دراسة مقارنة بين بعض الطرائق الحصينة في تقدير معلمات انموذج الانحدار الخطي باستخدام اسلوب المحاكاة التجريبي
في حالة وجود بيانات تتضمن مشاهدات شاذة

م. م. زياد زكي صالح
كلية الزراعة/ جامعة بغداد

الخلاصة

في الانحدار الخطي، الشواذ هي مشاهدات تمتلك بواقي كبيرة، بصيغة اخرى المشاهدات التي يكون فيها قيمة المتغير المعتمد غير عادية عن القيمة التقديرية للمتغيرات المتوقعة، والمشاهدة الشاذة يمكن ان تظهر عن طريق بيانات يتم ادخالها عن طريق الخطأ او عن طريق مشكلة اخرى.

المشاهدة مع قيم متطرفة على المتغير المتوقع تكون نقطة ذات قوة تباعد عالية، وقوة التباعد العالية هي قياس لبعد المسافة لانحراف المتغير المستقل عن وسطه الحسابي، حيث ان النقاط ذات قوة التباعد العالية ممكن ان تمتلك تأثيرا على تقديرات معلمات الانحدار .

التقدير الحصين لمعلمات الانحدار يتعامل مع حالات تمتلك قوة تباعد عالية والحالات الشاذة، التقدير الحصين جوهريا يحاول ان يساوي مابين الحالات التي تحاول تعديل الشواذ وعدم انتهاك الفروض لتقدير انحدار OLS، اي انه صيغة من المربعات الصغرى الموزونة تعامل بشكل تكراري وبكل خطوة يكون لدينا مجموعة اوزان تحدد على ضوء البواقي، بشكل عام تتناسب البواقي عكسيا مع الاوزان فكلما كبرت البواقي قلت الاوزان، لذلك الاوزان تعتمد على البواقي، والبواقي تعتمد على النموذج والنموذج يعتمد على الاوزان.

وبواسطة استخدام اسلوب المحاكاة التجريبي مع بيانات تولد من نموذج خطي مقترح وعن طريق جعل بعض البيانات لتكون مشاهدات شاذة، تم اجراء المقارنة بين ثلاثة طرائق مقدرات حصينة لدراسة الاختلافات بين طرائق التقدير المذكورة في عدة حالات وبظروف مختلفة .

Abstract

In linear regression, an outlier is an observation with large residual. In other words, it is an observation whose dependent-variable value is unusual given its values on the predictor variables. An outlier observation may indicate a data entry error or other problem .

An observation with an extreme value on a predictor variable is a point with high leverage. Leverage is a measure of how far an independent variable deviates from its mean. These leverage points can have an effect on the estimate of regression coefficients .

Robust estimation for regression parameters deals with cases that have very high leverage, and cases that are outliers. Robust estimation is essentially a compromise between dropping the case(s) that are moderate outliers and seriously violating the assumptions of OLS regression. It is a form of weighted least squares regression and is done iteratively. At each step a new set of weights are determined based on the residuals. In general, the larger the residuals, the smaller the weights. So the weights depend on residuals. At the same time, the residuals depend on the model and the model depends on the weights .

By using empirical simulation approach with data generated from suggesting linear model and by making some of data points to be outlier observations, the comparisons was made between three robust estimation methods to study the differences in many cases and conditions between these estimation methods .

المقدمة: Introduction

تقديرات معلمات الانحدار الخطي بطريقة المربعات الصغرى الاعتيادية تعد افضل التقديرات واقربها الى الواقع لتمثيل البيانات على شكل انموذج انحدار خطي لكن ذلك يكون حسب شروط معروفة يجب ان تتوافر حتى نحصل على افضل التقديرات بهذه الطريقة وفي حالة عدم توافر احد الشروط الاساسية للتقدير بطريقة OLS لمعلمات الانحدار الخطي سوف يؤدي ذلك الى التأثير في جودة التقديرات لمعلمات انموذج الانحدار ، ومن ضمن المشاكل التي تؤثر في احدى الشروط الاساسية لتقديرات OLS هي المشاهدات الشاذة في البيانات حيث ان المشاهدات الشاذة يمكن ان تغير القيم التقديرية لمعلمات انموذج الانحدار اذا ما تم حذفها او ازالتها من البيانات (بكلام اخر ان المشاهدات الشاذة هي مشاهدات مؤثرة في تقدير المعلمات) اي ان المشاهدات الشاذة يمكن ان تجعل تقديرات طريقة المربعات الصغرى الاعتيادية غير دقيقة وبعيدة عن الواقع، لذلك يتم اللجوء الى الطرائق الحصينة في تقدير معلمات انموذج الانحدار الخطي والتي تعالج مشكلة المشاهدات الشاذة من دون ان تحذفها حيث تعمل الطرق الحصينة على التقليل من تأثيرات المشاهدات الشاذة على تقدير المعلمات وهناك طرائق تقدير حصينة كثيرة تهتم في الكشف عن الشواذ اولا ثم معالجة المشاهدات الشاذة ثانيا ومن هذه الطرائق الحصينة التي سوف نتطرق اليها هي (الطريقة الحصينة باستخدام مصفوفة التغاير الحصينة لـ كامبل Campbell's Robust covariance matrix، الطريقة الحصينة باستخدام الوسيط للانحراف المطلق لـ هامبل Hamplel's median absolute deviation، الطريقة الحصينة لتحديد المشاهدات الشاذة باستخدام صيغة المسافة لكوك Cook's distance) .

مشكلة البحث : The Search Problem

المشاهدات الشاذة لها تأثير كبير على تقدير معلمات انموذج الانحدار الخطي حيث اذا ما تم استخدام طريقة المربعات الصغرى الاعتيادية في التقدير في ظل وجود المشاهدات الشاذة سوف نحصل على نتائج تقدير لمعلمات غير دقيقة وبعيدة عن الواقع التي تتصف به المشاهدات العامة، لذلك فأن الغرض الاساسي من الانحدار الحصين هو توفير نتائج لتقدير معلمات تقاوم (تثبت) بوجود الشواذ، ولغرض انجاز هذا الثبات، التقدير الحصين يحد من تأثير الشواذ في البيانات على تقدير المعلمات، من الناحية التاريخية يوجد ثلاثة اصناف من المشكلات التي تتعامل مع تقنيات الانحدار الحصين وهي :

- مشاكل مع الشواذ في المتغير المعتمد Dependent variable .
- مشاكل مع الشواذ المتعددة لمجموعة من المتغيرات المستقلة Independent variables .
- مشاكل مع الشواذ في كل من المتغير المعتمد و المتغيرات المستقلة .

وفي هذا البحث سوف نتطرق الى النوع الثاني فقط من المشاكل وهو (مشاكل مع الشواذ المتعددة لمجموعة من المتغيرات المستقلة)

اساليب معالجة المشكلة: Manners for treatment of problem

هناك طرائق كثيرة طورت لحل مشاكل المشاهدات الشاذة في التطبيقات الاحصائية من خلال تحديد المشاهدات الشاذة اولا ثم معالجتها ثانيا ومن هذه الطرائق :-

1- الطريقة الحصينة باستخدام مصفوفة التغاير الحصينة لـ كامبل Campbell's Robust covariance matrix [مصدر 3] .
2- الطريقة الحصينة باستخدام الوسيط للانحراف المطلق لـ هامبل Hamplel's median absolute deviation [مصدر 9] .
3- طريقة مقترحة من قبل الباحث وهي تعتبر طريقة حصينة لتحديد المشاهدات الشاذة باستخدام صيغة المسافة لكوك Cook's distance وبالاعتماد على [المصادر 5 و 6] .

