

تقدير النموذج الانحدار باستخدام طريقة المربعات الصغرى الموزونة الحصينة في حالة وجود مشكلة عدم تجانس التباين والقيم الشاذة مع تطبيق عملي

أ.د. سلمى ثابت ذاك *

علي قاسم محمد بندر **

المستخلص:

تعد مشكلة عدم تجانس التباين في النموذج الانحدار أحد أبرز المشاكل التي يمكن مواجهتها عند بناء مثل هذا النموذج ولا سيما في الظواهر الاقتصادية، وتعد طريقة المربعات الصغرى الموزونة (WLS) والتي تعد حالة خاصة عن طريقة المربعات الصغرى العمومية (GLS) وهي أحد أهم الوسائل البديلة لطريقة (OLS) في حالة وجود هذه المشكلة، إلا أن طريقة (WLS) تتطلب تقدير الوزن (W_i) الذي يتطلب هو الآخر معرفة هيكلية طبيعة عدم تجانس التباين الذي يكون في الغالب غير معروف، ولقد برزت في سنة 2001 طريقة موضوعية من قبل عدد من الباحثين والتي يرمز لها بـ (MPV) يمكن أن تضع مقدر الوزن ($\hat{W}_i$) بشكل يجعل تقديرات المربعات الصغرى الموزونة تتمتع بخصائص جيدة مثل عدم التحيز والاتساق والكفاءة تقاربياً، إلا أن طريقة (MPV) وغيرها من الطرائق البديلة لـ (OLS) لمعالجة مشكلة عدم تجانس التباين تخفق وبشكل كبير عند وجود القيم الشاذة إلى جانب وجود مشكلة عدم تجانس التباين الأمر الذي يجعل طريقة (MPV) متحيزة وليست ذات جدوى، وعلى هذا الأساس برزت منذ سنة 2008 طرائق مربعات صغرى موزونة (حصينة)، تضع تقديرات لا تتأثر بالقيم الشاذة وتبقى محتفظة بالخصائص المرغوب فيها، ولقد تناول هذا البحث طريقة (MPV) الحصينة (RMPV)، إذ أن طريقة (RMPV) لم تنال الاهتمام والدراسة من قبل الباحثين وعلى هذا الأساس فقد تم في هذا البحث تسليط الضوء على هذه الطريقة وعرض التفاصيل الخاصة بتطبيقها وذلك في ضوء بيانات واقعية لمجموعة بيانات (الدخل والإنفاق)، كما تم استعمال طرائق حصينة (غير تقليدية) وذلك للكشف عن القيم الشاذة وكذلك طرائق حصينة للكشف عن مشكلة عدم تجانس التباين في حالة وجود القيم الشاذة، ولقد تم تنفيذ كل هذه التطبيقات الخاصة بهذا البحث باستعمال برامج جاهزة مثل (SAS, Gretel, Statistic) ولقد تم التوصل إلى أن الطريقة الحصينة (RMPV) هي متفاربة جداً في نتائجها.

Abstract:

The non-Contrasted Homogeneity in an inclination model is considered as one of the supreme problems that one may confront while building such model, especially in economic phenomenon. Thus a method called (WLS)(Weighted Least Squares) which considered as a especial case from (GLS)(General Least Squares), one of the important Alternative means for (OLS), in case there is a problem. But the method (WLS) requires weight estimate also requires knowing the natural structure of non –Contrasted Homogeneity, which at most is un known. In (2001) a method made by a number of researchers, known as (MPV) This method can use weight estimate in a way made the (WLS) carry new good characteristics such as ; non-alignment and efficiency. Except that the (MPV) method and other Alternative methods for (OLS) failed enormously,

* الجامعة المستنصرية / كلية الادارة والاقتصاد .

** باحث .

مقبول للنشر بتاريخ 2017/10/30

مستل من رسالة ماجستير

to handle the problem of outliers, beside the problem of non-Contrasted Homogeneity. That made (MPV) method partial and useless. For that reason, in (2008) (robustness) (WLS) methods emerged and gave estimates that is not effected by outliers, and keep their wanted characteristics. This research took method (MPV) robustness (RMPV). The method (RMPV) did not interest Researchers who got all the attention, but both methods were highlighted and presented all the details for applying them on realistic data for a group of data of (income, expenditure). Also an unconventional robustness means were used to find out the outliers and non-Contrasted Homogeneity problem in case of an outlier. All these applications were executed by a ready-made programs like (statistics, Gretel , sas). And the they came to a conclusion that the results of (RMPV) are approximate. The (RMPV) was described as a complicated compared to the first method (RMPV).

الفصل الأول

1-1 المقدمة :

كما هو معروف فان العديد من الدراسات والبحوث تناولت الطرائق البديلة لـ (OLS) عند وجود مشكلة عدم تجانس التباين في الانموذج الانحدار وكانت ابرز هذه البدائل هي طريقة المربعات الصغرى الموزونة (WLS) الا انها تتطلب تقدير الوزن w_i الذي يتطلب بدوره معرفة طبيعة عدم تجانس التباين والذي في الغالب يكون غير معروف، وعليه فقد وضع مجموعة من الباحثين (Montgomery, peck and Vining) طريقة مربعات صغرى موزونة اطلقوا عليها (MPV) سنة 2001 تقوم على تقدير الوزن (w_i) في حالة عدم معرفة طبيعة عدم تجانس التباين، والى جانب هذه الطريقة هنالك أيضاً طريقة (HCCM) التي ظهرت عام 1980 من قبل الباحث (White) والتي تقوم على تقدير مصفوفة التباين والتباين المشترك الحصينة لمقدرات (OLS)، الا ان هذه البدائل جميعها تخفق وبشكل كبير عند وجود القيم الشاذة وعلى هذا الأساس كانت الحاجة الى وضع وتطوير طرائق مربعات صغرى موزونة حصينة لتقدير المعلمات عند وجود مشكلة عدم تجانس التباين والقيم الشاذة معاً. ولقد تناول هذا البحث طريقة (MPV) الحصينة والتي تم وضعها من قبل مجموعة الباحثين (IMON, MIDI and RANA) سنة (2009) الا ان طريقة (RMPV)، لم تنال الدراسة والاهتمام الكافي وعلى هذا الأساس فقد حاولنا في هذا البحث تسليط الضوء على طريقة (RMPV) وبيان الخطوات التفصيلية لمرحلة تطبيقها وذلك باستعمال بيانات واقعية لمجموعة (الدخل والنفقات) يتم فيها اختبار وجود القيم الشاذة أولاً، ومشكلة عدم تجانس التباين ثانياً وذلك وفق اختبارات وطرائق حصينة (غير تقليدية) للكشف عن القيم الشاذة، ومشكلة عدم تجانس التباين ومن ثم تطبيق طريقة (RMPV) في البيانات التي تتضمن وجود المشكلتين معاً.

1-2 مشكلة البحث:

تتمثل مشكلة البحث في كيفية تقدير معلمات انموذج الانحدار بطريقة المربعات الصغرى الموزونة وذلك في حالة وجود مشكلة عدم تجانس التباين ووجود القيم الشاذة في آن واحد، وبالشكل الذي يجعل تقديرات المربعات الصغرى الموزونة (الحصينة) لا تتأثر بوجود القيم الشاذة، وتبقى محتفظة بخصائصها المرغوب فيها كونها غير متحيزة وكفوءة تقاربياً ومتسقة وتبقى طريقة معتمدة لوضع المقدرات كوسيلة معالجة لمشكلة عدم تجانس التباين في انموذج الانحدار عند وجود القيم الشاذة.

1-3 هدف البحث:

يهدف البحث بالدرجة الأساس تسليط الضوء على طريقة تقدير المربعات الصغرى الموزونة (الحصينة) لمعلمات انموذج الانحدار وذلك في حالة وجود مشكلة عدم تجانس التباين والقيم الشاذة في آن واحد، والمتمثلة بطريقة (MPV) الحصينة وذلك من خلال بيان التفاصيل الخاصة بمراحل تطبيق هذه الطريقة.

