بعض مقدرات مصفوفة التباين المشترك المتسقة الحصينة في حالة عدم تجانس تباين الخطأ لأنموذج الانحدار الخطي العام

Some of the Robust Consistent Covariance Matrix Estimator in Case Heteroscedasticity of General Linear Regression Model

ا.د. سجى محمد حسين الباحث / قاسم محمد صاحب

كلية الادارة والاقتصاد - جامعة بغداد

مستخلص البحث

تعد مشكلة عدم تجانس التباين هي احدى مشاكل الانحدار الخطي ، التي تؤدي الى نتائج غير دقيقة عند استخدام طريقة المربعات الصغرى الاعتيادية (OLS) ، اضافة الى مشكلة عدم تجانس تباين الخطأ قد تحتوي البيانات على بعض القيم الشاذة المتطرفة في المتغيرات التوضيحية (**HLPs**) ، وجود تلك القيم يؤثر على كفاءة تقدير معلمات انموذج الانحدار الخطي أذ ما استخدمت الطرائق التقليدية ، و في حالة وجود مشكلة عدم تجانس التباين اضافة الى القيم الشاذة المتطرفة (HLPs) فأن الطرائق التقليدية سوف تعطي استدلالاً مضللا لا يمكن الاستناد اليه في تحديد فترت الثقة واختبار الفرضيات .

في هذا البحث سوف نستعرض بعض مقدرات مصفوفة التباين والتباين المشترك المتسقة الحصينة (RHCCM) وهي المقدر (HCw4) والمقدر (HCw5) ، هذه المقدرات تستند في ايجاد الاوزان والاخطاء على بعض الطرائق الحصينة التي تستخدم في الكشف عن القيم الشاذة المتطرفة (**HLPs**) ،و هي كل من مسافات مهلونوبس الحصينة على اساس اصغر حجم قطع ناقص **RMD(MVE)** ، مسافات مهلونوبس الحصينة على اساس اصغر محدد مصفوفة تباين مشترك **RMD(MCD)** ، تمت اجراء المقارنة بين المقدرات الحصينة لمعلمات انموذج الانحدار الخطي العام من خلال استعمال تجارب المحاكاة وفق اسلوب مونت كارلو ، والتي اعتمد فيها على وجود مشكلة عدم تجانس التباين مع اختلاف حدة عدم التجانس ، وكذلك وجود نسب مختلفة من (HLPs) في البيانات وايضا حجوم عينات مختلفة . وقد اعتمدت المقارنة على معيار الخطأ المعياري للمقدرين (HCw4) و (HCw5) و تم التوصل الى ان طريقة **RMD(MCD)** هي الافضل بالنسبة لكل من مقدر (HCw4) (HCw5).

المصطلحات الرئيسية للبحث / مقدر (HCw4) ، مقدر (HCw5) ، القيم الشاذة المتطرفة (HLPs) ، حدة عدم تجانس التباين ، RMD(MVE) ، RMD(MCD) .

Abstract

The problem of heteroscedastic errors is one of the problems of linear regression, which leads to inaccurate results when using the ordinary least squares method (OLS), in addition to the problem of heteroscedastic errors maybe the data contain some outliers extreme values in independent variables (High leverage points), presence These values effect the efficiency of estimating the parameters of the linear regression model if the classical methods are used, and in the case of the problem of heteroscedastic in addition to the outliers values (HLPS), the

classical methods will give a misleading inference that cannot be relied upon in determining confidence intervals and testing hypotheses.
In this paper we will review some of the Robust Heteroscedasticity Consistent Covariance Matrix (RHCCM) estimators are (HCW4) estimator and the (HCW5) estimator. These estimatiors are based on finding weights and errors on some robust methods that is used to detect of outliers extreme values (HLPS), which are both Mahalanobis Distance Based on Minimum Volume Ellipsoid RMD(MVE), Mahalanobis Distance Based on Minimum Determinant RMD(MCD), the comparison was done Among the robust Covariance estimators of the parameters for the general linear regression model by using simulation experiments according to the Montocarlo method, it was based on the presence of the problem of heteroscedastic with different severity of heterogeneity, as well as the presence of different percentage of (HLPS) in the data as well as different sample sizes. The comparison was based on the standard error criterion (SE) for estimators (HCW4), (HCW5) for the parameters of the linear regression model. It was concluded that the RMD (MCD) method is the best for both the (HCW4) (HCW5) estimators

١. المقدمة Introduction

الانحدار الخطي هو احد اهم الادوات الاحصائية واكثرها استخداما في نمذجة و تحليل العلاقة ما بين متغير الاستجابة (**Response Variable**) ومتغير واحد او اكثر من المتغيرات التوضيحية (**Independed Variable**) ، الا ان في كثير من الابحاث العلمية غالبا ما تكون بين متغير الاستجابة واكثر من متغير توضيحي والذي يعرف بانموذج الانحدار الخطي المتعدد وذلك لتعدد المتغيرات التوضيحية. تعتبر فرضية تجانس التباين هي احدى الفرضيات الاساسية التي تستند عليها طريقة المربعات الصغرى الاعتيادية (**OLS**) في تقدير معلمات انموذج الانحدار الخطي تتصف بالاتساق وعدم التحيز بالأضافة الى الكفاءة ، عدم تحقق هذه الفرضية تظهر لنا مشكلة عدم تجانس تباين الاخطاء ،وبالتالي فأن مقدرات (**OLS**) لا تمتلك صفة اقل تباين ممكن اي ان تلك المقدرات لن تكون افضل تقدير خطي غير متحيز (**BLUE**) وبذلك تكون فترات الثقة الموضوعة غير صحيحة . غالبا ما نجد مشكلة عدم تجانس التباين في البيانات المقطعية (**Cross Section Data**) ، كما ان وجود القيم الشاذة في البيانات يؤدي كذلك الى ظهور مشكلة عدم تجانس التباين .