هدف البحث (الاسلوب المقترح في البحث) :The search objective

كما لاحظنا ان الطرائق التي تهتم بمعالجة مشكلة المشاهدات الشاذة كثيرة ومتعددة وذلك يضعنا امام مجموعة من الاختيارات المتعددة لاختيار واحدة من تلك الطرائق والسؤال الذي يمكن ان يطرح نفسه هو (اي من هذه الطرائق التي تعالج المشاهدات الشاذة افضل لكي نستخدمها في عملية التقدير لمعلمات الانحدار الخطي)، ولأجل ان نحاول الوصول الى الاجابة عن هذا السؤال تم استخدام اسلوب المحاكاة التجريبي للمقارنة بين طرائق التقدير الحصينة والمعروضة في جانب اساليب معالجة المشكلة لدراسة مميزات وبالتالي التعرف خواص كل طريقة ضمن حالات افتراضية مختلفة من حيث (حجم العينات، عدد المشاهدات الشاذة، ...... وغيرها من الافتراضات التي سوف نتطرق اليها في جزء وصف تجارب المحاكاة في هذا البحث) .

سيتم في هذا البحث دراسة النوع الثاني فقط من المشاكل وهو (مشاكل مع الشواذ المتعددة لمجموعة من المتغيرات المستقلة) من خلال برنامج مكتوب بلغة فيجوال بيسك يتم استخدام اسلوب المحاكاة التجريبي في توليد بيانات لانموذج انحدار خطي مفترض ونعمل على وضع الية لجعل بعض المشاهدات للبيانات المولدة في المتغيرات المستقلة شاذة وبالتالي نقوم على تقدير معالم الانحدار لمعرفة تأثيرات المشاهدات الشاذة على تقديرات معلمات الانحدار بطرائق حصينة مختلفة (الطريقة الحصينة باستخدام مصفوفة التغاير الحصينة لـ كامبل Campbell's Robust covariance matrix ، الطريقة الحصينة باستخدام الوسيط للانحراف المطلق لـ هامبل Hampel's median absolute deviation، الطريقة الحصينة المقترحة لتحديد المشاهدات الشاذة باستخدام صيغة المسافة لكوك Cook's distance) اضافة الى طريقـة المربعات الصغرى الاعتيادية OLS التي لا تعالج الشواذ لمعرفة المدى التي تصل اليه شدة تأثير المشاهدات الشاذة على التقديرات في ظل ظروف مختلفة .

تحليل الانحدار الخطي : Linear regression analyses

في اغلب الاحيان نحن نرغب في توضيح التغيرات التي تحصل في المتغير المعتمد الممثل بالمتجه Y كمتغير استجابة (متغير معتمد) Dependent variable الى التغيرات التي تحصل في المتغيرات التوضيحية الممثلة بالمصفوفة X (متغيرات مستقلة) Independent variables سواء كانت (متغير مفرد او متغيرات متعددة) حيث نفرض ان العلاقة بين (X و Y) تكون خطية اي ان البيانات يمكن ان توصف كما يلي :

$$Y = b_1X_1 + b_2X_2 + \cdots \ldots\ldots + b_mX_m \qquad ----(1)$$

$$\underline{Y} = X\underline{b} \qquad ----(2)$$

لذلك نحن نحصل على مجموعة بيانات هي (متجه $\underline{Y}$ من درجة (n x 1) ومصفوفة X من درجة ( ) n x m) حيث ان n>=m حيث ان مجموعة البيانات هذه يمكن ان تقدم لنا نظام يتكون من n من المعادلات فيه m من المجاهيل غير المعلومة ممثلة بمتجه $\underline{b}$ من درجة (m x 1) حيث ان المتجة $\underline{b}$ يحاول ان يوضح العلاقة بين المتغيرات المعتمدة والمتغيرات المستقلة بشكل مضبوط او مطابق لكن متجه الخطأ $\underline{e}$ (البواقي) Residuals والذي هو من درجة (nx1) يجعل المعادلة (2) تكون كما يلي :-

$$\underline{Y} = X\underline{b} + \underline{e} \qquad ----(3)$$

تقديرات المربعات الصغرى الاعتيادية OLS لمتجة المعلمات b في انموذج الانحدار الخطي :
المعادلة (3) يمكن ان تكون كما يلي :

$$X^{-g}\underline{Y} = X^{-g}X\underline{b} + X^{-g}\underline{e} \qquad ----(4)$$

حيث ان X^{-g} هي مصفوفة المعكوس العام (generalized inverse) للمصفوفة X

حيث ان $X^{-g} = (X'X)^{-1}X'$ وبذلك يكون $X^{-g}X = (X'X)^{-1}X'X = I$

وبذلك يمكن لنا ان نعرض المعادلة (4) كما يلي :

$$(X'X)^{-1}X'\underline{Y} = \underline{b} + (X'X)^{-1}X'\underline{e}$$

وبما اننا نفرض ان المصفوفة (X) ومتجه الاخطاء ($\underline{e}$) غير مرتبطان مع بعض (uncorrelated) بحيث ان $X'\underline{e} = 0$ ، نحصل على تقديرات متجه المعلمات $\underline{b}$ بطريقة المربعات الصغرى الاعتيادية OLS كما يلي :

$$\underline{\hat{b}_{ols}} = (X'X)^{-1}X'\underline{Y} \qquad ------(5)$$

المشاهدات الشاذة في البيانات المشاهدة: Outliers in observed data

تقدير معلمات المتجه b بطريقة OLS يعتبر افضل تقدير لكن في حالة وجود قيم شاذة في البيانات المشاهدة تصبح تقديرات هذه الطريقة ضعيفة ولا تمثل البيانات افضل تمثيل لذلك نلجأ الى استخدام الطرق الحصينة Robust Methods في التقدير لمعلمات المتجة $\underline{b}$ والتي تقوم على التقليل من اثر القيم الشاذة على البيانات المدروسة لتعطينا تقدير لمتجه المعلمات $\underline{b}$ يمكن ان يمثل البينات افضل تمثيل حيث تقوم فكرة عمل هذه الطرق بشكل عام على تحديد المشاهدات الشاذة اولا ثم العمل على تقليل تأثيرها في عملية التقدير حيث ان الطرائق التي سوف تعرض في هذه الدراسة تستخدم صيغة المسافة لمهلنبوس Mahalanobis [مصدر 14] لتحديد القيم الشاذة اولا ومحاولة تقليل التأثير الذي تؤثر به القيم الشاذة على عملية تقدير المعلمات ثانيا.

صيغة المسافة لمهلنبوس : Mahalanobis Distance form

يمكن ان نعرف هذه الصيغة العامة للمسافة في الفروق بين قيم m من المتغيرات عن مراكزها بالمعادلة التالية [مصدر 14] :

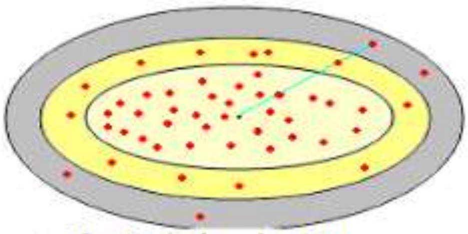
$$d_i = [(x_i - E(x))' \quad S^{-1} \quad (x_i - E(x))]^{1/2} \quad --(6) \qquad i=1,2,\dots\dots,n$$

حيث ان :

n : عدد المشاهدات الكلي (حجم العينة) .

x_i : متجة الى m من المتغيرات المستقلة من درجة (mx1) للمشاهدة i .