الفصل الثاني الجانب النظري

2-1 المقدمة :

يتضمن هذا الفصل الجانب النظري لمحورين أساسيين، الأول ويتمثل بوجود القيم الشاذة في بيانات نموذج الانحدار، وكل ما يتعلق بالمفاهيم الأساسية للقيم الشاذة والطرائق التقليدية والحصينة لاكتشافها وكذلك الطرائق الحصينة البديلة للطرائق التقليدية في تقدير معاملات الانحدار، أما المحور الثاني فيتمثل بوجود مشكلة عدم التجانس في التباين في نموذج الانحدار حيث سيتم توضيح المفاهيم الأساسية لهذه المشكلة وآثارها وكيفية التعامل مع مشكلة عدم التجانس عندما تتواجد القيم الشاذة في آن واحد في الانحدار، وذلك فيما يتعلق بالطرائق البديلة الحصينة للكشف عن مشكلة عدم التجانس، والطرائق الحصينة لمعالجتها.

2-2 المحور الأول: القيم الشاذة في نموذج الانحدار: [1]-[4]

تعرف القيمة الشاذة على أنها تلك القيمة التي تبتعد عن نمط (Pattern) أو سلوك المشاهدات أو القيم الأخرى في الانحدار، أو بعبارة أخرى هي المشاهدات التي تكون غير متسقة مع بقية المشاهدات، وهناك تعريف آخر لها هي القيمة التي تبتعد عن مركز البيانات. إن مشكلة وجود القيم الشاذة في بيانات نموذج الانحدار يمكن تصنيفها بالشكل الآتي:

- 1- مشكلة وجود القيم الشاذة (Outlier) في المتغير المعتمد (y) أما يسمى بمتغير الاستجابة.
- 2- مشكلة وجود القيم الشاذة في المتغيرات التوضيحية (x-space) وعندئذ فإنها تسمى بـ (Leverage) points، وهناك نوعين الأول ويسمى (Bad Leverage)، وهي القيم الشاذة التي تبتعد عن نمط قيم المتغيرات التوضيحية (x-space) وغير متسقة معها فضلاً عن أنها تبتعد عن خط الانحدار المقدر، أي بعبارة أخرى إن القيم الشاذة في المتغيرات التوضيحية تبتعد عن مركز البيانات لهذه المتغيرات وبنفس الوقت تؤثر على تقدير متغير الاستجابة (y). والثاني ويسمى (good Leverage) فهي القيم التي تبتعد عن نمط القيم الخاصة بالمتغيرات التوضيحية ولكنها تبقى قريبة من خط الانحدار المقدر أو على الاتجاه الخطي أو بعبارة أخرى غير مؤثرة على تقدير متغير الاستجابة (y).
- 3- مشكلة تواجد القيم الشاذة (Outlier) في المتغير المعتمد (y) وكذلك وجودها في المتغيرات المستقلة (Leverage) معاً.

وفي ضوء إمكانية تواجد القيم الشاذة في نموذج الانحدار وفق أحد الاشكال الثلاثة المشار إليها ، فقد ظهرت العديد من طرائق الكشف عن وجود القيم الشاذة، وظهرت العديد من الطرائق الحصينة البديلة للطرائق التقليدية بغية وضع تقديرات حصينة لمعاملات الانحدار تكون بديلة عن الطرق التقليدية المتمثلة بـ (OLS) والتي وكما هو معروف تتأثر بشكل كبير عند وجود القيم الشاذة وذلك من حيث التقديرات الموضوعة للمعاملات فضلاً عن أن ابتعاد التوزيع الخطأ العشوائي عن التوزيع الطبيعي في حالة وجود هذه القيم يعمل على عدم إمكانية وضع الاستدلال الصحيح الأمر الذي استوجب معه ظهور ما يسمى بالانحدار الحصين وتنوعت أشكاله بما يتناسب ووجود هذه القيم الشاذة بالأشكال الثلاثة أعلاه، كما تنوعت طرق الكشف عنها بالشكل الذي يأخذ بنظر الاعتبار تواجدها وفق أحد الاشكال الثلاثة المشار إليها .

2-2-1 طرائق الكشف عن القيم الشاذة : [2],[10],[11],[5]

لقد ظهرت العديد من الطرائق للكشف عن القيم الشاذة في بيانات نموذج الانحدار ويمكن تلخيصها في الآتي:

- الطرائق التقليدية.
- الطرائق الحصينة.

والآتي عرض مفصل للطرائق الحصينة التي تم استخدامها للكشف عن القيم الشاذة في هذا البحث.

2-2-1-1 الطرائق الحصينة في الكشف عن القيم الشاذة: [8]

تتمثل هذه الطرائق بما يأتي: -

أ - البواقي الحصينة (Robust residuals).

ب - المسافة الحصينة لمهلا نوبس (RMD).

ج - الدمج بين البواقي الحصينة والمسافة الحصينة.

وتتصف هذه الطرائق بكونها غير حساسة لوجود القيم الشاذة وذلك لأنها تعتمد على طرائق تقدير حصينة سواء بالنسبة للبواقي أم المؤشرات الموقع والقياس الخاصة باحتساب المسافة، وبالتالي فإن ذلك يمكنها في الكشف عن القيم الشاذة بصورة يمكن الاعتماد عليها ولا تعمل على حجب، أو اغراق هذه القيم عند تشخيصها كما أنها كفوءة وخاصة عندما يكون هنالك بيانات بحجم كبير، وباتجاهات متعددة، وعليها اتجهت معظم الدراسات والبحوث في الآونة الأخيرة إلى اعتمادها، والآتي شرح مفصل لكل منها: -

أ- البواقي الحصينة (Robust residuals): -

وهي البواقي التي يتم الحصول عليها من الانحدار الحصين الذي يتم فيه تقدير المعلمات بطرائق حصينة لا تتأثر بوجود القيم الشاذة والتي تقود الى الحصول على بواقي حصينة، ومن هذه الطرائق هي طريقة (M- M Estimation)، وطريقة (LTS)، وطريقة (LMS)، وطريقة (S) وغيرها، ويتم اعتماد هذه البواقي لتشخيص القيم الشاذة في بيانات المتغير المعتمد (y) وذلك وفق الأسلوب الآتي: -
لتكن البواقي r_i اذ ان $i=1,2,...,n$ والمستحصل عليها من التقديرات الحصينة للانحدار فان متغير القيم الشاذة (Outlier) يمكن تعريفه بالشكل الآتي:-

$$\text{Outlier} = \begin{cases} 0 & \text{if } |e| \leq k_\sigma \\ 1 & \text{otherwise} \end{cases}$$

اذ ان e البواقي القياسية (standardized residuals)

ب- المسافة الحصينة لمهلا نوبس Robust Mahalinobes distant: [12]

كما هو معروف فان أحد الطرائق التقليدية للكشف عن (Leverage) هو استخدام مؤشر (Mahalinobes) والذي يرمز له (MD) والذي يعبر عن المسافة وابتعاد المشاهدة عن مركز البيانات والصيغة الخاصة به هي:

$$MD_i = \sqrt{(x_i - \hat{\mu})^T C^{-1} (x_i - \hat{\mu})} \quad (2-1)$$

اذ ان: C تمثل مصفوفة التباين المشترك التقليدية المحتسبة من العينة وان

$$X_i = (X_{1i}, \dots, X_{pi}) \quad , \quad i=1, \dots, n$$

وفي اتمودج الانحدار الخطي التقليدي فان عناصر القطر h_{ij} للمصفوفة H والتي تسمى (Hat) Matrix المبينة ادناه:

$$H = x(x^T x)^{-1} x^T$$

تستعمل للكشف عن النقاط او المشاهدات البعيدة عن مركز (Leverage)، ولقد أشار الباحثان (Rousseeuw and Van Driessen) سنة (1999) الى ان كلاً من (MDi), (hii) غير آمنة ولا يمكن الاعتماد عليها كلياً في معرفة نقاط (Leverage) في البيانات وذلك لأنه في العديد من الحالات قد تكون القيم بمسافات كبيرة الا انها لا تؤثر على انها (Leverage)، بسبب التأثيرات المخفية (masking) والتي لا تظهر والتي يكون السبب الأساس في هذا الامر هو كون (MDi) محتسب على أساس مؤشرات غير حصينة اذ ان تقديرات (μ) ومصفوفة التباين المشترك C هي غير حصينة، وعليه فقد اقترح كلاً من الباحثين سنة (1999) احتساب (MDi) حصينة وسميت بالمسافة الحصينة (Robust distant) ويرمز لها ب (RD) والتي تقوم على أساس تقديرات لمؤشرات الموقع (Location)، ومصفوفة تباين مشترك حصينة اسموها (scatter matrix) يمكن احتسابها من (MCD). وهي أصغر او أدنى محدود لمصفوفة التباين المشترك (minimum Covariance determinant) وقد وضع هذان الباحثين سنة (1999) خوارزمية سريعة لاحتسابها. والصيغة الخاصة باحتساب المسافة الحصينة هي:

$$RD_i = \sqrt{(x_i - T(X))^T C(x)^{-1} (x_i - T(X))} \quad (2-2)$$

اذ ان $T(x)$ ، $C(x)$ هي تقديرات (location)، ومصفوفة التباين المشترك (Scatter) على التوالي الحصينة المحتسبة على أساس MCD. وفيما يلي ادناه كيفية تشخيص (Leverage):
ليكن $C(p)$ هو قيمة القطع cut off، والذي يحسب بالشكل الآتي:

$$C(p) = \sqrt{\chi^2_{p;(1-\alpha)}} \quad (2-3)$$

اذ ان:

