, Inc, Americas, New York.
12. Weisberg, Sanford. (2005)," Applied Linear Regression" Third Edition, USA, A John Wiley & Sons, Inc., Publication, in Canada.