S : مصفوفة التباين و التباين المشترك الى m من المتغيرات المستقلة من درجة (X m) والتي تقيس مسافة الانحرافات القيم عن مراكزها الممثلة بـ $E(x)$.



The Mahalanobis Distance

فأذا كانت قيمة d_i اذا كانت اكبر من القيمة المفترضة ان تكون لبقية المشاهدات يمكن ان تعتبر المشاهدة **i** مشاهدة شاذة **outlier** عن بقية المشاهدات .

الطريقة الحصينة الاولى باستخدام مصفوفة التغاير الحصينة لـ كامبل:

Campbell's Robust covariance matrix :

باستخدام صيغة المسافة لمهلبونس كما في المعادلة (6) كمقياس للفروق عن مركز البيانات حيث قام الباحث **Campbell** [مصدر 3] بالحصول على مصفوفة التغاير الحصينة كما يلي :

تستخدم الوسط الحسابي الموزون $\bar{x}$ ومصفوفة تباين وتباين مشترك موزونة S من درجة **(m x m)** حيث ان في بداية اول تكرار الاوزان تكون w_i مساوية الى ($w_i = \frac{1}{n}$) ومجموع الاوزان يكون ($\sum_{i=1}^{n} w_i = 1$) ويتم تعريف مايلي :

$$d_0 = \sqrt{m} + a_1/\sqrt{2} \quad , \quad a_1 = 2 \quad , \quad a_2 = 1.25$$

لذلك نحن نحصل على :

$$\bar{x} = \sum_{i=1}^{n} w_i\, x_i / \sum_{i=1}^{n} w_i \quad ،$$

$$S = \sum_{i=1}^{n} w_i\,(x_i - \bar{x})^{/}(x_i - \bar{x}) / [\sum_{i=1}^{n} w_i^{\,2} - 1]$$

$$d_i = \{(x_i - \bar{x}) \quad S^{-1} \quad (x_i - \bar{x})^{/}\}^{1/2} \qquad -------(7)$$

$$w_i = \frac{\omega(d_i)}{d_i}$$

$$\omega(d_i) = d_i \quad \text{if} \quad d_i \le d_0 \quad \text{else} \quad \omega(d_i) = d_0 \exp\left[-0.5(d_i - d_0)^2/a_2^2\right]$$

حيث اذا كانت قيمة d_i تساوي صفر او قيمة قريبة جدا من الصفر فأن $w_i = 1$

بعد الحصول على مصفوفة الاوزان القطرية **W** من درجة **(nxn)** والتي عناصرقطرها تكون ممثلة بـ w_i يتم استخدام طريقة المربعات الصغرى الموزونة **WLS** بأوزان المصفوفة **W** للحصول على تقديرات معالم الانحدار $\hat{b}$ وكما يلي :

$$\underline{\hat{b}} = (WX^{/}WX)^{-1} WX^{/}W\underline{Y} \qquad -------(8)$$

الطريقة الحصينة الثانية باستخدام الوسيط للانحراف المطلق لـ هامبل :

Hamplel's median absolute deviation :

الباحث **Hamplel** [مصدر 7] يعرف الوسيط للانحرافات المطلقة (للوسيط) بالصيغة التالية :

$$S_H^*(d_i^*) = \underset{i}{median} |d_i^*|$$

$$d_i^* = \left| d_i^{\,m} - \underset{i}{median} (d_i^{\,m}) \right|$$

$$d_i^{\,m} = \{(x_i - x_{median}) \quad S^{-1} \quad (x_i - x_{median})^{/}\}^{1/2}$$

$$S = \sum_{i=1}^{n} w_i (x_i - x_{median})^{/} (x_i - x_{median}) / [\sum_{i=1}^{n} w_i^{\,2} - 1]$$

حيث ان $S_H = S_H^* (d_i^*)/0.6745$ والذي هو يكون القياس الحصين للانحرافات ، باستخدام S_H نستطيع تحديد اوزان الى نقاط بيانات مختلفة في بداية اول تكرار كما يلي :

- لقيم d_i^* التي تقع ضمن مدى $d_i^* \leq S_H$ تكون دالة الوزن w_i مساويـة الى ($w_i = 1$) .
- لقيم d_i^* التي تقع ضمن مدى $S_H \leq d_i^* \leq 2S_H$ تكون دلة الوزن w_i مساوية الى ($w_i = (1/2)^2$) .
- لقيم d_i^* التي تقع ضمن مدى $2S_H \leq d_i^* \leq 3S_H$ تكون دلة الوزن w_i مساوية الى ($w_i = (1/2)^4$) .
- وهكذا ....

بعد الحصول على مصفوفة الاوزان القطرية **W** من درجة **(nxn)** والتي عناصرقطرها تكون ممثلة بـw_i يتم استخدام استخدام طريقة المربعات الصغرى الموزونة **WLS** بأوزان المصفوفة **W** كما في المعادلة **(8)** للحصول على تقديرات معالم الانحدار $\hat{b}$.

الطريقة الحصينة الثالثة باستخدام صيغة المسافة لكوك (Cook's distance) :

يتم من خلال هذه الطريقة تحديد مواقع المشاهدات الشاذة من خلال صيغة المسافة للباحث **Cook** [المصادر 5 و 6] التالية للمشاهدة **i** :

$$D_i = \frac{\sum_{j=1}^{n} \left(\hat{y}_j - \hat{y}_j(i)\right)^2}{(m+1)\hat{\sigma}^2} \qquad i = 1,2,\dots,n \qquad ----(9)$$

حيث ان :

$\hat{\sigma}^2$: تقدير الوسط التربيعي للاخطاء بطريقة **OLS** .

$\hat{y}_j$: مقدر الانحدار الى الوسط الشرطي $E\left(y_j \middle| x_{1j}, \dots, x_{mj}\right)$ باستخدام تقديرات معالم انحدار بطريقة **OLS** .

$\hat{y}_j(i)$: مقدر الانحدار الى الوسط الشرطي $E\left(y_j \middle| x_{1j}, \dots, x_{mj}\right)$ للمشاهدة **i** باستخدام تقديرات معالم انحدار بطريقة **OLS** .

حيث $\underline{\hat{y}} = \underline{X\hat{b}_{ols}}$

ثم يتم استخراج الوسيط لقيم D_i كما يلي :

$$D^{med} = \text{median}(D_i)$$

ثم يتم استخراج الوسيط لكل متغير من المتغيرات المستقلة في النموذج (2) وكما يلي :

$$x_k^{med} = \text{median}(X_{ki}) \qquad k = 1,2,\dots,m \qquad i = 1,2,\dots,n$$