P: عدد المتغيرات التوضيحية بالاتمودج و α مستوى المعنوية وعليه يمكن تعريف متغير Leverage كالآتي: -

$$\text{Leverage} = \begin{cases} 0 & \text{if } RD_i \leq C(p) \\ 1 & \text{otherwise} \end{cases}$$

ج- الدمج بين البواقي الحصينة والمسافة الحصينة: [14]

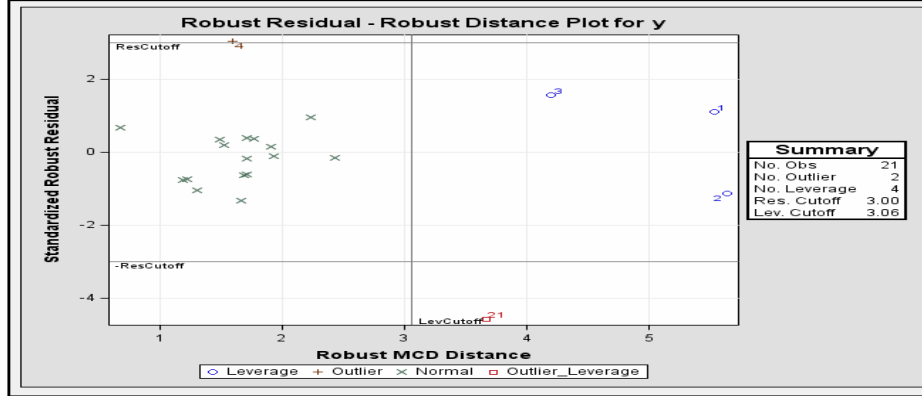
ان تعريف النقاط (Leverage) يستند كلياً على كون (x_i) بعيدة عن المركز وليس له علاقة بقيم الاستجابة (y)، وبإدخال قيمة (y_i) في التعريف لقيم (Leverage) فان الباحثان (Rousseeuw and Van Driessen) سنة (1999) ميزا بين (Leverage) السيء والجيد وكالاتي:-

1- النقاط (good Leverage) هي النقاط ($x_i, y_i()$) التي تكون قريبة من خط الانحدار ففي هذه الحالة نقاط (good Leverage) تحسن دقة معاملات الانحدار.

2- النقاط (Bad Leverage) وهي النقاط ($x_i, y_i()$) التي تكون بعيدة عن خط الانحدار وفي هذه الحالة نقاط (Bad Leverage) تقلل من دقة الانحدار معاملات الانحدار، فعندما تؤثر (x_i) على أساس (RD_i) على انها

(Leverage) والقيمة (y_i) المقابلة لها تؤثر في ضوء البواقي الحصينة (Robust residuals)، على أنها قيمة شاذة ففي هذه الحالة فإن (x_i) هو (Bad Leverage)، أما عندما يؤثر (RD_i) على (x_i) فهي (Leverage) إلا أن قيمة (y_i) المقابلة في ضوء البواقي الحصينة هي ليست قيمة شاذة فإنه (x_i) تؤثر أنها (good Leverage).

وقد اقترح الباحثان إيجاد العرض البياني للبواقي القياسية الحصينة للانحدار الحصين مثل [(LTS) أو (LMS)..... وغيرها]، مقابل المسافات الحصينة (RD) وذلك بأن يكون هنالك خطان أفقيان مناظرة إلى (k_σ) تكون مفيدة للتمييز بين البواقي الصغيرة والكبيرة، وخط عمودي واحد مناظر إلى $\sqrt{\chi^2_{p,(1-\alpha)}}$ يست للتمييز بين المسافات الصغيرة والكبيرة وكما مبين في الشكل (1) ادناه:-



الشكل (1)

تشخيص القيم الشاذة و (Leverage) عن طريق الدمج بين البواقي الحصينة والمسافة الحصينة

2-2-2 الانحدار الحصين: (Robust Regression) [3]

كما هو معروف فإن تقديرات طريقة (OLS) لأنموذج الانحدار هي حساسة وبشكل كبير جداً لوجود القيم الشاذة حتى وإن كان مقدار تلوث البيانات صغيراً إذ أن وجود مفردة واحدة شاذة سوف يكون لها تأثيراً كبيراً على تقدير المعلمات. إن نقطة الانهيار (Breakdownpoint) التي هي عبارة عن نسبة التلوث بالبيانات تعتبر مؤشر لمدى حصانة المقدرات وعدم تأثرها ومقاومتها (resistant)، لوجود القيم الشاذة بهذه النسبة، ونجد أن طريقة (OLS) يكون (Breakdown point) الخاص بها يساوي $\frac{1}{n}$ والذي يقترب من الصفر. وعلى هذا الأساس تم وضع العديد من الطرائق الحصينة التي تهدف إلى الحصول على نقاط انهيار عالية (high Breakdown)، أو كفاءة عالية أو كلاهما بغية الحصول على تقديرات لمعاملات الانحدار يمكن الاعتماد عليها، والآتي بعض من طرائق التقدير الحصين:-

- طريقة M (M-Estimation).
- طريقة (LTS).
- طريقة (LTS-RLS).
- طريقة (LMS).
- طريقة (R).
- طريقة (S).
- طريقة (MM).

والآتي التفاصيل الخاصة بطريقة (LTS) وطريقة (LTS-RLS) لما لها علاقة بمجريات هذا البحث.

2-2-2-1 تقدير المربعات الصغرى المشرذمة الحصين [3],[7],[14],[15]

(Robust Least Trimmed squares Estimation)

إن تقدير المربعات الصغرى المشرذمة والتي يرمز لها بـ (LTS) والمقترحة من قبل الباحث (Rousseeuw) يمكن أن تعرف بالشكل الآتي، وذلك في ضوء أنموذج الانحدار الخطي المفترض معادلة (2-4)، والمقدر معادلة (2-6):

$$\underline{b} = \arg \min \sum_{i=1}^h e_i^2 \quad (2-4)$$

اذ أن: $e_1^2 \leq e_2^2 \leq \dots \leq e_n^2$. وعليه فإن طريقة (LTS) قائمة على أساس المجموعة الجزئية (sub set) التي تتضمن (H) من المشاهدات [من المجموع الكلي للمشاهدات n] والتي تكون تقديرات المربعات الصغرى للمعاملات المحتسبة على أساسها تحقق اصغر مجموع مربعات بواقي، إذ أن:-

$$\frac{n}{2} + 1 \leq H \leq \frac{3n + p + 1}{4}$$

حيث p عدد المتغيرات المستقلة .
وان نقطة الانهيار :

$$\text{Break point} = \frac{n - H}{n} \quad (2 - 5)$$

الخطأ المعياري المقدّر ($\hat{\sigma}_{LTS}$) والذي يرمز له (S_{LTS}) يحسب بالشكل الاتي :

$$S_{LTS}(x, y) = d_{H,n} \sqrt{\frac{1}{H} \sum_{i=1}^H e_i^2} \quad (2 - 6)$$

$$d_{H,n} = \frac{1}{\sqrt{1 - \frac{2n}{H \cdot C_{H,n} \phi\left(\frac{1}{C_{H,n}}\right)}}} \quad (2 - 7)$$