في اغلب البيانات الحقيقية نجد بأنها تحتوي على بعض المشاهدات التي تتبع نمط معين يختلف عن نمط المشاهدات البقية ومنها القيم الشاذة المتطرفة في المتغير التوضيحي (النقاط الرافعة العليا) (HLPs) ، ان سبب وجود المشاهدات الشاذة قد يكون نتيجة اخطاء في القياس او التسجيل ، لكن هذا لا يستبعد قد تكون هذه المشاهدة الشاذة عبارة عن معلومة قيمة تتمثل بجودة البيانات كونها معلومة حقيقية حصلت في ظروف آنية ، الا ان وجود (HLPs) يمكن ان يؤدي تأثيرها الى استنتاجات غير صحيحة بسبب تقديرات مشوهه لكل من مقدرات الموقع والشكل ، كما ان وجود (**HLPs**) يعتبر اكثر تأثيرا من حدة عدم تجانس التباين في سلوك العينات المحدودة ولمختلف مقدرات مصفوفة التباين والتباين المشترك المتسقة (**HCCM**) **.**

ان وجود مشكلة عدم تجانس التباين و القيم الشاذة المتطرفة (**HLPs**) في البيانات معا دفعنا الى استخدام مقدرات مصفوفة التباين والتباين المشترك المتسقة الحصينة (RHCCM) ، كونها اقل حساسية تجاه (**HLPs**) وايضا تحد من تأثير مشكلة عدم تجانس التباين .

اهتم بعض الباحثين بهذا الموضوع منهم الباحث (Furno) [4] الذي استخدم البواقي الحصينة في ايجاد مقدرات مصفوفة التباين المشترك المتسقة (HCCM) عند وجود القيم الشاذة المتطرفة في المتغيرات التوضيحية (**HLPs**) ، والذي تطلب احتساب الاوزان الحصينة للحد من تأثيرات تلك القيم . فكرة استخدام البواقي الحصينة هي تقليل التحيز لمقدرات (HCCM) والحصول على مقدرات متسقة وحصينة لمصفوفة التباين المشترك (RHCCM) .

في عام ٢٠٠٤ (Neto) [3] قدم مقدر **HC4** وهو مقدر متسق جديد لمصفوفة التباين المشترك عند وجود مشكلة عدم تجانس التباين وكذلك وجود القيم الشاذة المتطرفة في المتغيرات التوضيحية (النقاط الرافعة العليا) (HlPs) في العينات المحدودة ، مقدر **HC4** يستند الى مقدر **HC3** الا انه يأخذ بالحسبان النسبة بين قياس قوة الرفعة $\mathbf{h_i}$ ومتوسطها ، اظهرت النتائج ان الاستدلال المستند الى المقدر المقترح هو اكثر واقعية من الاستدلال المستند الى مقدر **HC3** .

في عام ٢٠٠٧ اقترح (Neto) واخرون [2] مقدر **HC5** وهو مقدر متسق جديد لمصفوفة التباين والتباين المشترك عند وجود مشكلة عدم تجانس التباين و وجود القيم الشاذة المتطرفة المتعددة في المتغيرات التوضيحية (النقاط الرافعة العليا) (HIPs) ، يعتبر هذا المقترح اول مقدر يأخذ بنظر الاعتبار القيمة العظمى (maximum leverage) للقيم الشاذة المتطرفة في المتغير التوضيحي (HIPs) اضافة الى مميزات مقدر **HC4** ، من خلال استخدم بيانات حقيقة تبين ان الاستدلال بالاستناد الى المقدر المقترح هو اكثر ثقة من الاستدلال على بقية مقدرات مصفوفة التباين والتباين المشترك الاخرى .

٢ - هدف البحث The Aim of the Research

استعراض بعض الطرائق الموزونة الحصينة على اساس مقاييس اكتشاف القيم الشاذة المتطرفة في المتغير التوضيحي (النقاط الرافعة العليا) (**HLPs**) وهي كل من مسافة مهلونوبس الحصينة على اساس اصغر حجم قطع ناقص RMD(MVE) ومسافة مهلونوبس الحصينة على اساس اصغر محدد مصفوفة تباين مشترك RMD(MCD) واستعمالها في طريقة المربعات الصغرى الموزونة لتقدير معلمات انموذج الانحدار الخطي العام ومن ثم حساب مقدرات مصفوفة التباين و التباين المشترك المتسقة الحصينة (RHCCM) مثل (HC_{W4}) (HC_{W5}) عند وجود مشكلة عدم تجانس التباين والقيم الشاذة المتطرفة في المتغير التوضيحي (النقاط الرافعة العليا) (**HLPs**) وبالتالي المقارنة بين التقديرات الحصينة لمعلمات انموذج الانحدار الخطي العام من خلال المعيار **SE** للمقدرين (HC_{W4}) (HC_{W5}) .

٣ - الجانب النظري

تعتمد دقة تقدير معلمات انموذج الانحدار الخطي عند أستخدام طريقة المربعات الصغرى الاعتيادية (**OLS**) على توفر فرضية تجانس التباين للاخطاء ، سواء كان نموذج انحدار خطي بسيط او متعدد ، وان عدم تحقق فرضية تجانس التباين تؤدي الى ان المعلمات المقدرة لن تكون افضل تقدير خطي غير متحيز (**BLUE**) اي ان المعلمات المقدرة لا تمتلك اقل تباين ممكن .

نفرض لدينا نموذج الانحدار الخطي الاتي :

$$\underline{Y} = X\underline{\beta} + \underline{U} \qquad (1)$$

حيث ان :

$\underline{Y}$: موجه متغير الاستجابة من الدرجة $(n * 1)$

X : مصفوفة المتغيرات التوضيحية من الدرجة $[n * [(p + 1)]$

$\underline{\beta}$: موجه معلمات الانموذج من الدرجة $[(p + 1) * 1]$

$\underline{U}$: موجه حد الخطأ العشوائي من الدرجة $(n * 1)$

ولايجاد تقديرات معلمات الانموذج عند وجود مشكلة عدم تجانس التباين نستخدم اسلوب المربعات الصغرى الموزونة كما في المعادلة الاتية :

$$\widehat{B}_{WLS} = (X'W^{-1}X)^{-1}X'W^{-1}Y \qquad (2)$$

تعرف طريقة التقدير هذه بطريقة مقدر اتكن (**Aitkin Estimator**) ويعرف هذا الاسلوب بطريقة المربعات الصغرى العامة (**Generalized Least Squares**) ، ولاعتمادها على الاوزان (w_i) فأنها تعرف بطريقة المربعات الصغرى الموزونة (**WLS**). **[1 pp.304]**.