بعد ذلك يتم تحديد مواقع المشاهدات الشاذة في البيانات ضمن صفوف المصفوفة X والتي هي من درجة **(nxm)** وذلك من خلال اخذ الفرق المطلق بين الوسيط D^{med} وبين كل قيمة من قيم D_i حيث ان كل فرق اكبر من *0.5* يتم فيه اعتبار المشاهدة **i** مشاهدة شاذة الى المتغيرات $(x_{1i}, x_{2i}, \dots x_{mi})$ في المشاهدة *i* وعند تحقق الفرق $(|D_i - D^{med}|) > 0.5$ يتم اختبار الفرق لكل قيمة من المتغيرات $(x_{1i}, x_{2i}, \dots x_{mi})$ للمشاهدة *i* والتي تحقق $(x_{ki} - \mathbf{x}_k^{med}) > 2$ فرق اكبر من 2 عن الوسيط X_k^{med} لـ *n* من مشاهدات المتغير *k* $(x_{k1}, x_{k2}, \dots x_{kn})$ ليتم ابدال المشاهدة التي تحقق الشرط الاخير بقيمة الوسيط X_k^{med} بدلا المشاهدة الشاذة x_{ki} وكما يلي :

$$\forall_1 \; if \begin{cases} (|D_i - D^{med}|) > 0.5 \xrightarrow{i^{th}\ data\ is\ outlier} \forall_k \; if \begin{cases} (x_{ki} - \mathbf{x}_k^{med}) > 2 \xrightarrow{x_{ki}\ is\ outlier\ then} x_{ki} = \mathbf{x}_k^{med} \\ (x_{ki} - \mathbf{x}_k^{med}) < 2 \xrightarrow{x_{ki}\ ihave\ no\ outlier\ then} x_{ki} = x_{ki} \end{cases} \\ (|D_i - D^{med}|) < 0.5 \xrightarrow{i^{th}\ data\ have\ no\ outlier} x_{ki} = x_{ki} \end{cases}$$

$$i = 1, 2, \dots, n \qquad k = 1, 2, \dots, m$$

ثم بعد ذلك يتم استخدام طريقة المربعات الصغرى الاعتيادية **OLS** كما في المعادلة (5) وبذلك يتم التخفيف من اثر المشاهدات الشاذة في البيانات بعد تقريبها الى قيمة الوسيط .

وصف تجارب المحاكاة **Describes of Simulation experiments :**

من خلال استخدام برنامج مكتوب بلغة فيجوال بيسك يتم من خلاله توليد **1000** عينة باحجام **(25 ، 50 ، 100)** لمشاهدات تولد على ضوء الانموذج الخطي في المعادلة (3) لمتغيرين مستقلين فقط (x_1, x_2) يتم توليدهما من خلال استخدام معادلة التوليد للتوزيع الطبيعي والى اثنان من المتغيرات المستقلة بمتوسطات (μ_{x1} و μ_{x2}) على التوالي و تباينات (σ_{x1}^2 و σ_{x2}^2) على التوالي { اي $x_{i1} \sim N(\mu_{x_1}, \sigma_{x1}^2)$ و $x_{i2} \sim N(\mu_{x_2}, \sigma_{x2}^2)$ } ، اما توزيع الاخطاء يتم توليدها حسب معادلة التوليد للتوزيع الطبيعي بمتوسط $(\mu_e = 0)$ وتباين افتراضي (σ_e^2) { اي $e_i \sim N(\mu_e = 0, \sigma_e^2)$ } ، (حيث ان **i=1,2, …..n** وان **n** يمثل حجم العينة المفترض)،

وفيما يلي عرض للقيم الافتراضية لتجارب المحاكاة وكما يلي :-

- القيم الافتراضية لمعالم الانحدار النموذج (3) سوف نفرضها ونثبتها عند الاعداد الاتية :-

$$(\; b_0 = 1 \,, b_1 = 0.5 \,, b_2 = 0.2 \;)$$

- القيم الافتراضية لمعلمة متوسط مربعات الخطأ لانموذج الانحدار {اي تباين توزيع الاخطاء e_i } سوف نفترضه في حالتين وهي :
  - الحالة الاولى : $\sigma_e^2 = 0.5$
  - الحالة الثانية : $\sigma_e^2 = 5$
- القيم الافتراضية للمتوسطات في التوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2) سنثبتها عند العدد **10** وكما يلي :-

$$(\mu_{x1} = 10 \;, \mu_{x2} = 10 \;)$$

اما القيم الافتراضية للتباينات في التوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2) سوف نفرضها في حالتين و كما يلي :-

- الحالة الاولى : $(\sigma_{x1}^2 = 1 \ , \ \sigma_{x2}^2 = 1)$
- الحالة الثانية : $(\sigma_{x1}^2 = 5 \ , \ \sigma_{x2}^2 = 5)$

باستخدام معادلة Box Muller لتوليد التوزيع الطبيعي يتم توليد الاخطاء e_i وكذلك يتم توليد قيم المتغيرات المستقلة (x_{i1}, x_{i2}) ، وحسب النموذج (3) يتم اضافة قيمة e_i المولدة مع اضافة حاصل ضرب ماتم توليده للمتغيرات المستقلة (x_{i1}, x_{i2}) في معالم الانحدار المفترضة لنحصل على قيم المتغير المعتمد Y_i .

بعد توليد المشاهدات للمتغيرات المستقلة (x_{i1}, x_{i2}) والمتغير المعتمد (Y_i) وحسب النموذج (3) وللقيم الافتراضية اعلاه نعمل على نعمل على فرض حالتين من القيم الشاذة وهي :-

- الحالة الاولى : وضع اثنان من مشاهدات المتغيرات المستقلة شاذة (المشاهدة رقم 1 من المتغير X_1 والمشاهدة رقم 3 من المتغير X_2) .
- الحالة الثانية: وضع اربعة من مشاهدات المتغيرات المستقلة شاذة (المشاهدة رقم (1) ورقم (6) من المتغير X_1 والمشاهدة رقم (3) , ورقم (10) من المتغير X_2) .

حيث سنجعل قيمة المشاهدات شاذة وذلك بضرب قيمة المشاهدة المولدة الاصلية بالعدد (20) لجعلها تختلف عن مثيلتها الاصلية المولدة في تجربة المحاكاة .

نتائج تجارب المحاكاة: Results of simulation experiments

من خلال استخدام 1000 تجربة من البيانات المولدة في تجارب المحاكاة للقيم الافتراضية الموضحة وبعد جعل المشاهدات التي تم توليدها تمتلك قيم متطرفة (شاذة) وكما تم توضيح ذلك في حالتين في جزء وصف تجربة المحاكاة تم تقدير معلمات أنموذج الانحدار الخطي وبالطرائق الاتية :-

1. المربعات الصغرى الاعتيادية OLS .
2. الطريقة الحصينة الاولى باستخدام مصفوفة التغاير الحصينة لـ كامبل Campbell .
3. الطريقة الحصينة الثانية باستخدام الوسيط للانحراف المطلق لـ هامبل Hamplel.
4. الطريقة الحصينة الثالثة باستخدام صيغة المسافة لكوك Cook .

حيث تم ايجاد الاتي من الالف تجربة من تجارب المحاكاة :-

- الاوساط الحسابية لقيم معامل التحديد Coefficient of determination .
- الاوساط الحسابية لتقديرات قيم متوسط مربعات الخطأ Mean square error لإنموذج الانحدار .
- الاوساط الحسابية لتقديرات معلمات الانحدار .
- مقدار التحيز المطلق لتقديرات معلمات الانحدار .
- قيم متوسط مربعات الاخطاء لتقديرات معلمات الانحدار.