$$C_{H,n} = \frac{1}{\Phi^{-1}\left(\frac{H+n}{2n}\right)} \quad (2 - 8)$$

اذ ان Φ, ϕ هي دالة التوزيع ، وكثافة التوزيع ، للتوزيع الطبيعي القياسي (standard Normal distribution) على التوالي . وقد أشار الباحث (Rousseeuw) سنة (1987) انه في حالة كون حجم العينة صغيراً $p=1$ فانه يتم استخدام كافة العينات الجزئية وبعدد C_p^n . اما عندما يكون حجم العينة كبيراً ، $p \geq 2$ فانه يتم استخدام خوارزمية سريعة سميت (Fast – LTS algorithm) والمقترحة من قبل الباحثين (Rousseeuw and Van Driessen) سنة (1999) . يتصف تقدير (LTS) بكونه يمتلك نقطة انهيار عالية ويساوي 0.50 ، كما انه متسق وطبيعي تقاربياً وبذلك فهو يتميز عن التقديرات بنقاط الانهيار العالية الأخرى مثل (LMS) الا ان تقدير (LTS) يتصف بكون الكفاءة النسبية له هي (0.08) وهي نسبة صغيرة وعلى هذا الأساس فقد ظهرت العديد من الطرائق الأخرى لتحسين كفاءة (LTS) منها طريقة إعادة المربعات الصغرى الموزونة (Reweighted least squares) والتي يرمز لها (RLS).

2-2-2-2 تحسين كفاءة أداء (LTS) [9],[12],[13],[15]

(تقدير المربعات الصغرى الموزونة المشددة LTS-RLS)

لتحسين كفاءة (LTS) ، ظهرت العديد من المقترحات وكان أولها سنة (1984) والموضوعة من قبل الباحث (Rousseeuw) والتي سماها طريقة المربعات الصغرى الموزونة المعادة (Reweighted least squares) والتي يرمز لها (RLS) اذ ان الاوزان تستند على التقديرات الأولية لطريقة (LMS)، وتم تطويرها من قبل (Rousseeuw and Leroy) سنة (1987) وفي سنة (1997) وضع الباحثان (Rousseeuw and Hubert) طريقة (LTS) القائمة على أساس المربعات الصغرى الموزونة المعادة (LTS-RLS) والمبينة تفاصيلها في الخطوات الاتية:-

- 1- يتم إيجاد مقدر موجه للمعطيات بطريقة (LTS) ويرمز له بـ (b_{LTS}) وكذلك مقدر الخطأ المعياري (S_{LTS}) وفق المعادلة (2-6) ، وليكن $\hat{\sigma}_0$.
- 2- يتم إيجاد البواقي القياسية u_{0i} كالآتي:-

$$u_{0i} = \frac{(y_i - x'_i b_{LTS})}{\hat{\sigma}_0} \quad (2 - 9)$$

- 3- يتم إيجاد الوزن (W_i) حيث يعرف بالشكل الاتي:-

$$W_i = \begin{cases} 0 & \text{if } |u_{0i}| > k \\ 1 & \text{otherwise} \end{cases} \quad i = 1, 2, \dots, n$$

حيث k اما تساوي (2.5) او (3).

- 4- يتم تعديل تقدير (Scale) وكالاتي:-

$$\hat{\sigma}_1^2 = \frac{1}{\sum_{i=1}^n W_i - p} \sum_{i=1}^n W_i (y_i - x'_i b_{LTS})^2 \quad (2 - 10)$$

- 5- يتم إيجاد البواقي القياسية مجددة وليكن u_{1i} وكالاتي :-

$$u_{1i} = \frac{(y_i - x_i' b_{LTS})}{\hat{\sigma}_1} \quad (2 - 11)$$

6- يتم اختيار المشاهدات نسبة الى $|u_{1i}| \leq k$ ، ثم يتم تطبيق (OLS) على هذه العينة التي تم تشذيبها (clean sample) ثم يتم احتساب الخطأ المعياري () $\hat{\sigma}_{OLS}$ من هذه العينة ايضاً .

ولقد تم إيجاد طريقة المربعات الصغرى الموزونة (LTS-RLS) بطريقة أخرى تم وضعها من قبل الباحثين (Marona, Martin and yohai) سنة (2006) وذلك وفق الخطوات المبينة ادناه:-

1- يتم تقدير موجه المعلمات بطريقة (LTS) وليكن b_{LTS} ثم يتم إيجاد البواقي :

$$e_{i,LTS} = (y_i - x_i' b_{LTS}) \quad i = 1, 2, \dots, n$$

2- يتم تقدير معلمة القياس الحصينة (Robust scale) $\hat{\sigma}_0$ ، وفق الاتي :-

$$\tilde{\sigma}_0 = \frac{1}{\Phi^{-1}(0.75)} \text{median}(|e_{1,LTS}|, |e_{2,LTS}|, \dots, |e_{n,LTS}| \text{ for } e_{i,LTS} \neq 0) \quad (2 - 12)$$

$$\tilde{\sigma}_0 = \frac{MAD}{\Phi^{-1}(0.75)} \quad (2 - 13)$$

3- يتم إيجاد البواقي القياسية $u_{0,i}$:

$$u_{0,i} = \frac{e_{i,LTS}}{\tilde{\sigma}_0} \quad i = 1, 2, \dots, n \quad (2 - 14)$$

4- يتم اختيار المشاهدات نسبة الى $|u_{0,i}| \leq k$ ، ثم يتم تطبيق (OLS) على هذه العينة المشدبة (clean sample)

2-3 المحور الثاني : مشكلة عدم تجانس التباين (Heteroscedasticity): [4],[5]

من المعروف بان مقدرات (OLS) لمعاملات النموذج الانحدار الخطي تتصف بكونها تمتلك كافة الخصائص المرغوب بها اذ انها كفوء وغير متحيزة ومتسقة وذلك في حالة تحقق عدد من الفرضيات الاساسية في النموذج، وبرز هذه الفرضيات هي فرضية تجانس التباين فعلى فرض انه لدينا النموذج الاتي :-

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\epsilon} \quad (2 - 15)$$

اذ ان :

$\underline{Y}$: موجه بـ $(n * 1)$ من المشاهدات لمتغير الاستجابة .

$\underline{X}$: $[n * (p + 1)]$ من المتغيرات المستقلة .

$\underline{\beta}$: موجه $[(p + 1) * 1]$ لمعاملات النموذج .

$\underline{\epsilon}$: موجه $(n * 1)$ الأخطاء .

فعند تحقق فرضية تجانس التباين (Homoscedastic) فان :

$$E \underline{\epsilon} \underline{\epsilon}' = \sigma^2 In \quad (2 - 16)$$

في حين عند مخالفة الفرضية يكون التباين غير متجانس (Heteroscedasticity) وان :

$$E \underline{\epsilon} \underline{\epsilon}' = \sigma^2 \Omega = \begin{bmatrix} \frac{\sigma^2}{W_1} & 0 & \dots & 0 \\ 0 & \frac{\sigma^2}{W_2} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{\sigma^2}{W_n} \end{bmatrix} = \begin{bmatrix} \sigma^2_1 & 0 & \dots & 0 \\ 0 & \sigma^2_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2_n \end{bmatrix}$$

ان وجود مشكلة عدم تجانس التباين، تنعكس على مقدرات (OLS) حيث تفقد الكثير من خصائصها، اذ انها (not BLUE) فضلاً عن التباين المقدر للمعاملات المقدرة بـ (OLS) تكون (متحيزة)، وهذا ينعكس أيضاً على جعل الاستدلال الاحصائي مضلل ولا يمكن الاعتماد عليه. وعلى هذا الأساس ظهرت طرائق تقدير لمعالجة مشكلة عدم تجانس التباين في أنموذج الانحدار.