كما تعرف مصفوفة التباين والتباين المشترك للمعلمات المقدرة بالصيغة التالية :

$$var - cov(\widehat{B}_{WLS}) = \sigma^2(X'W^{-1}X)^{-1} \qquad (3)$$

ان وجود القيم الشاذة المتطرفة في المتغيرات التوضيحية (**HLPs**) يجعل من توزيع تلك البيانات ثقيل الذيل (**Heavy Tailed Distribution**) وينتهك عندها افتراض التوزيع الطبيعي لها **[7]**.

وبذلك فأن وجود بعض القيم الشاذة في البيانات من الممكن ان تجعل من الطرائق التقليدية للتقدير ومنها طريقة المربعات الصغرى الموزونة غير مجدية لكونها سوف تعطي استدلالا غير واقعي .

٤ - تقدير معلمات انموذج الانحدار الخطي

Estimates of the Regression linear model parameters

تقدير معلمات انموذج الانحدار عند وجود مشكلة عدم تجانس التباين يتطلب ايجاد مصفوفة الاوزان في طريقة المربعات الصغرى الموزونة ، الا ان وجود مشكلة عدم تجانس التباين اضافة الى وجود القيم الشاذة المتطرفة في المتغيرات التوضيحية (**HLPs**) سوف نحصل على نتائج مضللة نتيجة تأثير تلك القيم ، لذلك تطلب ايجاد اوزان حصينة تطبق في طريقة المربعات

الصغرى الموزونة ، تم ايجاد تلك الاوزان عن طريق بعض الطرائق الحصينة التي تشترك في نقطة وهي اعطاء وزن اقل للمشاهدة الشاذة الغرض من ذلك هو تقليل تأثيرها ، من تلك الطرائق نذكر :

1-4 مسافة مهلونوبس الحصينة على اساس اصغر حجم قطع ناقص Robust Mahalanobis Distance Based on Minimum Volume Ellipsoid RMD(MVE)

تعتبر طريقة مسافات مهلونوبس التي تستند الى اصغر حجم قطع ناقص **RMD(MVE)** احدى الطرائق الحصينة في اكتشاف القيم الشاذة ، فكرة طريقة **RMD(MVE)** هي البحث عن اصغر حجم مجسم قطع ناقص من بين جميع القطوع الناقصة التي تغطي على الاقل نصف البيانات، وتعويض مقدرات الموقع والتشتت لذلك القطع في صيغة حساب مسافات مهلونوبس ثم الكشف عن القيم الشاذة . [7][11]

تمتاز هذه الطريقة بانها تستطيع تحمل نسب عالية من وجود القيم شديدة التطرف يعود سبب ذلك الى امتلاك مقدرات **(MVE)** خصائص المقدرات الحصينة ومنها تساوي التغاير وايضا نقطة انهيار عالية تقترب الى ان تصل الى ٥٠٪ ، كما وتعتبر ايجاد مقدرات **(MVE)** طريقة توافقية اذ تعتمد على سحب عينات جزئية من المشاهدات وبشكل متكرر من خلال اسلوب اعادة المعاينة ، الغرض من ذلك هو الحصول على عينة جزئية يؤمل عليها ان تكون خالية من المشاهدات الشاذة .
قدمت الكثير من الخوارزميات في حساب مقدرات الموقع والتشتت لطريقة **(MVE)** ، الا ان اعتمدنا على الخوازمية التقريبية والتي يمكن ايجازها في عدد من الخطوات :

١. سحب جميع العينات الجزئية الممكنة **J** بحجم **(P+1)** من المشاهدات وفق التوافيق :

$$\binom{\mathbf{n}}{\mathbf{P+1}} \qquad J = 1, 2, \ldots\ldots.., P+1 \qquad (4)$$

٢. حساب مسافات مهلونوبس لمشاهدات العينة بأكملها بالاعتماد على مقدرات الموقع والتشتت لكل عينة جزئية وفق الصيغة الاتية :

$$MD_i = \sqrt{\left(x_i - \underline{\hat{\mu}_i}\right)\ (cv)^{-1}\left(x_i - \underline{\hat{\mu}_i}\right)'} \qquad (5)$$

حيث ان : $\underline{\hat{\mu}_i}$ يمثل مقدر الموقع و هو موجه الوسط الحسابي لكل عينة جزئية

cv تمثل مقدر التشتت وهي مصفوفة التباين والتباين المشترك لكل عينة جزئية

٣. نعمل على ترتيب قيم مسافات مهلونوبس ترتيبا تصاعديا والقيمة التي تمثل تسلسل $\mathbf{h} = \frac{n+p+1}{2}$ تعرف بعامل التعظيم الصحيح $(\mathbf{m_j})$ الذي يستخدم في تعظيم حجم العينة الجزئية ليحتوي على **(h)** من المشاهدات .

٤. حساب حجم المجسم للقطع الناقص لجميع العينات الجزئية وفق الصيغة الاتية :

$$\text{volume} = \sqrt{det\ (m_j{}^2\ cv)} = m_j{}^p\ det(cv) \qquad (6)$$

٥. العينة الجزئية التي تمتلك اقل حجم مجسم قطع ناقص من بين جميع العينات الجزئية وفق المعادلة **(6)** تعتبر العينة الجزئية المثلى ، ثم نعمل على اعادة الحساب عدة مرات مع الاحتفاظ بأصغر حجم قطع في كل مرة . يعتبر متجه الوسط الحسابي لاصغر حجم قطع ناقص من بين تلك القطوعات هو مقدر الموقع لطريقة **RMD(MVE)** .

٦. حساب مصفوفة التباين المشترك من العينة الجزئية التي تمتلك اصغر حجم مجسم قطع ناقص وفق الصيغة الاتية :

$$\Sigma = (\chi^2{}_{p,0.5})^{-1} * C_{n,p}{}^2 * m_j * cv \qquad (7)$$

حيث ان :

$(\chi^2{}_{p,0.5})$: يعتبر ثابت تحدد قيمته وفقا الى التوزيع المفترض لأغلب البيانات ، ويمثل وسيط توزيع χ^2 بدرجة حرية **p**.