وذلك لكل طريقة من الطرق اعلاه وللقيم الافتراضية الموضحة من حيث احجام العينات والقيم الافتراضية لمعلمة الخطأ لنموذج الانحدار وللقيم الافتراضية لتباينات المتغيرات المستقلة وللحالات المفترضة من القيم الشاذة في تجارب المحاكاة، وكانت النتائج كما موضح في الجداول الاتية :-

جدول رقم (1) يعرض لنا الوسط الحسابي لقيم معامل التحديد لإنموذج الانحدار

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size	R^2_{ols}	$R^2_{Campbell}$	R^2_{Hampel}	R^2_{Cook}
2	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	0.0805	0.8642	0.9143	0.3265
			50	0.0416	0.8542	0.9121	0.3334
			100	0.0217	0.7649	0.9114	0.3448
		$\sigma^2_e = 5$	25	0.0825	0.5104	0.7554	0.1065
			50	0.0406	0.3953	0.7527	0.0812
			100	0.0211	0.2570	0.7479	0.0652
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	0.0776	0.7269	0.9010	0.6092
			50	0.0441	0.8839	0.9135	0.6371
			100	0.0290	0.8468	0.9140	0.6850
		$\sigma^2_e = 5$	25	0.0810	0.4655	0.7566	0.2227
			50	0.0422	0.4670	0.7610	0.2151
			100	0.0564	0.7431	0.8076	0.1808
4	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	0.0851	0.8812	0.9287	0.2756
			50	0.0412	0.8584	0.9174	0.2994
			100	0.0203	0.8390	0.9145	0.3226
		$\sigma^2_e = 5$	25	0.0833	0.5552	0.7732	0.0934
			50	0.0400	0.4932	0.7561	0.0744
			100	0.0206	0.3591	0.7437	0.0645
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	0.0878	0.7313	0.9040	0.4700
			50	0.0446	0.6891	0.9160	0.5697
			100	0.0262	0.7974	0.9167	0.6525
		$\sigma^2_e = 5$	25	0.0851	0.4841	0.7641	0.1785
			50	0.0418	0.4193	0.7637	0.1906
			100	0.0218	0.4087	0.7589	0.2001

من خلال نتائج المحاكاة تبين من الجدول رقم (1) ان الاوساط الحسابية لقيم معامل التحديد بطريقة Hamplel اظهرت اقرب النتائج الى العدد واحد الصحيح تليها طريقة Campbell اظهرت ثاني اقرب قيم الى العدد واحد الصحيح لمتوسطات معامل التحديد وتليها طريقة Cook والتي اظهرت ثالث اقرب قيم الى العدد واحد لمتوسطات معامل التحديد اما طريقة OLS فقط اظهرت اقل متوسطات لقيم معامل التحديد وهي اقرب الى الصفر وذلك لكافة احجام العينات (25 ، 50 ، 100) ولكل الحالات التي تم فرض فيها القيم الافتراضية لمعلمة الخطأ لانموذج الانحدار (0.5 ، 5) وكافة الحالات التي تم فيها فرض قيم افتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2) وكافة الحالات التي تم فيها وضع عدد مشاهدات شاذة في تجارب المحاكاة (2) من المشاهدات شاذة و (4) من المشاهدات شاذة) حيث نستنتج من ذلك ان طريقة Hamplel في التقدير للمعلمات تقدم لنا افضل معادلة انحدار تمثل البيانات افضل تمثيل .

جدول رقم (2) يعرض قيم الاوساط الحسابية لتقديرات قيم متوسط مربعات الخطأ لإنموذج الانحدار

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size	σ^2_{ols}	$\sigma^2_{Campbell}$	σ^2_{Hampel}	σ^2_{Cook}
2	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	0.7218	0.7631	1.2091	0.5187
			50	0.7585	0.4775	1.2226	0.5213
			100	0.7700	0.4798	1.2192	0.5126
		$\sigma^2_e = 5$	25	4.8122	4.6789	4.8310	4.1783
			50	5.0767	4.8528	4.6224	4.1912
			100	5.1722	4.9346	4.7982	4.2524
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	1.7968	1.8291	1.4508	0.7368
			50	1.8669	0.5060	1.2516	0.6959
			100	1.8846	0.4833	1.2331	0.6046
		$\sigma^2_e = 5$	25	5.8826	5.8336	4.3026	4.9233
			50	6.1863	4.6475	4.1811	5.0323
			100	6.2893	1.2522	4.8879	5.4483
4	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	0.7177	0.7242	1.0411	0.5611
			50	0.7591	0.7732	1.1574	0.5482
			100	0.7712	0.5219	1.1837	0.5306
		$\sigma^2_e = 5$	25	4.8084	4.4595	3.9711	4.7516
			50	5.0801	4.9025	4.1638	4.8893
			100	5.2070	4.9098	4.5901	4.9731
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	1.7761	1.8277	1.4602	1.0090
			50	1.8660	1.9228	1.2296	0.8260
			100	1.8906	0.8660	1.2045	0.6653
		$\sigma^2_e = 5$	25	5.8527	5.5684	4.2959	5.2153
			50	6.1899	6.0419	4.1839	5.1943
			100	6.2921	5.1002	4.2337	5.1280

من خلال نتائج المحاكاة تبين من الجدول رقم (2) نلاحظ ما يلي :

نلاحظ ان الاوساط الحسابية لتقديرات قيم متوسطات مربعات الخطأ لإنموذج الانحدار بطريقة **Hample** وطريقة **Campbel** اظهرتا قيما اقرب الى القيم الافتراضية لها في تجارب المحاكاة وذلك في معظم الحالات الافتراضية والتي تكون بها القيمة الافتراضية لمعلمة الخطأ لنموذج الانحدار تساوي $\sigma^2_e = 5$ وخصوصا في حالة التباينات المفترضة للمتغيرات المستقلة ($\sigma^2_{x1} = 1$ و $\sigma^2_{x2} = 1$).

جدول رقم (3) يعرض قيم الاوساط الحسابية لتقديرات معلمات الانحدار (b_0 , b_1 , b_2)
لإنموذج الانحدار وذلك في حالة وجود اثنان من القيم الشاذة في المتغيرات المستقلة

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size		OLS	Campbell	Hamplel	Cook
2	$\mu_{x_1} = 10, \sigma^2_{x1} = 1$ $\mu_{x_2} = 10, \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	b_0	7.9893	8.4493	7.2214	2.0831
				b_1	0.0003	-0.0326	0.0575	0.4230
				b_2	0.0001	-0.0131	0.0053	0.1688
			50	b_0	7.9824	1.3365	3.8144	1.7560
				b_1	0.0008	0.4657	0.2913	0.4437
				b_2	0.0003	0.2004	0.1098	0.1806
			100	b_0	7.9733	1.0102	1.5611	1.4876
				b_1	0.0013	0.4968	0.4463	0.4667
				b_2	0.0007	0.2021	0.1802	0.1844
		$\sigma^2_e = 5$	25	b_0	7.9894	8.4448	7.1792	2.1396
				b_1	0.0007	-0.0327	0.0582	0.4061
				b_2	0.0001	-0.012	0.0090	0.1806
			50	b_0	7.9725	1.3022	3.8575	1.6474
				b_1	0.0014	0.4675	0.2892	0.4502
				b_2	0.0005	0.2018	0.1071	0.1851
			100	b_0	7.9651	1.0327	1.5331	1.5622
				b_1	0.0016	0.4901	0.4480	0.4515
				b_2	0.0015	0.2067	0.1816	0.1925
	$\mu_{x_1} = 10, \sigma^2_{x1} = 5$ $\mu_{x_2} = 10, \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	b_0	7.9618	7.4204	4.6161	1.8913
				b_1	0.0007	-0.003	0.2358	0.4345
				b_2	0.0001	0.0582	0.0868	0.1767
			50	b_0	7.9339	1.1773	1.8781	1.7725
				b_1	0.0029	0.4821	0.4259	0.4452
				b_2	0.0012	0.2000	0.1688	0.1773
			100	b_0	7.8776	1.0164	1.1722	1.3863
				b_1	0.0069	0.4974	0.4747	0.4731
				b_2	0.0030	0.2009	0.1908	0.1878
		$\sigma^2_e = 5$	25	b_0	7.9626	7.4119	4.5609	2.0540
				b_1	0.0011	-0.0032	0.2372	0.4224
				b_2	0.0001	0.0599	0.0914	0.1724
			50	b_0	7.9235	1.1602	1.9198	1.7387
				b_1	0.0035	0.4829	0.4230	0.4490
				b_2	0.0015	0.2006	0.1670	0.1770
			100	b_0	7.8691	1.0273	1.1553	1.5780
				b_1	0.0072	0.4943	0.4763	0.4568
				b_2	0.0038	0.2030	0.1911	0.1853