2-3-1 معالجة مشكلة عدم تجانس التباين : [2],[6]

يمكن معالجة مشكلة عدم التجانس في التباين وذلك باستخدام طريقة المربعات الصغرى الموزونة (Weighted Least Square) (وهي حالة خاصة من طريقة المربعات الصغرى العمومية GLS) لتقدير معلمات النموذج وذلك وفق الصيغة الآتية:-

$$\hat{\beta}_{WLS} = (x' \Omega^{-1} x)^{-1} x' \Omega^{-1} y \quad (2-17)$$

اذ ان مصفوفة Ω^{-1} تعرف:

$$\Omega^{-1} = \begin{bmatrix} W_1 & 0 & \dots & 0 \\ 0 & W_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & W_n \end{bmatrix}$$

وان W_i هو الوزن. فعندما يكون الوزن معلوم فان طريقة (WLS) تكون مثلى، ولكن في الواقع العملي يكون غير معلوم ولا بد من تقديره وعندئذ فان طريقة المربعات الصغرى الموزونة تسمى (FWLS)

$$\hat{\beta}_{FWLS} = (x' \hat{\Omega}^{-1} x)^{-1} x' \hat{\Omega}^{-1} y \quad (2-18)$$

$$\hat{\Omega}^{-1} = \begin{bmatrix} \hat{W}_1 & 0 & \dots & 0 \\ 0 & \hat{W}_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{W}_n \end{bmatrix}$$

ففي حالة كون طبيعة هيكلية نموذج عدم التجانس معروف يمكن تقديره ولكن في كثير من الاحيان تكون غير معروفة، ولذلك فقد ظهرت العديد من البحوث تتناول كيفية تقدير W_i في مثل هذه الحالات وهي كالآتي:

2-3-1-1 طريقة (MPV): [10]

وهي الطريقة التي تم وضعها من قبل مجموعة الباحثين (Montgomery, Peck and Vining) سنة (2001) وسميت باسمهم (MPV) كمختصر، حيث يتم تقدير الوزن (W_i) بغية استخدام (WLS) لتقدير المعلمات وفق المعادلة (2-19). وخطوات هذه الطريقة هي كالآتي:

$$y_i = \beta_0 + \beta_1 x_1 + \epsilon_i, \quad i = 1, 2, \dots, n \quad (2-19)$$

- 1- يتم ترتيب قيم $x_i()$ تصاعدي، ثم تنظم قيم (y_i) المناظرة للترتيب التصاعدي في $x_i()$.
- 2- يتم تقسيم البيانات الى مجاميع كل مجموعة تشتمل على القيم المتقاربة مع بعضها البعض، فنحصل على عدد من المجاميع ولتكن (g).
- 3- يتم إيجاد الوسط الحسابي ($\bar{x}$) لكل مجموعة في (x).
- ويتم إيجاد التباين (S^2) لكل مجموعة في (y). فنحصل على:

$$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_g$$

$$S_1^2, S_2^2, \dots, S_g^2$$

4- يتم إيجاد انحدار المتغير المعتمد (S_j^2) على المتغير التوضيحي $\bar{x}_j()$ وكالآتي:-

$$S_j^2 = \alpha_0 + \alpha_1 \bar{x}_j, \quad j = 1, 2, \dots, g \quad (2-20)$$

5- باستخدام (OLS) يتم تقدير المعلمات α_1, α_0 في المعادلة (2-20).

6- نجد التقدير ($\hat{S}_i^2$) وفق النموذج الآتي:

$$\hat{S}_i^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_i, \quad i = 1, 2, \dots, n \quad (2-21)$$

7- نجد تقدير الاوزان ($\hat{w}_i$) وكالآتي:

$$\hat{w}_i = \frac{1}{|\hat{S}_i^2|} \quad (2-22)$$

8- يتم تطبيق طريقة المربعات الصغرى الموزونة معادلة (2-18) وفق هذه الاوزان المقدرة ($\hat{w}_i$) لتقدير موجه المعلمات وإيجاد ($\hat{\beta}_{FWLS}$).

ان طريقة المربعات الصغرى الموزونة (MPV) تكون غير ذات جدوى وليس لها معنى ولا يمكن الاعتماد عليها في حالة وجود قيم شاذة، حيث تتأثر هذه الطرائق وبشكل كبير بالقيم الشاذة، وعلى هذا الأساس كانت هنالك حاجة الى طرائق تقدير مربعات صغرى موزونة حصينة لمعالجة مشكلة عدم تجانس التباين وبوجود

القيم الشاذة في آن واحد. ولقد ظهرت في الآونة الأخيرة ومنذ سنة (2008) بحوث تهدف الى وضع مثل هذه الطرائق الحصينة.

2-3-2 طرائق تقدير المربعات الصغرى الموزونة الحصينة: [11]

تقوم طرائق تقدير المربعات الصغرى الموزونة الحصينة على أساس وضع تقدير للوزن ($\hat{w}_i$) حصين للقيم الشاذة، ولا يتأثر بوجودها والذي يقود بالتالي الى جعل تقدير المربعات الصغرى الموزونة هي الأخرى حصينة يمكن اعتمادها في وضع تقدير (β) عند تزامن وجود مشكلة عدم التجانس في التباين مع وجود القيم الشاذة، من هذه الطرائق نذكر الآتي:

2-3-2-1 طريقة (MPV) الحصينة: [2] (Robust MPV)

كما اوضحنا في بحث (2-3-1-1) فان طريقة (MPV) في تقدير الوزن ($\hat{w}_i$)، بالأساس تقوم على إيجاد مؤشرين هما $\bar{x}_j$ ، S_j^2 لكل مجموعة واذا ان كلا المؤشرين ليست لهما حصانة للقيم الشاذة، اذ يتأثر كلا منهما وبشكل كبير بوجود القيم الشاذة وبالتالي فان الوزن المقدر ($\hat{w}_i$) غير حصين، ويقود الى تقدير لمربعات الصغرى الموزونة غير كفوءة، وعلى هذا الأساس كانت الفكرة باستبدال هذين المؤشرين بأخرى حصينة مثل الوسيط (Median)، ووسيط مطلق الانحرافات (MAD) في تطبيق خطوات طريقة (MPV)، ويرمز لهذه الطريقة (RMPV) ويمكن بيان خطواتها كالآتي:

1- يتم ترتيب القيم تصاعدياً لـ $x_i()$ وتعيين $y_i()$ المناظرة لها، وإيجاد المجاميع $x_j()$ ، $y_j()$ اذ ان: $j=1,2,\dots,g$

2- نجد الوسيط (Median) لكل مجموعة في (x) .

3- نجد وسيط الانحرافات المطلقة MAD لكل مجموعة في (y) . نحصل على:

$$\text{Median}(x_1), \text{Median}(x_2), \dots, \text{Median}(x_g), \\ \text{MAD}(y_1), \text{MAD}(y_2), \dots, \text{MAD}(y_g),$$

4- يتم إيجاد انحدار المتغير $[MAD(y_i)]^2$ على المتغير التوضيحي $\text{Median}(x_j)$ وكالاتي:

$$[MAD(y_i)]^2 = \alpha_0 + \alpha_1 \text{Median}(x_j) \quad (2-23)$$

5- يتم استخدام طريقة (OLS) لتقدير α_0, α_1 في معادلة (2-23).

6- نجد التقدير $[MAD(y_i)]^2$ وفق الانموذج الآتي:

$$[MAD(y_i)]^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_i, \quad i = 1, 2, \dots, n \quad (2-24)$$

7- نجد الاوزان ($\hat{w}_i$) الحصينة وكالاتي:

$$\hat{w}_i = \frac{1}{[MAD(y_i)]^2} \quad (2-25)$$

8- يتم تطبيق طريقة المربعات الصغرى الموزونة وفق العادلة (2-19) وذلك على أساس الوزن الحصين معادلة (2-25) أعلاه، لتقدير موجه المعلمات (β).

وتجدر الإشارة الى ان طريقة (MPV) الحصينة لم تنال الاهتمام الكافي ولم تدرس من قبل الباحثين بالرغم من كونها قائمه على أساس منطقي ومتين، وعلى هذا الأساس فقد حاولنا في هذا البحث تسليط الضوء على هذه الطريقة ودراسة ادائها.

2-3-3 طرائق الكشف عن مشكلة عدم تجانس التباين: [4], [10], [17]

هنالك العديد من الطرائق التي يمكن الاعتماد عليها في الكشف عن مشكلة عدم تجانس التباين في انموذج الانحدار الخطي، والآتي التفاصيل الخاصة باختبار وايت لما له علاقة مباشرة في هذا البحث فضلاً عن انه تم تطبيقه في الجانب العملي.