$C_{n,p}{}^2$: يمثل عامل تصحيح مناسب يوفر الاتساق لمصفوفة التباين المشترك ، يستخدم للعينات الصغيرة بالاعتماد على حجم العينة **(n)** وابعادها **(p)** ،والذي استخدم في بعض الدراسات عندما يكون **(P ≤ 8)** : [9][5]

$$C_{n,p}{}^2 = \left(1 + \frac{15}{n-p}\right)^2$$

٧. تعويض مقدرات الوسط الحسابي لأصغر حجم مجسم قطع ناقص وايضا مصفوفة التباين المشترك في المعادلة **(5)** بدل الوسط الحسابي ومصفوفة التباين المشترك في صيغة مهلونوبس التقليدية ، عند ذلك نحصل على صيغة مسافة مهلونوبس الحصينة **RMD(MVE)** .

بعد حساب مسافات مهلونوبس الحصينة نستطيع معرفة القيم الشاذة في حالة متعدد المتغيرات عن طريق اختبارها بنقطة قطع مناسبة ، وعلى هذا الاساس فان اي مسافة مهلونوبس حصينة تزداد عن قيمة القطع تعتبر قيمة شاذة متطرفة في المتغير التوضيحي **(HLP)**، تعرف نقطة القطع لمسافات مهلونوبس الحصينة بـ **(cd)** و تحسب بالصيغة الاتية: **[11][8][9]**

$$cd = median(RMD_i) + 3MAD(RMD_i) \qquad (8)$$

حيث ان :

(RMD_i) : مسافات مهلونوبس الحصينة

$median(RMD_i)$: قيمة الوسيط لمسافات مهلونوبس الحصينة .

$MAD(RMD_i)$: الوسيط لمطلق انحرافات مسافات مهلونوبس عن وسيطها ويحسب :

$$MAD(RMD_i) = median\left|(RMD_i) - median(RMD_i)\right|$$

الاوزان الحصينة التي سوف نحصل عليها من خلال طريقة **RMD(MVE)** تعطى وفق الصيغة الاتية:

$$w_{ir} = min\left(1, \frac{cd}{RMD_i}\right) \qquad (9)$$

القيم الشاذة المتطرفة في فضاء **X** او ما تعرف ايضا بنقاط الرافعة العليا **(HLPs)** سوف تأخذ الوزن $\left(\frac{cd}{RMD_i}\right)$ ، اما القيم الاعتيادية الاخرى فأنها تأخذ الوزن (١) **[8]**.

2-4 مسافات مهلونبس الحصينة على اساس اصغر محدد مصفوفة تباين مشترك Robust Mahalanobis Distance Based on Minimum Covariance Determinant RMD(MCD)

تعتبر طريقة مسافات مهلونوبس التي تستند الى اصغر محدد مصفوفة تباين مشترك **RMD(MCD)** هي الطريقة البديلة لـ **RMD(MVE)** ، والتي تمتلك مقدرات لها نفس نقطة الانهيار ، وبذلك تعتبر مقدرات اصغر محدد مصفوفة تباين مشترك **(Minimum Covariance Determinant) (MCD)** من ضمن اهم المقدرات الحصينة في تقدير متجه الوسط الحسابي ومصفوفة التباين المشترك في البيانات متعددة المتغيرات ، تستند فكرة طريقة **RMD(MCD)** على ايجاد جميع العينات الجزئية التي تحتوي على الاقل نصف المشاهدات تقريبا **(h)** ، وايجاد العينة الجزئية التي لها اصغر محدد مصفوفة تباين مشترك من بين العينات الجزئية ، والاعتماد على مقدراتها لكل من الموقع والتشتت حسابيا في ايجاد مسافات مهلونوبس ثم الكشف عن القيم الشاذة **[6]**.

ويمكن تلخيص أيجاد مقدرات **(MCD)** وتوضيح عملها في الكشف عن المشاهدات الشاذة في متغيرات X في الخوارزمية التالية :

١. سحب جميع العينات الجزئية الممكنة **J** بحجم **(h)** من المشاهدات من خلال ايجاد التوافيق التالي :

$$\binom{n}{h} = \frac{n!}{h!\,(n-h)!} \qquad J = 1, 2, \ldots\ldots, h \qquad (10)$$

وتشير **h** تقريبا الى نصف البيانات والتي تساوي :

$$h = \frac{n+p+1}{2} \qquad (11)$$

٢. ايجاد افضل عينة جزئية **J** وهي العينة الجزئية التي تمتلك اصغر محدد مصفوفة تباين مشترك من بين جميع العينات وتحقق :

$$\mathcal{J} = arg_J\ min\left|cv\right| \qquad (12)$$

حيث ان: cv تمثل مصفوفة التباين المشترك للعينة الجزئية

٣. بعد ايجاد افضل عينة جزئية تعتبر مقدراتها لكل من موجه الوسط الحسابي ومصفوفة التباين المشترك بعد ضربها بعامل الاتساق $(C_{n,p}{}^2)$ كما في طريقة **RMD(MVE)** هي مقدرات اصغر محددة مصفوفة تباين مشترك **(MCD)** ، من ثم نعمل على ايجاد مسافات مهلونوبس التربيعية لجميع مشاهدات العينة بالاعتماد على تلك المقدرات وفق الصيغة الاتية: **[10]**

$$RMD_{iMCD}{}^2 = \left(x_i - \underline{\hat{\mu}}_{j_{MCD}}\right)\,(cv_{MCD})^{-1}\left(x_i - \underline{\hat{\mu}}_{j_{MCD}}\right)' \qquad (13)$$

٤. لأيجاد القيم الشاذة المتطرفة (HLPs) سوف نختبر المسافات الحصينة المحسوبة وفق نقطة القطع المستخدمة في طريقة **RMD(MVE)** في المعادلة **(8)**. وبذلك ان اي مسافة حصينة تزداد عن قيمة القطع سوف تعتبر قيمة شاذة متطرفة في المتغير التوضيحي **(HLP)** وخلاف ذلك فأنها اعتيادية . الاوزان الحصينة التي نحصل عليها من خلال طريقة **RMD(MCD)** سوف تكون وفق المعادلة الاتية :

$$w_{im} = min\left(1, \frac{cd}{RMD_i}\right) \qquad (14)$$

المشاهدات التي تقابل القيم الشاذة المتطرفة في المتغيرات التوضيحية **(HLPs)** سوف تأخذ الوزن ($\frac{cd}{RMD_i}$) ، اما المشاهدات الاعتيادية الاخرى فأنها تأخذ الوزن (١) . **[8]**.