جدول رقم (4) يعرض قيم الاوساط الحسابية لتقديرات معلمات الانحدار (b_0 , b_1 , b_2)
لانموذج الانحدار وذلك في حالة وجود اربعة من القيم الشاذة في المتغيرات المستقلة

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size		OLS	Campbell	Hampel	Cook
4	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	b_0	7.9874	8.4954	7.8605	2.7754
				b_1	0.0002	-0.0196	0.0175	0.3697
				b_2	0.0001	-0.0302	-0.0140	0.1525
			50	b_0	7.9854	8.5244	4.9878	2.2645
				b_1	0.0005	-0.0331	0.2115	0.4090
				b_2	0.0001	-0.0199	0.0750	0.1644
			100	b_0	7.9805	2.0249	1.9507	1.8321
				b_1	0.0009	0.3963	0.4201	0.4412
				b_2	0.0004	0.2011	0.1684	0.1756
		$\sigma^2_e = 5$	25	b_0	7.9945	8.5026	7.8590	2.8674
				b_1	0.0004	-0.0195	0.0175	0.3453
				b_2	0.0001	-0.0302	-0.0128	0.1687
			50	b_0	7.9804	8.5175	4.9635	2.2087
				b_1	0.0009	-0.0330	0.2122	0.4102
				b_2	0.0001	-0.0194	0.0765	0.1687
			100	b_0	7.9756	2.0096	1.9181	1.8046
				b_1	0.0011	0.3930	0.4231	0.4369
				b_2	0.0008	0.2060	0.1690	0.1828
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	b_0	7.9412	8.3671	5.8392	2.9996
				b_1	0.0010	-0.0163	0.1556	0.3549
				b_2	0.0003	-0.0226	0.0487	0.1446
			50	b_0	7.9412	7.9969	2.4465	2.2718
				b_1	0.0021	-0.0116	0.3873	0.4127
				b_2	0.0007	0.0093	0.1528	0.1598
			100	b_0	7.9190	2.6157	1.2619	1.6724
				b_1	0.0039	0.3533	0.4683	0.4575
				b_2	0.0016	0.1847	0.1890	0.1753
		$\sigma^2_e = 5$	25	b_0	7.9487	8.3740	5.8358	3.1019
				b_1	0.0012	-0.0162	0.1558	0.3437
				b_2	0.0002	-0.0227	0.0499	0.1465
			50	b_0	7.9359	7.9878	2.4315	2.3545
				b_1	0.0024	-0.0113	0.3881	0.4044
				b_2	0.0008	0.0098	0.1533	0.1597
			100	b_0	7.9141	2.6026	1.2476	1.8722
				b_1	0.0041	0.3521	0.4693	0.4357
				b_2	0.0020	0.1873	0.1898	0.1771

جدول رقم (5) يعرض لنا مقدار التحيز المطلق لتقديرات معلمات الانحدار لإنموذج الانحدار عن القيم الافتراضية لها في تجربة المحاكاة وذلك في حالة وجود اثنان من القيم الشاذة في المتغيرات المستقلة

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size		OLS	Campbell	Hamplel	Cook
2	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	b_0	6.9893	7.4493	6.2214	1.0831
				b_1	0.4996	0.5326	0.4424	0.0769
				b_2	0.2000	0.2131	0.1946	0.0311
			50	b_0	6.9824	0.3365	2.8144	0.7560
				b_1	0.4991	0.0342	0.2086	0.0562
				b_2	0.1996	0.0004	0.0901	0.0193
			100	b_0	6.9733	0.0102	0.5611	0.4876
				b_1	0.4986	0.0031	0.0536	0.0332
				b_2	0.1992	0.0021	0.0197	0.0155
		$\sigma^2_e = 5$	25	b_0	6.9894	7.4448	6.1792	1.1396
				b_1	0.4992	0.5327	0.4417	0.0938
				b_2	0.2000	0.2120	0.1909	0.0193
			50	b_0	6.9725	0.3022	2.8575	0.6474
				b_1	0.4985	0.0324	0.2107	0.0497
				b_2	0.1994	0.0018	0.0928	0.0148
			100	b_0	6.9651	0.0327	0.5331	0.5622
				b_1	0.4983	0.0098	0.0519	0.0484
				b_2	0.1984	0.0067	0.0183	0.0074
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	b_0	6.9618	6.4204	3.6161	0.8913
				b_1	0.4992	0.5030	0.2641	0.0654
				b_2	0.2000	0.1417	0.1131	0.0232
			50	b_0	6.9339	0.1773	0.8781	0.7725
				b_1	0.4970	0.0178	0.0740	0.0547
				b_2	0.1987	0.0001	0.0311	0.0226
			100	b_0	6.8776	0.0164	0.1722	0.3863
				b_1	0.4930	0.0025	0.0252	0.0268
				b_2	0.1969	0.0009	0.0091	0.0121
		$\sigma^2_e = 5$	25	b_0	6.9626	6.4119	3.5609	1.0540
				b_1	0.4988	0.5032	0.2627	0.0775
				b_2	0.1999	0.1400	0.1085	0.0275
			50	b_0	6.9235	0.1602	0.9198	0.7387
				b_1	0.4964	0.0170	0.0769	0.0509
				b_2	0.1984	0.0006	0.0329	0.0229
			100	b_0	6.8691	0.0273	0.1553	0.5780
				b_1	0.4927	0.0056	0.0236	0.0431
				b_2	0.1961	0.0030	0.0088	0.0146

جدول رقم (6) يعرض لنا مقدار التحيز المطلق لتقديرات معلمات الانحدار لإنموذج الانحدار عن القيمة الافتراضية فى المحاكاة وذلك فى حالة وجود اربعة من القيم الشاذة فى المتغيرات المستقلة