2-3-3-1 اختبار وايت: (Whites Test): [17]

اقترح هذا الاختبار من قبل الباحث (halbert .white) في عام (1980)م ويعد هذا الاختبار من الاختبارات المستخدمة للكشف عن مشكلة عدم تجانس تباين الخطأ في أنموذج الانحدار، و أن هذا الاختبار يفترض أن (σ_i^2) دالة خطية من المتغيرات التوضيحية (X_s) من الدرجة الثانية أو أكثر ومن حاصل ضربهما (Cross Products)، كما أن هذه الطريقة لا تفترض أن (ϵ_i) يتوزع توزيعاً طبيعياً وليبيان خطوات هذه الطريقة نفترض أن لدينا أنموذج الانحدار في المعادلة التالية:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad (2-26)$$

وتم احتساب ما يأتي:

1- نجد (e_i) بطريقة (OLS).

2- يتم وضع أنموذج انحدار مساعد (Auxiliary) وعلى افتراض وجود متغيرين (X_2, X_1) فقط في الأنموذج الأصلي وكما مبين أدناه:

$$\hat{e}_i^2 = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \alpha_3 x_{1i}^2 + \alpha_4 x_{2i}^2 + \alpha_5 x_{1i} x_{2i} + e_i \quad (2-27)$$
 وتجدر الإشارة الى أن الأنموذج المشار إليها (Auxiliary) يتضمن المعلمة الثابتة (α_0) بغض النظر فيما إذا أن الأنموذج الأصلي المفترض في معادلة (2-26) يتضمن معلمة ثابتة أو لا.

3- يتم إيجاد (R^2) لأنموذج الانحدار المساعد (Auxiliary) إذ أن:

$$nR^2 \sim \chi^2_{(d \cdot f)} \quad (2-28)$$

و أن $(d \cdot f)$ هي درجة الحرية التي تمثل عدد المعلمات باستثناء المعلمة (α_0) و في الأنموذج المساعد المفترض فإن عدد المعلمات هي (5) وتقارن قيمة (χ^2) المحسوبة في الخطوة (3) مع قيمة (χ^2) الجدولية فإذا كانت أكبر هنالك مشكلة عدم تجانس التباين، أما إذا كانت أصغر فانه لا توجد مشكلة عدم تجانس التباين وأن فرضية العدم هي :

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

والجدير بالذكر أن بعض المصادر تسمي هذه الطريقة باختبار عدم تجانس العام (White's General Heteroscedasticity).

3-4-2 الطرائق الحصينة للكشف عن مشكلة عدم تجانس التباين: [4], [16]

لقد ظهرت العديد من الدراسات والبحوث تناولت مسألة اختبار وجود مشكلة عدم تجانس التباين في حالة وجود القيم الشاذة وعندما يبتعد توزيع الخطأ عن التوزيع الطبيعي، إلا أن أبرز هذه الطرائق هي تلك التي وضعت في السنوات الأخيرة وهي كالآتي: -

1-4-3-2 التعديل الحصين لاختبار كولد فيلد-كوانت: [4], [16]

(Robust Modification of Goldfeld-Quandt Test)

أن هذه الطريقة الحصينة تم وضعها من قبل مجموعة الباحثين (Rana, Midi and Imon) سنة (2008)، تنصب على تعديل إحصاء اختبار (GQ) وتحسينها من خلال معرفة أي من مركبات اختبار (GQ) قد تأثرت بالقيم الشاذة، ومن ثم يتم استبدال هذه المركبات بأخرى حصينة، ويرمز لهذا الاختبار (MGQ) اختصاراً لـ (Modification of Goldfeld-Quandt) وهذا الاختبار أكثر قوة من الاختبار التقليدي (GQ) في حالة وجود القيم الشاذة، أما خطوات هذا الاختبار فهي كالآتي: -

1- كما في الاختبار التقليدي (GQ) يتم ترتيب القيم تصاعدياً لـ (x_i) ويتم ترتيب $y_i()$ المناظرة لهذه القيم المرتبة

2- يتم حذف (c) من المشاهدات المركزية ويتم تقسيم البيانات المتبقية $(n-c)$ الى مجموعتين كل منها بـ $\frac{(n-c)}{2}$ من المشاهدات.

3- يتم فحص وجود القيم الشاذة باستخدام الانحدار الحصين (LTS) في كل مجموعة ثم يتم احتساب البواقي (deletion residuals) لبيانات المجموعة بعد حذف القيم الشاذة منها والمؤشرة بـ (LTS).

أن (deletion residuals)، والذي يرمز له بـ (di) يمكن أن يعرف وفق المعادلة الآتية:-

$$di = y_i - \hat{y}_{i(i)} \quad (2-29)$$

اذ أن:

y_i : يمثل استجابة الملاحظة المفردة i^{th} .

$\hat{y}_{i(i)}$: يمثل الاستجابة المقدرة للمفردة i^{th} على أساس الأنموذج المقدر بعد حذف الملاحظة i^{th} .

والفكرة الأساس لـ (deletion residuals) هي بحذف مشاهدة واحدة في كل مرة ومن ثم يتم تقدير معلمات الأنموذج وإيجاد الأنموذج المقدر على أساس المشاهدات المتبقية $(n-1)$ ، وهذه تعتبر طريقة مطولة حيث يتم إيجاد $(n-1)$ من الانحدارات المقدرة لاحتساب (deletion residuals)، ويمكن تطبيق معادلة أخرى أكثر سهولة لاحتساب (di) والتي تعتمد على عناصر قطر مصفوفة (Hat-matrix) وكالآتي:-

$$di = \frac{e_i}{1 - h_{ii}} \quad (2-30)$$

اذ أن e_i بواقي (OLS) وان h_{ii} العنصر (i) في قطر مصفوفة (Hat).

4- لكلا المجموعتين يتم احتساب وسيط مربع البواقي (deletion residuals) والذي يرمز له بـ

(MSDR) وهو اختصار لـ (Median of Squared Deletion Residual)، ثم يتم احتساب

النسبة الآتية:-

$$MGQ = \frac{MSDR_2}{MSDR_1} \quad (2 - 31)$$

اذ ان: $MSDR_1$ ، $MSDR_2$ هي وسيط مربعات البواقي (deletion) للمجموعة الصغيرة والكبيرة على التوالي. اذ ان احصائه (MGQ) تتبع توزيع F بدرجة حرية كل من البسط والمقام مساوية الى $\frac{(n-c-2k)}{2}$ فاذا كانت F_c المحتسبة اكبر من F الجدولية توجد هنالك مشكلة عدم التجانس في التباين.

2-3-4-2 طريقة تعديل البيانات:

ان المنهجية الخاصة بهذه الطريقة هي في البحث واكتشاف القيم الشاذة في البيانات باستعمال الطرائق الحصينة، ومن ثم تقسيم البيانات الى جزئين الأول وهو الأكبر والذي يكون (خالياً) من القيم الشاذة ويسمى بالجزء النظيف (Clean Part)، والجزء الآخر وهو الذي يحتوي على القيم الشاذة، اذ ان الجزء النظيف (Clean Part) يمكن استخدامه لمعرفة فيما اذا كانت هنالك مشكلة عدم تجانس التباين ام لا وذلك بتطبيق كافة الاختبارات التقليدية المعروفة في الكشف عن مشكلة عدم التجانس وكما هي موضحة في مبحث (2-3-3)، وتكون هذه الاختبارات معتمده ويمكن استعمالها للكشف عن مشكلة عدم التجانس في التباين.

الفصل الثالث التطبيق العملي

3-1 المقدمة:

يتضمن هذا الفصل التطبيقات العملية للمفاهيم الخاصة باكتشاف القيم الشاذة باستعمال الطرائق الحصينة المتمثلة بطريقة المربعات الصغرى المشرومة (LTS) وكذلك العروض البيانية المعتمدة على الربط بين البواقي الحصينة والمسافات الحصينة لـ (Mahalanobis) والاختبارات الحصينة للكشف عن مشكلة عدم تجانس التباين في حالة وجود القيم الشاذة وطرائق تقدير المعلمات لأنموذج الانحدار في حالة وجود هاتين المشكلتين معاً باستخدام طريقة المربعات الصغرى الموزونة الحصينة (RMPV) لبيانات (الدخل والانفاق الاستهلاكي).