5 - تقدير مصفوفة التباين والتباين المشترك المتسقة الحصينة لمتجه المعلمات المقدرة لأنموذج الانحدار الخطي العام

Robust Heteroscedastic Consistent Covariance Matrix Estimate for Estimate of the Regression linear model parameters (RHCCM)

اقترح **(white)[12]** اول مقدر متسق لمصفوفة التباين المشترك يعتمد على تقديرات **OLS** عند وجود مشكلة عدم تجانس التباين ، اطلق عليه بمقدر مصفوفة التباين المشترك المتسقة ، و يعرف بالصيغة الاتية :

$$\mathbf{HC_0} = (X'\mathbf{X})^{-1}X'\hat{\boldsymbol{\psi}}_0\,\mathbf{X}\,(X'\mathbf{X})^{-1} \qquad (15)$$

صيغة $\mathbf{HC_0}$ تضمنت البواقي التربيعية $\mathbf{e_i}^2$ لطريقة المربعات الصغرى بدلا من $\boldsymbol{\sigma^2 i}$ ، اي ان $\hat{\boldsymbol{\psi}}_0$ تساوي :

$$\hat{\boldsymbol{\psi}}_0 = diag\,(\mathbf{e_i}^2) \qquad i = 1, 2, \dots\dots n \qquad (16)$$

يعتبر مقدر $\mathbf{HC_0}$ متسق تحت كلا الحالتين التجانس وعدم التجانس للتباين ، الا انه يعاب عليه بأنه مقدر متحيز في العينات المحدودة ، وانه يفشل في تقدير تباينات يعتمد عليها عندما تتضمن مصفوفة المتغيرات المستقلة قيم شاذة متطرفة **(HLPs)** **[2]** .

بعد ذلك اقترحت المقدرات التالية $(\mathbf{HC_3}, \mathbf{HC_2}, \mathbf{HC_1})$ ، من ثم اقترح **(Cribari–Neto)[3]** مقدر $(\mathbf{HC_4})$ الذي بني على اساس مقدر $\mathbf{HC_3}$ ، يمتاز مقدر $\mathbf{HC_4}$ بأنه يأخذ بالحسبان النسبة بين قياس قوة الرفعة $\mathbf{h_i}$ ومتوسطها ، يعتبر مقدر $\mathbf{HC_5}$ مقترح جديد من قبل **(Cribari–Neto)** واخرون **[2]** ، هذا المقدر يأخذ في الاعتبار الحد الاقصى لقوة الرفعة اضافة الى مميزات $\mathbf{HC_4}$.الا ان وجود مشكلة عدم تجانس التباين و القيم الشاذة المتطرفة في المتغير التوضيحي **(HLPs)** سويا يجعل من تقديرات معلمات طريقة المربعات الصغرى متحيزة ، وبالتالي يصبح الاستدلال غير واقعي .

(Furno)[4]اقترح مصفوفة التباين والتباين المشترك المتسقة الحصينة **(RHCCM)** في حالة وجود مشكلة عدم تجانس التباين و القيم الشاذة المتطرفة **(HLPs)** ، اعتمد فيها الانحدار الحصين الذي يسمح في السيطرة على تأثيرات القيم الشاذة على التقديرات وبالتالي الحد من التحيز الحاصل **.[4][8]**

فكرة **(Furno)** تكمن في استخدام بواقي المربعات الصغرى الموزونة **(WLS)** بدلا من بواقي المربعات الصغرى الاعتيادية **OLS** . اعتمدنا في دراستنا على كل من المقدرين $\mathbf{HC_4}$ و $\mathbf{HC_5}$ وذلك لكونهما يأخذان في الحسبان وجود القيم الشاذة المتطرفة في المتغير التوضيحي **(HLPs)** والاستناد الى طريقة **(Furno)** الموزونة في حساب مصفوفة التباين والتباين المشترك المتسقة الحصينة لمعلمات الانموذج المقدرة وحسب الصيغ الاتية وعلى التوالي :

$$\mathbf{HC_{4W}} = (\mathbf{X'WX})^{-1}\mathbf{X'}W\hat{\boldsymbol{\psi}}_{4w}\,\mathbf{WX}\,(\mathbf{X'}W\mathbf{X})^{-1} \qquad (17)$$

حيث ان : $\mathbf{W}$ مصفوفة مربعة قطرية عناصر قطرها $\boldsymbol{wi}$

$$\hat{\boldsymbol{\psi}}_4 = diag\left\{\frac{\mathbf{e_i}^2}{(1 - h_i{}^*)^{\delta i}}\right\} \qquad i = 1, 2, \dots\dots n$$

$$\delta i = min\left\{4, \frac{h_i{}^*}{\bar{h}^*}\right\} \qquad , \bar{h}^* = \frac{\sum h_i{}^*}{n}$$

$$\mathbf{HC_{5W}} = (\mathbf{X'WX})^{-1}\mathbf{X'W}\hat{\boldsymbol{\psi}}_{5w}\,\mathbf{WX}\,(\mathbf{X'WX})^{-1} \qquad (18)$$

حيث ان :

$$\widehat{\Psi}_5 = diag\left\{\frac{e_i^{\,2}}{\sqrt{(1-h_i^{*})^{\alpha i}}}\right\} \qquad i = 1,2,\dots\dots..n$$

$$\alpha i = min\left\{\frac{h_i^{*}}{\bar{h}^{*}}, max\left\{4, \frac{kh^{*}_{max}}{\bar{h}^{*}}\right\}\right\}$$

$$h_i^{*} = \sqrt{W}\, x_i\, (X'WX)^{-1} x_i'\, \sqrt{W}$$

حيث h_i^{*} تمثل العناصر القطرية لمصفوفة القبعة المرجحة والتي تحسب وفق الصيغة الاتية :

$$(19) H_W = \sqrt{W}X\,(X'WX)^{-1}X'\sqrt{W}$$

وان k ثابت محدد مسبقا بين (٠ < k < ١)، اعتمده اغلب الباحثين مساوي الى (٠،٧).[2]

٥ - الجانب التجريبي

تم الاعتماد على اسلوب المحاكاة (Simulation) وفق طريقة (Monte Carlo) في الحصول على تجارب عديدة وبأحجام عينات مختلفة وكذلك بقيم مختلفة لحدة عدم تجانس التباين وايضا بنسب مختلفة للقيم الشاذة المتطرفة (HLPs) ، تمت المقارنة بين كل من الطريقتان RMD(MVE) RMD(MCD) وفق معيار الخطاء المعياري (SE) من خلال مصفوفة التباين والتباين المشترك المتسقة الحصينة (RHCCM) ولكل من مقدر (HC_{W4}) (HC_{W5}) .