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size		OLS	Campbell	Hamplel	Cook
4	$\mu_{x_1} = 10$, $\sigma^2_{x1} = 1$ $\mu_{x_2} = 10$, $\sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	b_0	6.9874	7.4954	6.8605	1.7754
				b_1	0.4997	0.5196	0.4824	0.1302
				b_2	0.2000	0.2302	0.2140	0.0474
			50	b_0	6.9854	7.5244	3.9878	1.2645
				b_1	0.4994	0.5331	0.2884	0.0909
				b_2	0.1998	0.2199	0.1249	0.0355
			100	b_0	6.9805	1.0249	0.9507	0.8321
				b_1	0.4990	0.1036	0.0798	0.0587
				b_2	0.1995	0.0011	0.0315	0.0243
		$\sigma^2_e = 5$	25	b_0	6.9945	7.5026	6.8590	1.8674
				b_1	0.4995	0.5195	0.4824	0.1546
				b_2	0.2000	0.2302	0.2128	0.0312
			50	b_0	6.9804	7.5175	3.9635	1.2087
				b_1	0.4990	0.5330	0.2877	0.0897
				b_2	0.1998	0.2194	0.1234	0.0312
			100	b_0	6.9756	1.0096	0.9181	0.8046
				b_1	0.4988	0.1069	0.0768	0.0630
				b_2	0.1991	0.0060	0.0309	0.0171
	$\mu_{x_1} = 10$, $\sigma^2_{x1} = 5$ $\mu_{x_2} = 10$, $\sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	b_0	6.9412	7.3671	4.8392	1.9996
				b_1	0.4989	0.5163	0.3443	0.1450
				b_2	0.1996	0.2226	0.1512	0.0553
			50	b_0	6.9412	6.9969	1.4465	1.2718
				b_1	0.4978	0.5116	0.1126	0.0872
				b_2	0.1992	0.1906	0.0471	0.0401
			100	b_0	6.9190	1.6157	0.2619	0.6724
				b_1	0.4960	0.1466	0.0316	0.0424
				b_2	0.1983	0.0152	0.0109	0.0246
		$\sigma^2_e = 5$	25	b_0	6.9487	7.3740	4.8358	2.1019
				b_1	0.4987	0.5162	0.3441	0.1562
				b_2	0.1997	0.2227	0.1500	0.0534
			50	b_0	6.9359	6.9878	1.4315	1.3545
				b_1	0.4975	0.5113	0.1118	0.0955
				b_2	0.1991	0.1901	0.0466	0.0402
			100	b_0	6.9141	1.6026	0.2476	0.8722
				b_1	0.4958	0.1478	0.0306	0.0642
				b_2	0.1979	0.0126	0.0101	0.0228

من خلال نتائج المحاكاة تبين من الجداول (3,4,5,6) ما يلي :

1- ان الاوساط الحسابية لتقديرات قيم معلمات أنموذج الانحدار بطريقة Cook اظهرت افضل اقرب قيم الى القيم الافتراضية لها في تجارب المحاكاة للعينات ذات الحجم (25) وذلك لكافة الحالات الافتراضية المستخدمة في تجارب المحاكاة من حيث (القيم الافتراضية لمعلمة الخطأ لنموذج الانحدار والقيم الافتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2) والحالات التي تم فيها وضع عدد مشاهدات (2 ، 4) شاذة في تجارب المحاكاة) .

2- في حالة عدد المشاهدات الشاذة (2) واحجام العينات (50 ، 100) تبين ان الاوساط الحسابية لتقديرات قيم معلمات أنموذج الانحدار بطريقة Campbell اظهرت افضل اقرب قيم الى القيم الافتراضية لها في تجارب المحاكاة وذلك لكافة الحالات الافتراضية المستخدمة في تجارب المحاكاة من حيث (القيم الافتراضية لمعلمة الخطأ لنموذج الانحدار والقيم الافتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2)) .

3- في حالة عدد المشاهدات الشاذة (4) واحجام العينات (50 ، 100) تبين ان الاوساط الحسابية لتقديرات قيم معلمات أنموذج الانحدار بطريقة Cook اظهرت افضل اقرب قيم الى القيم الافتراضية لها في تجارب المحاكاة وذلك لكافة الحالات الافتراضية المستخدمة في تجارب المحاكاة من حيث (القيم الافتراضية لمعلمة الخطأ لنموذج الانحدار والقيم الافتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2)).

4- في حالة عدد المشاهدات الشاذة (2) تبين ان الاوساط الحسابية لتقديرات قيم معلمات أنموذج الانحدار بطريقة Hamplel اظهرت ثاني افضل قيم اقرب الى القيم الافتراضية لها في تجارب المحاكاة للعينات ذات الحجم (25) وذلك لكافة الحالات الافتراضية المستخدمة في تجارب المحاكاة من حيث (القيم الافتراضية لمعلمة الخطأ لنموذج الانحدار والقيم الافتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2)) .

جدول رقم (7) يعرض لنا قيم متوسط مربعات الاخطاء لتقديرات معلمات الانحدار لإنموذج الانحدار وذلك في حالة وجود اثنان من القيم الشاذة في المتغيرات المستقلة

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size		OLS	Campbell	Hampel	Cook
2	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	b_0	48.8974	55.5939	39.2611	8.9968
				b_1	0.2496	0.2836	0.1986	0.0598
				b_2	0.0400	0.0460	0.0399	0.0280
			50	b_0	48.7782	2.2687	10.3176	5.4170
				b_1	0.2492	0.0112	0.0551	0.0374
				b_2	0.0399	0.0110	0.0188	0.0140
			100	b_0	48.6414	1.0570	2.5815	3.0043
				b_1	0.2486	0.0051	0.0136	0.0201
				b_2	0.0397	0.0054	0.0122	0.0086
		$\sigma^2_e = 5$	25	b_0	49.1726	56.044	41.2293	46.5221
				b_1	0.2493	0.2839	0.2096	0.2703
				b_2	0.0401	0.0482	0.0494	0.2011
			50	b_0	48.7729	20.919	24.1611	22.3421
				b_1	0.2487	0.0970	0.1202	0.1237
				b_2	0.0399	0.1095	0.0857	0.1019
			100	b_0	48.6058	10.5709	17.707	11.9443
				b_1	0.2485	0.0516	0.0867	0.0669
				b_2	0.0395	0.0544	0.0915	0.0529
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	b_0	48.5819	41.8081	14.5294	4.9020
				b_1	0.2492	0.2536	0.0780	0.0377
				b_2	0.0400	0.0243	0.0175	0.0115
			50	b_0	48.1399	0.6934	1.7576	3.8485
				b_1	0.2470	0.0040	0.0101	0.0292
				b_2	0.0395	0.0022	0.0054	0.0070
			100	b_0	47.3327	0.2243	0.5666	1.7857
				b_1	0.2431	0.0012	0.0032	0.0136
				b_2	0.0388	0.0010	0.0028	0.0039
		$\sigma^2_e = 5$	25	b_0	48.8713	42.6851	18.0535	13.5981
				b_1	0.2490	0.2542	0.0970	0.0836
				b_2	0.0402	0.0309	0.0352	0.0477
			50	b_0	48.1338	4.6900	7.5587	7.2063
				b_1	0.2466	0.0226	0.0365	0.0453
				b_2	0.0395	0.0219	0.0332	0.0244
			100	b_0	47.2992	2.1425	4.0224	4.5289
				b_1	0.2430	0.0104	0.0196	0.0296
				b_2	0.0387	0.0108	0.0206	0.0133