3-2 الكشف عن القيم الشاذة باستخدام الطرائق الحصينة:

لقد تم استخدام الطرائق الحصينة في الكشف عن القيم الشاذة و Leverage في كلاً من قيم المتغير المعتمد والمتغير المستقل على التوالي وفق المفاهيم النظرية الموضحة في المبحث (2-2-2-1) في الجانب النظري، وعليه فقد تم تطبيق طريقة المربعات الصغرى المشرومة (LTS) لاحتساب البواقي الحصينة على أساسها، كما تم احتساب المسافات الحصينة RD الموضوع على أساس التقديرات الحصينة لـ (Location) و (Scatter Matrix) المحتسبة بطريقة MCD، وباستخدام البرنامج الجاهز (sas 9.3).
❖ البيانات الخاصة بالدخل والانفاق الاستهلاكي الشهري للعراق للفترة (1970-2014)
ان الامودج المفترض لهذه المجموعة هو:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

1 - 3

اذ ان: y: هو الدخل، x: هو الانفاق الاستهلاكي
ولقد تم تطبيق طريقة المربعات الصغرى المشرومة (LTS) على بيانات هذه المجموعة وكما هو موضح في الجدول (1) و(2) و(3)، وكما مبينة في ادناه تباعاً:-

The ROBUSTREG Procedure

Model Information	
Data Set	WORK. Ali 1
Dependent Variable	Y
Number of Independent Variables	1
Number of Observations	45
Method	LTS Estimation

جدول (1)

LTS Parameter Estimates		
Parameter	DF	Estimate
Intercept	1	-118073
X	1	0.1783
Scale (S_{LTS})	0	512652.6
Scale (W_{scale})	0	323880.6

جدول (2)
القيم الشاذة وقيم Leverage باستخدام طريقة (LTS) لبيانات الدخل والاتفاق

Diagnostics						
Obs	x	Mahalanobis Distance	Robust MCD Distance	Leverage	Robust Residual	Outlier
30	31381049	0.1305	1.2095		-4.9287	*
31	46634635	0.0875	1.9475		-6.9508	*
33	34677723	0.0833	1.3690		5.7297	*
35	46923316	0.0916	1.9614		16.5542	*
36	65798567	0.3613	2.8746	*	9.4818	*
37	85431539	0.6418	3.8244	*	-0.3955	
38	100100817	0.8514	4.5341	*	9.7064	*
39	147641254	1.5307	6.8340	*	-0.1977	
40	120428411	1.1419	5.5175	*	19.0380	*
41	146453469	1.5138	6.7765	*	14.4167	*
42	192237070	2.1679	8.9915	*	8.7868	*
43	227221851	2.6678	10.6840	*	5.4585	*
44	243518659	2.9007	11.4724	*	13.7694	*
45	230310053	2.7119	10.8334	*	0.7247	

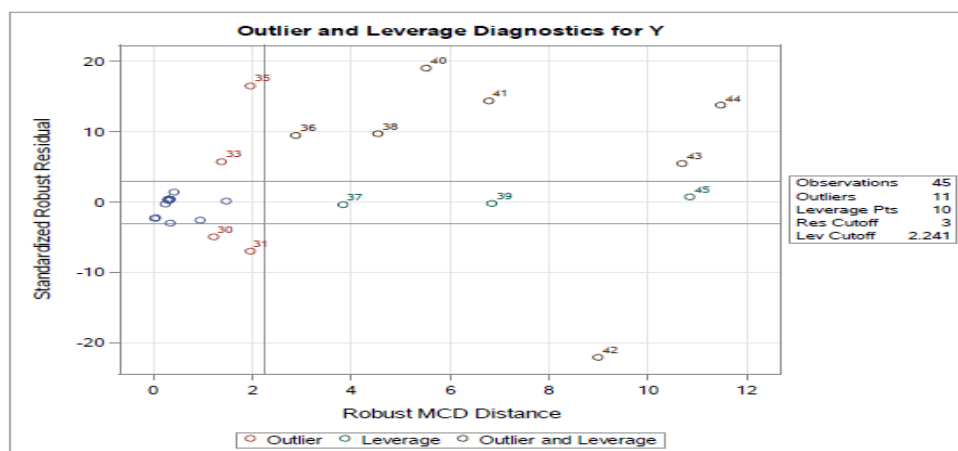
جدول (3)

Parameter Estimates for Final Weighted Least Squares Fit							
Parameter	DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept	1	-118073	50848.31	-217734	-18412.5	5.39	0.0202
x	1	0.1783	0.0010	0.1763	0.1803	30593.6	<.0001
Scale	0	279399.6					

في ضوء نتائج جدول (2) نجد ان هنالك قيم شاذة ممثلة في المشاهدات رقم (30) و(31) و(33) و(35) و(36) و(38) و(40) و(41) و(42) و(43) و(44)، في حين ان المشاهدات رقم (37) و(39) و(45) فيها (good Leverage) ولا يؤثر على استجابة (y).

3-3 الكشف عن القيم الشاذة و [Leverage] باستعمال العرض البياني للبواقي الحصينة والمسافات الحصينة:

لقد تم تشخيص القيم الشاذة والـ Leverage باستعمال العرض البياني للبواقي الحصينة والمسافات الحصينة (RD) وذلك بالنسبة لمجموعة بيانات الدخل والاتفاق باستخدام برنامج (sas version 9.3) وكما هي موضحة في الشكل (2)، وذلك وفق المفاهيم النظرية الموضحة في الجانب النظري مبحث (2-2-1-2).



الشكل (2)

القيم الشاذة و (Leverage) باستخدام المؤشرات الحصينة

في ضوء ملاحظة العروض البيانية نجد ان شكل (2) قد وضح ان المشاهدات رقم (30) و(31) و(33) و(35) هي شاذة، وانه قد وضح ان المشاهدات رقم (36) و(38) و(40) و(41) و(42) و(43) و(44) هي Leverage سيئة، في حين ان المشاهدات رقم (37) و(39) و(45) فيها (good Leverage) ولا يؤثر على استجابة (y).

3-4 الاختبار الحصين لـ (White) للكشف عن مشكلة عدم تجانس التباين في حالة وجود قيم الشاذة:

لقد تم تطبيق الاختبار الحصين (White) للكشف عن مشكلة عدم تجانس التباين في حالة وجود القيم الشاذة وذلك وفق المفاهيم النظرية المذكورة في المبحث (2-3-3-1) في الجانب النظري بالنسبة لبيانات (الدخل والانفاق) وذلك باستعمال برنامج (Gretl) وفيما يلي ادناه النتائج التي تم التوصل اليها كما هي موضحة في جدول (4).

جدول (4)

نتائج اختبار (White) في الكشف عن مشكلة عدم تجانس التباين في البيانات المعدلة

الاحصاء	White	
	χ^2_c	P
مجموعة البيانات	6.827514	0.032917
الدخل والانفاق لاستهلاكي		

في ضوء نتائج جدول (4) نجد ان مجموعة بيانات (الدخل والانفاق الاستهلاكي) فان قيمة الاحتمال P هو أصغر من (0.05) وان الاختبار يشير الى وجود مشكلة عدم تجانس التباين الى جانب القيم الشاذة، الامر الذي يقود الى ضرورة استعمال طريقة المربعات الصغرى الموزونة الحصينة لتقدير معلمات الانموذج وذلك بالنسبة لهذه المجموعة.

3-4-2 اختبار كولد فليد – كوانت الحصين:

(A Robust Modification of the Goldfeld-Quandt Test)

فضلاً عن اختبار (White) أعلاه فقد تم تطبيق طريقة (Goldfeld) الحصين بالنسبة للمجموعة (الدخل والانفاق الاستهلاكي) وذلك وفق الخطوات التي تم توضيحها في الجانب النظري مبحث (2-3-4-1) والخاصة بهذا الاختبار حيث تم الحصول على قيمة (F) المحتسبة وكانت تساوي $MGQ = 62856762.2$ ومقارنتها مع (F) الجدولية بدرجة حرية $(n-c-2k) / 2$.

$$\frac{n - c - 2k}{2} = \frac{45 - 7 - 2(2)}{2} = 17$$

$$F(17,17,0.05) = 1.86$$

قيمة (MGQ) أكبر من احصاء F الجدولية

اذن نرفض فرضية العدم ويوجد عدم تجانس التباين.