تخلل جانب المحاكاة عدد من الخطوات في توليد البيانات هي :

١) توليد المتغيرات المستقلة (x_1, x_2, x_3, x_4) وفق التوزيع الطبيعي القياسي وحسب الصيغة الاتية:

$X \sim N(0,1)$

٢) معلمات الانموذج الاصلي افترضت بانها مساوية الى القيم الاتية :

$$B_0 = 1.5\ , \qquad B_1 = B_4 = 1\ , \qquad B_2 = 0.6\ , \qquad B_3 = 0.4$$

٣) توليد الاخطاء العشوائية وفق التوزيع الطبيعي وحسب الصيغة الاتية :

$$\epsilon_i \sim (o, \sigma^2{}_i) \qquad , i = 1,2,\dots,n$$

حيث ان :

$$\sigma^2{}_i = \exp\{c * x_{i1}\}$$

حيث $(\sigma^2{}_i)$ تمثل دالة توليد عدم تجانس تباين الاخطاء . تعتمد حدة (قوة)عدم تجانس التباين على مقدار قيمة ثابت الاضطراب (C) فاذا كانت قيمة $(C=0)$ يؤدي ذلك الى تجانس التباين وبزيادة مقدار (C) تزداد حدة عدم تجانس التباين ، تجارب المحاكاة في دراستنا سوف تأخذ درجتين مختلفتين لحدة عدم تجانس التباين هما منخفضة و مرتفعة بالاعتماد على قيمة (C) ، عندما نعتمد قيمة (C=0.20) نحصل على حدة عدم تجانس تباين منخفضة ، اما اذا كانت قيمة (C=0.40) تكون حدة عدم تجانس التباين مرتفعة ، وتحسب حدة عدم تجانس التباين وفق الاتي: [4]

$$\lambda = \frac{\max(\sigma^2{}_i)}{min(\sigma^2{}_i)} \qquad , i = 1,2\dots,n \qquad (20)$$

٤) حساب المتغير المعتمد من خلال ضرب مصفوفة المتغيرات المستقلة مع متجه المعلمات الافتراضية مضافة اليهم قيمة حد الخطأ العشوائي .

٥) استبدال بعض المشاهدات في المتغيرات المستقلة بشكل عشوائي بقيم شاذة متطرفة (HLPs) وبنسب محددة هي (10% , 20%) من حجم العينة الكلي ، وان القيم الشاذة المتطرفة (HLPs) تتبع التوزيع الطبيعي وحسب الصيغة الاتية : [8]

$X \sim N(15,1)$

1-5 نتائج تجارب المحاكاة Results of the Simulation Experiment

جدول رقم (1) يبين تقديرات المربعات الصغرى الموزونة الحصينة بأستخدام الطرائق الحصينة RMD(MVE) ,RMD(MCD) ومعيار الخطأ المعياري (SE) للمقدرين (HC_{W4}) و (HC_{W5}) عندما تكون (n=20) ونسبة القيم الشاذة المتطرفة (HLP=10%)

Method	حدة عدم تجانس التباين عندما (C=0.20)			حدة عدم تجانس التباين عندما (C=0.40)		
	Estimate	$Se.HC_{W4}$	$Se.HC_{W5}$	Estimate	$Se.HC_{W4}$	$Se.HC_{W5}$

b0.MVE	1.497326	0.245554	0.228445	1.494678	**0.246826***	**0.229863***
b1. MVE	0.99556	0.294905	0.25825	0.997127	**0.278882***	**0.251313***
b2. MVE	0.59998	0.181568	**0.146294***	0.602775	**0.212116***	**0.171326***
b3. MVE	0.397208	0.319033	0.289941	0.395423	**0.235999***	**0.202655***
b4. MVE	1.004374	0.218792	0.200603	0.999096	**0.142227***	**0.125702***
b0.MCD	1.496641	**0.243635***	**0.22748***	1.494106	0.269439	0.246726
b1. MCD	0.997125	**0.288024***	**0.24387***	0.998053	0.337993	0.306565
b2. MCD	0.599774	**0.17732***	0.146878	0.602817	0.229089	0.175188
b3. MCD	0.399197	**0.309011***	**0.26831***	0.395255	0.292064	0.249591
b4. MCD	1.003685	**0.190818***	**0.174063***	0.997681	0.198848	0.181461

جدول رقم (2) يبين تقديرات المربعات الصغرى الموزونة الحصينة بأستخدام الطرائق الحصينة RMD(MVE) ومعيار الخطأ المعياري (SE) للمقدرين (HCw4) و (HCw5) عندما تكون (n=20) ونسبة القيم الشاذة المتطرفة (HLP=20%)،RMD(MCD)

Method	حدة عدم تجانس التباين عندما (C=0.20)			حدة عدم تجانس التباين عندما (C=0.40)		
	Estimate	Se.HCw4	Se.HCw5	Estimate	Se.HCw4	Se.HCw5
b0.MVE	1.496685	**0.259138***	**0.238697***	1.504231	0.365918	**0.312665***
b1. MVE	0.998332	**0.369367***	0.319415	1.003176	0.59634	0.459145
b2. MVE	0.603239	**0.276461***	**0.237511***	0.596368	**0.193681***	**0.165814***
b3. MVE	0.397544	**0.310978***	0.274788	0.398869	0.24917	0.220099
b4. MVE	1.00038	**0.132039***	0.115155	1.001247	0.197266	0.172988
b0.MCD	1.496743	0.261333	0.239528	1.503909	**0.365621***	0.313121
b1. MCD	0.99882	0.376752	**0.306214***	1.003485	**0.577461***	**0.450268***
b2. MCD	0.603195	0.2789	0.24328	0.596755	0.195831	0.167328
b3. MCD	0.39806	0.316207	**0.254579***	0.399014	**0.240293***	**0.208748***
b4. MCD	1.000362	0.136361	**0.110262***	1.000995	**0.163519***	**0.136607***

تشير علامة (*) الى اقل قيمة لمعيار المقارنة الخطأ المعياري SE للمعلمة المقدرة مقارنة مع نظيراتها من الطرائق الاخرى .

جدول رقم (3) يبين تقديرات المربعات الصغرى الموزونة الحصينة بأستخدام الطرائق الحصينة RMD(MVE) ومعيار الخطأ المعياري (SE) للمقدرين (HCw4) و (HCw5) عندما تكون (n=60) ونسبة القيم الشاذة المتطرفة (HLP=10%)،RMD(MCD)

Method	حدة عدم تجانس التباين عندما (C=0.20)	حدة عدم تجانس التباين عندما (C=040)

	Estimate	Se.HCw4	Se.HCw5	Estimate	Se.HCw4	Se.HCw5
b0.MVE	1.499143	**0.137921***	0.134204	1.500977	0.142953	0.138849
b1. MVE	0.999865	0.157976	0.149843	0.998663	0.172952	0.16484
b2. MVE	0.599856	0.10797	0.10394	0.601184	0.080867	0.077666
b3. MVE	0.398193	**0.160871***	0.153004	0.396231	0.146689	0.138916
b4. MVE	0.999366	**0.090415***	**0.0874***	1.00053	0.089471	0.086473
b0.MCD	1.499193	0.137934	**0.134144***	1.500957	**0.142707***	**0.138762***
b1. MCD	0.999863	**0.15787***	**0.149687***	0.99871	**0.172664***	**0.164634***
b2. MCD	0.599806	**0.105999***	**0.101582***	0.601037	**0.078327***	**0.075237***
b3. MCD	0.398184	0.161085	**0.152687***	0.396227	**0.146198***	**0.138681***
b4. MCD	0.999349	0.090757	0.087687	1.000418	**0.084612***	**0.081709***

جدول رقم (4) يبين تقديرات المربعات الصغرى الموزونة الحصينة بأستخدام الطرائق الحصينة RMD(MVE) ,RMD(MCD)ومعيار الخطأ المعياري (SE) للمقدرين (HCw4) و (HCw5) عندما تكون (n=60) ونسبة القيم الشاذة المتطرفة (HLP=20%)

Method	حدة عدم تجانس التباين عندما (C=0.20)			حدة عدم تجانس التباين عندما (C=040)		
	Estimate	Se.HCw4	Se.HCw5	Estimate	Se.HCw4	Se.HCw5
b0.MVE	1.501682	0.141809	0.138983	1.499935	0.171703	0.1666
b1. MVE	0.999156	**0.116706***	**0.110942***	0.996355	0.318095	0.294171
b2. MVE	0.599728	0.08514	0.082211	0.600398	0.096372	0.092987
b3. MVE	0.40079	**0.079883***	**0.076731***	0.401312	0.099919	0.096164
b4. MVE	1.000487	0.08735	0.083784	0.999882	0.100055	0.095996
b0.MCD	1.501698	**0.14148***	**0.138726***	1.500057	**0.16743***	**0.163118***
b1. MCD	0.999123	0.117302	0.111031	0.996859	**0.292018***	**0.273731***
b2. MCD	0.599772	**0.082525***	**0.079616***	0.600395	**0.092834***	**0.089879***
b3. MCD	0.400727	0.080418	0.077417	0.401188	**0.089595***	**0.086094***
b4. MCD	1.0004	**0.084595***	**0.08095***	0.999899	**0.097909***	**0.093976***

تشير علامة (*) الى اقل قيمة لمعيار المقارنة الخطأ المعياري SE للمعلمة المقدرة مقارنة مع نظيراتها من الطرائق الاخرى .

جدول رقم (5) يبين تقديرات المربعات الصغرى الموزونة الحصينة بأستخدام الطرائق الحصينة RMD(MVE) ,RMD(MCD)ومعيار الخطأ المعياري (SE) للمقدرين (HCw4) و (HCw5) عندما تكون (n=100) ونسبة القيم الشاذة المتطرفة (HLP=10%)

Method	حدة عدم تجانس التباين عندما (C=0.20)			حدة عدم تجانس التباين عندما (C=0.40)		
	Estimate	Se.HCw4	Se.HCw5	Estimate	Se.HCw4	Se.HCw5
b0.MVE	1.500891	0.115489	0.113597	1.254403	**0.100049***	**0.098144***
b1. MVE	1.000894	0.114751	0.111664	0.836199	**0.100913***	**0.098151***
b2. MVE	0.599538	0.063487	0.062242	0.501297	0.054155	0.053034
b3. MVE	0.398716	0.114551	0.111785	0.334133	0.087797	0.085566
b4. MVE	1.000519	0.06886	0.067042	0.835571	**0.058059***	**0.056781***
b0.MCD	1.500893	**0.115407***	**0.11357***	1.254399	**0.100049***	0.09815
b1. MCD	1.00089	**0.114669***	**0.111636***	0.836198	0.100915	0.098159
b2. MCD	0.599608	**0.058054***	**0.056903***	0.501313	**0.052665***	**0.051577***
b3. MCD	0.398711	**0.114474***	**0.111757***	0.33413	**0.08777***	**0.085545***
b4. MCD	1.000499	**0.066542***	**0.064814***	0.83557	0.058628	0.05734

جدول رقم (6) يبين تقديرات المربعات الصغرى الموزونة الحصينة بأستخدام الطرائق الحصينة RMD(MVE) ,RMD(MCD)ومعيار الخطأ المعياري (SE) للمقدرين (HCw4) و (HCw5) عندما تكون (n=100) ونسبة القيم الشاذة المتطرفة (HLP=20%)