ونجد ان النتائج تشير الى ان F_C المحتسبة $< F_T$ الجدولية وبذلك فان كافة الاختبارات تشير الى وجود مشكلة عدم تجانس التباين في مجموعة (الدخل والانفاق).

3-5 طريقة المربعات الصغرى الموزونة الحصينة لتقدير معلمات الانموذج:

لقد تم تطبيق طريقة المربعات الصغرى الموزونة الحصينة لتقدير معلمات الانموذج في حالة وجود مشكلتي عدم تجانس التباين والقيم الشاذة معاً في الانموذج، وذلك بالنسبة الى مجموعة (الدخل والانفاق الاستهلاكي). ولقد تم تطبيق هذه الطريقة وذلك في ضوء المفاهيم والصيغ النظرية المذكورة في الجانب النظري في المبحث (2-3-1-1). ولقد تم استعمال برنامج (Gretl و statistica) في إيجاد النتائج الخاصة لتطبيق هذه الطريقة، والآتي النتائج التي تم التوصل اليها.

لقد تم الحصول على النتائج الخاصة لتطبيق هذه الطريقة وكما هي موضحة في الجدول (5) و(6).

جدول (5)

نتائج طريقة MPV الحصينة على مجموعة الدخل والانفاق

	Coefficient	Std. Error
β_0	-74508.3	51709.56
β_1	0.2	0.01

جدول (6)
ANOVA Table

Effect	Sums of Squares	df	Mean Squares	F	P-value
Regress	201.7580	1	201.7580	1423.616	0.000000
Residual	6.0941	43	0.1417		
Total	207.8521				

وقد تم احتساب معامل التحديد (R^2) للامودج حيث كانت قيمته

R^2
0.97068081

الفصل الرابع

4-1 الاستنتاجات:

- ان اهم الاستنتاجات التي يمكن الخروج بها من هذا البحث تتلخص بما يأتي: -
- 1- لقد تم بناء امودج لتقدير ظاهرة الانفاق في ضوء الدخل لسلسلة البيانات (2014-1970) يمكن الاعتماد عليه حيث تم تجاوز مشكلتي عدم تجانس التباين ومشكلة القيم الشاذة وتزامن وجودها في البيانات وذلك من خلال تطبيق طريقة المربعات الصغرى الموزونة الحصينة (RMPV).
 - 2- إمكانية الاعتماد على الطريقة الحصينة لاختبار كولد-فيلد (MGQ) وطريقة تعديل البيانات، في الكشف عن مشكلة عدم تجانس التباين عندما يتزامن وجودها مع وجود القيم الشاذة في آن واحد، إذ ان كلاً منها تقود الى نفس النتائج.
 - 3- تتصف الطريقة الحصينة القائمة على (تعديل البيانات) في الكشف عن مشكلة عدم تجانس التباين بالتزامن مع وجود القيم الشاذة، بسهولة قياس الى الطريقة الحصينة لاختبار كولد-فيلد (MGQ) التي تتطلب خطوات تنفيذية أطول.
 - 4- ان استبدال المقاييس الغير حصينة في طريقة (MPV) (التقليدية) المتمثلة بـ $(\bar{X})$ و (S^2) بأخرى حصينة والمتمثلة بـ (MAD, Median) على التوالي، كان كافياً لوضع وزن حصين يمكن الاعتماد عليه في تقدير المعلمات بطريقة المربعات الصغرى الموزونة لمعالجة مشكلة عدم تجانس التباين بالتزامن مع وجود القيم الشاذة، إذ ان هذا الاجراء يغني تماماً عن عملية التوسع باستخدام طريقة (LTS) لتقدير الوزن (W_i) الحصين ومعاملاته بأوزان (Huber).

4-2 التوصيات:

- 1- نوصي باستخدام الطريقة الحصينة القائمة على (تعديل البيانات) المذكورة في هذا البحث في الكشف عن مشكلة عدم تجانس التباين بالتزامن مع وجود القيم الشاذة، لكونها طريقة سهلة وبالإمكان تطبيقها، وذلك لتوفر البرامجيات الجاهزة مثل (sas, Gretel)، وتعطي ذات النتائج للطرائق الأخرى الأكثر تعقيداً.
- 2- استخدام طريقة المربعات الصغرى الموزونة الحصينة (RMPV) لتقدير معلمات الامودج في حالة تزامن وجود مشكلة عدم تجانس التباين والقيم الشاذة معاً في البيانات حيث يمكن تجاوز كافة السلبيات التي تنعكس على خصائص مقدرات المعلمات والحصول على مقدرات معلمات بخصائص يمكن الاعتماد عليها.

المصادر

- 1- احمد صدقي محمد الديب ، 2009 (دراسة مقارنة لبعض الطرق الحصينة للكشف عن المشاهدات الشاذة في نماذج الانحدار الخطي) جامعة الإسكندرية التجارة الإحصاء التطبيقي 2009 .
- 2- Bello Abdulkadir Rasheeda , Robiah Adnana, Seyed Ehsan Saffarib, Kafi dano Patia , (2014), " Robust Weighted Least Squares Estimation of Regression Parameter in the Presence of Outliers and Heteroscedastic Errors", Jurnal Teknologi .
- 3- Chun Yu, Weixin Yao, and Xue Bai , 2014 , (Robust Linear Regression: A Review and Comparison) , Kansas State University, Manhattan, Kansas, USA 66506-0802 .
- 4- Ekele Alih and Hong Choon Ong , (2015), "An Outlier – resistant test for heteroscedasticity in linear models" , Journal of Applied Statistics, 42: 8, 1617-1634, ISSN 0266-4763 .
- 5- Habshah Midi, Md. Sohel Rana, A. H. M. Rahmatullah Imon, (2009), "The Performance of Robust Weighted Least Squares in the Presence of Outliers and Heteroscedastic Errors", WSEAS TRANSACTIONS on MATHEMATICS .
- 6- Huseyin Tastan ,(2012), " HETEROSCEDASTICITY " , Yildiz Technical University Department of Economics .

- 7- Jurgen A. Doornik , 2011 (Robust Estimation Using Least Trimmed Squares) , Institute for Economic Modelling, Oxford Martin School, and Economics Department, University of Oxford, UK .
- 8- Marco Riani, Anthony C. Atkinson and Domenico Perrotta , (2014), "A Parametric Framework for the Comparison of Methods of Very Robust Regression" , Statistical Science 2014, Vol. 29, No. 1, 128–143 DOI: 10.1214/13-STS437 c Institute of Mathematical Statistics .
- 9- Maronna, R. A., Martin, R. D. and Yohai, V. J. (2006), Robust Statistics. John Wiley.
- 10- Montgomery, D. C. , Peck, E. A. and Vining, G.G . ,(2001) ,"Introduction to Linear Regression Analysis" , 3rd ed, Wiley, New York .
- 11- MULLER, C. (2002) , "Robust estimators for discontinuous function" <http://www.searchfor Science.com> .
- 12- Rousseeuw, P. J. (1984). Least median of squares regression. Journal of the American Statistical Association, 79, 871–880 .
- 13- Rousseeuw, P. J., and Hubert, M. (1997). Recent developments in PROGRESS. IMS Lecture Notes, 31, 201–214 .
- 14- Rousseeuw, P.J. and Van Driessen, K. (1999), "A Fast Algorithm for the Minimum Covariance Determinant Estimator," Technometrics,41,212-223 .
- 15- Rousseeuw, P.J., and A. Leroy, 1987, *Robust Regression and Outlier Detection*, Wiley, New York, 1987.
- 16- Sohail Rana . Md , Habshah Midi and A.H.M. Rahmatullah Imon , (2008) , "A Robust Modification of the Goldfeld-Quandt Test for the Detection of Heteroscedasticity in the Presence of Outliers " , Journal of Mathematics and Statistics 4 (4): 277-283, ISSN 1549-3644 .
- 17- White, H., (1980), "A heteroscedasticity- consistent covariance estimator and a direct test for heteroscedasticity" *econometrica* 48, 817-838.

.....
.....
.....