Method	حدة عدم تجانس التباين عندما (C=0.20)			حدة عدم تجانس التباين عندما (C=0.40)		
	Estimate	Se.HCw4	Se.HCw5	Estimate	Se.HCw4	Se.HCw5
b0.MVE	1.500266	0.108837	0.107699	1.499623	**0.117228***	**0.116022***
b1. MVE	0.999009	0.107841	0.104998	1.002056	**0.268309***	**0.25665***
b2. MVE	0.599793	0.069904	0.068533	0.600159	0.101169	0.098202
b3. MVE	0.39974	0.066559	0.065181	0.399751	**0.083246***	**0.081473***
b4. MVE	1.000183	0.067978	0.066582	1.001074	**0.072795***	**0.07094***
b0.MCD	1.50028	**0.108753***	**0.107634***	1.499628	0.117669	0.116465
b1. MCD	0.998969	**0.107006***	**0.103848***	1.002115	0.277932	0.26502
b2. MCD	0.599776	**0.069677***	**0.068302***	0.600131	**0.100832***	**0.097683***
b3. MCD	0.399705	**0.065639***	**0.064255***	0.399749	0.083825	0.081984
b4. MCD	1.000195	**0.067018***	**0.065641***	1.001106	0.074852	0.072919

تشير علامة (*) الى اقل قيمة لمعيار المقارنة الخطأ المعياري SE للمعلمة المقدرة مقارنة مع نظيراتها من الطرائق الاخرى .

2-5 مناقشة نتائج تجارب المحاكاة

discuss the results of the simulation experiment

من خلال النتائج المبينة في الجداول (١)(٢)(٣)(٤)(٥)(٦) نلاحظ ما يلي :

١. في حالة وجود نسبة (HLP_S=10%) في البيانات مع اختلاف حدة عدم تجانس التباين نلاحظ ان طريقة RMD(MVE) هي الافضل وفق معيار الخطأ المعياري SE للمقدرات (HC_{W4}) (HC_{W5}) عندما تكون حجم العينة (n=20) ، الا ان مع زيادة حجوم العينات نجد ان طريقة RMD(MCD) هي الافضل عندما تكون (n=60,100) .
٢. اما في حالة وجود نسبة (HLP_S=20%) في البيانات مع اختلاف حدة عدم تجانس التباين نجد ان كل من الطريقتان RMD(MVE) RMD(MCD) لهما نفس الافضلية وفق معيار الخطأ المعياري SE لمتجه المعلمات المقدرة لانموذج الانحدار الخطي العام ولكل من المقدرات (HC_{W4}) و (HC_{W5}) عندما تكون حجم العينة (n=20) ، وتظهر افضلية طريقة RMD(MCD) عندما تكون (n=60,100) .
٣. كما تبين نتائج تجارب المحاكاة ان اعتماد كل من مقدر (HC_{W4}) و (HC_{W5}) على طريقة RMD(MCD) في اكتشاف (HLP_S) وتحديد الاوزان الحصينة هي الافضل وفق معيار الخطأ المعياري SE مع اختلاف حجوم العينات و حدة عدم تجانس التباين ونسب (HLP_S) في البيانات.

٦ ـ الاستنتاجات والتوصيات

1-6 الاستنتاجات

١. اظهرت نتائج تجارب المحاكاة ان طريقة RMD(MCD) هي الافضل وفق معيار الخطأ المعياري (SE) لمتجه المعلمات المقدرة لانموذج الانحدار ومن خلال كل من مقدر (HC_{W4}) و (HC_{W5})
٢. تبين ان طريقة RMD(MCD) اكثر كفاءة عند زيادة حجوم العينات مع اختلاف نسب (HLP_S) في البيانات وحدة عدم تجانس التباين .

2-6 التوصيات

١. اعتماد طريقة RMD(MCD) في تقدير معلمات انموذج الانحدار الخطي وتقدير الاوزان الحصينة في حالة حجوم العينات الكبيرة عند وجود (HLP_S) في البيانات .
٢. استعمال طريقة RMD(MVE) في حالة حجوم العينات الصغيرة عند وجود (HLP_S) في البيانات.

7 ـ المصادر

7 -1 المصادر العربية

١. كاظم ، أموري هادي ، الدليمي ، محمد مناجد (١٩٨٨) " مقدمة في تحليل الانحدار الخطي " جامعة الموصل – مديرية الكتب للطباعة والنشر .

7 - 2 المصادر الاجنبية

[2]- Cribari-Neto, F., Souza, T.C.& Vasconcellos, K.L.P. "Inference under heteroskedasticity and leveraged data" . Theory Methods , vol (36) ,pp. 1877–1888 (2007). Errata :vol (37),pp. 3329–3330 (2008)

[3]- Cribari-Neto, F. (2004). "Asymptotic inference under heteroskedasticity of unknown form" .Computational Statistics and Data Analysis, vol (45), pp. 215–233.

[4]- Furno, M. (1996) " Small sample behavior of a robust heteroskedasticity consistent covariance matrix estimator" Journal of Statistical Computation and Simulation, vol (54), pp. 115-128.

[5]- Hadi , A.S (1992) "Identifying Multiple Outliers in Multivariate Data" journal of the royal ststistical society , vol (54) ,pp. 761-771

[6]- Jensenl ,W.A,. Birch ,J.B,. & Woodall ,W.H. (2006)" High Breakdown Estimation Methods for Phase I Multivariate Control Charts" Qual. Reliab. Engng. Int. Vol (23) ,pp.615–629.

[7]- Lim, H.A., Midi , H. (2016) "Diagnostic Robust Generalized Potential Based on Index Set Equality (DRGP (ISE)) for the identification of high leverage points in linear model", Comput Stat, Vol 31, pp. 859–877

[8]- Midi, H ., Sani ,M ., & Arasan , J. (2018) " Robust Heteroscedasticity Consistent Covariance Matrix Estimator based on Robust Mahalanobis Distance and Diagnostic Robust Generalized Potential Weighting Methods in Linear Regression" Journal of Modern Applied Statistical Methods. Vol (17) iss.1

[9]- Rousseeuw, P.J ., Van Zomeren , B.C (1990) " Unmasking multivariate outliers and leverage points ", J Am Stat Assoc, vol 85(411), PP. 633–639

[10]- Rousseeuw, P.J ., Hubert, M.,(2018) "Anomaly detection by robust Statistics", Wiley Periodicals, Inc. vol (8) ,pp.1-14.

[11]- Rousseeuw, P.J., Aelst, S.F.(2009) " Minimum volume ellipsoid " John Wiley & Sons, Inc. vol (1) , pp. 71–82

[12]- White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". Econometrica, vol 48(4),pp. 817-838.