جدول رقم (8) يعرض لنا قيم متوسط مربعات الاخطاء لتقديرات معلمات الانحدار لإنموذج الانحدار وذلك في حالة وجود اربعة من القيم الشاذة في المتغيرات المستقلة

no. outlier	$x_{i1} \sim N(\mu_{x_1}, \sigma^2_{x1})$ $x_{i2} \sim N(\mu_{x_2}, \sigma^2_{x2})$	$e_i \sim N(\mu_e = 0, \sigma^2_e)$	Sample size		OLS	Campbell	Hamplel	Cook
4	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 1$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 1$	$\sigma^2_e = 0.5$	25	b_0	48.8726	56.2342	47.2858	12.4626
				b_1	0.2497	0.2700	0.2337	0.0852
				b_2	0.0400	0.0530	0.0465	0.0303
			50	b_0	48.8194	56.6709	17.2362	8.1494
				b_1	0.2494	0.2842	0.0900	0.0556
				b_2	0.0399	0.0487	0.0212	0.0171
			100	b_0	48.7394	3.1093	2.8550	4.4178
				b_1	0.2490	0.0257	0.0154	0.0324
				b_2	0.0398	0.0059	0.0111	0.0108
		$\sigma^2_e = 5$	25	b_0	49.2518	56.6127	48.2902	48.6003
				b_1	0.2496	0.2699	0.2375	0.2909
				b_2	0.0401	0.0530	0.0496	0.1907
			50	b_0	48.8764	56.7825	24.524	24.6674
				b_1	0.2491	0.2841	0.1254	0.1404
				b_2	0.0400	0.0495	0.0554	0.1020
			100	b_0	48.7378	11.2488	16.064	13.7302
				b_1	0.2489	0.0592	0.0765	0.0756
				b_2	0.0397	0.0561	0.0799	0.0557
	$\mu_{x_1} = 10,\ \sigma^2_{x1} = 5$ $\mu_{x_2} = 10,\ \sigma^2_{x2} = 5$	$\sigma^2_e = 0.5$	25	b_0	48.2995	54.4195	24.3087	10.957
				b_1	0.2489	0.2667	0.1232	0.0769
				b_2	0.0399	0.0497	0.0253	0.0179
			50	b_0	48.2385	49.2107	2.8927	6.3665
				b_1	0.2479	0.2619	0.0168	0.0463
				b_2	0.0397	0.0383	0.0056	0.0102
			100	b_0	47.9005	5.8897	0.5565	3.0495
				b_1	0.2461	0.0510	0.0033	0.0216
				b_2	0.0393	0.0037	0.0025	0.0064
		$\sigma^2_e = 5$	25	b_0	48.6816	54.805	26.3122	18.6734
				b_1	0.2488	0.2666	0.1320	0.1196
				b_2	0.0400	0.0498	0.0335	0.0502
			50	b_0	48.2929	49.3500	7.3759	10.4701
				b_1	0.2476	0.2617	0.0379	0.0683
				b_2	0.0397	0.0396	0.0262	0.0269
			100	b_0	47.8989	7.2577	3.8831	5.9179
				b_1	0.2459	0.0567	0.0186	0.0398
				b_2	0.0392	0.0128	0.0191	0.0155

من خلال نتائج المحاكاة تبين من الجداول (8,7) ما يلي :

1- **ان قيم متوسط مربعات الاخطاء لتقديرات قيم معلمات أنموذج الانحدار بطريقة Cook اظهرت اقل قيم في تجارب المحاكاة للعينات ذات الحجم (25) وذلك لمعظم الحالات الافتراضية المستخدمة في تجارب المحاكاة من حيث (القيم الافتراضية لمعلمة الخطأ لنموذج الانحدار والقيم الافتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2) والحالات التي تم فيها وضع عدد مشاهدات شاذة في تجارب المحاكاة) .**

2- **في حالة عدد المشاهدات الشاذة (2) واحجام العينات (50 ، 100) تبين ان قيم متوسط مربعات الاخطاء لتقديرات قيم معلمات أنموذج الانحدار بطريقة Campbell اظهرت اقل قيم في تجارب المحاكاة وذلك لمعظم الحالات الافتراضية المستخدمة في تجارب المحاكاة من (القيم الافتراضية لمعلمة الخطأ لنموذج الانحدار والقيم الافتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2)) .**

3- **في حالة عدد المشاهدات الشاذة (4) واحجام العينات (50 ، 100) تبين ان قيم متوسط مربعات الاخطاء لتقديرات قيم معلمات أنموذج الانحدار بطريقة Hamplel اظهرت اقل قيم في تجارب المحاكاة في وذلك لمعظم الحالات الافتراضية المستخدمة في تجارب المحاكاة من حيث (القيم الافتراضية لمعلمة الخطأ لنموذج الانحدار والقيم الافتراضية للتباينات للتوزيع الطبيعي للمتغيرات المستقلة (x_1, x_2)) .**

الاستنتاجات : **Conclusions**

- ✓ قيم معامل التحديد بطريقة Hamplel اظهرت اقرب النتائج الى العدد واحد الصحيح تليها طريقة Campbell اظهرت ثاني اقرب قيم الى العدد واحد الصحيح لمتوسطات معامل التحديد و تليها طريقة Cook والتي اظهرت ثالث اقرب قيم الى العدد واحد .
- ✓ الاوساط الحسابية لتقديرات قيم متوسطات مربعات الخطأ لإنموذج الانحدار بطريقة **Hample** وطريقة **Campbel** اظهرتا قيما اقرب الى القيم الافتراضية لها في تجارب المحاكاة وذلك في معظم الحالات والتي تكون يها القيمة الافتراضية لمعلمة الخطأ لنموذج تكون كبيرة وخصوصا في حالة التباينات ذات التشتت القليل للمتغيرات المستقلة) .
- ✓ في ظل وجود اثنان من المشاهدات الشاذة اظهرت تقديرات قيم معلمات أنموذج الانحدار بطريقة Cook افضل اقرب قيم الى القيم الافتراضية لها في تجارب المحاكاة للعينات ذات الحجم 25 اما في العينات (50 ، 100) فقد كانت تقديرات قيم معلمات أنموذج الانحدار بطريقة Campbell اظهرت افضل اقرب قيم الى القيم الافتراضية لها في تجارب المحاكاة .
- ✓ في ظل وجود اربعة من المشاهدات الشاذة اظهرت تقديرات قيم معلمات أنموذج الانحدار بطريقة Cook افضل اقرب قيم الى القيم الافتراضية لها في تجارب المحاكاة للعينات ذات الحجم (50 ، 100) .

التوصيات :**Recommendations**

- ✓ طريقة Hamplel تعد افضل طريقة ممكن ان نعتمدها في الحصول على معلمات انموذج الانحداريمثل البيانات افضل تمثيل حيث نحصل من تقديرات هذه الطريقة على افضل قيمة لمعامل التحديد قريب من العدد واحد الصحيح .
- ✓ طريقة **Hample** وطريقة **Campbel** تعد افضل طرائق اذا ما رغبنا في الحصول على اقرب تقدير لمتوسط مربعات الخطأ لإنموذج الانحدار من القيمة الحقيقية له في المجتمع وذلك في الحالات والتي تكون يها قيمة معلمة الخطأ لنموذج الانحدار تكون كبيرة وخصوصا في حالة التباينات ذات التشتت القليل للمتغيرات المستقلة .
- ✓ طريقة Cook اثبتت افضليتها في تقدير معلمات انموذج الانحدار في حالتين (العينات الصغيرة، عدد المشاهدات الشاذة قليل)

References: المصادر

1. Alam, M.A., Nasser, M. and Imon, A.H.M.R. (2008). Sensitivity and Influence Analysis of Estimators of Correlation Coefficients, Journal of Probability and Statistics, Vol. 3, No. 1, pp. 119-136.
2. Bagheri, A., Habshah, M. and Imon, A.H.M.R. (2009). Two-Step Robust Diagnostic Method for Identification of Multiple High leverage Points, (Accepted for Publication) Journal of Mathematics and Statistics.
3. Campbell , N,A. (1980) " Robust procedures in multivariate analysis I : robust covariance estimation " , Applied statistics 29(3) :231-237.
4. Colin Chen, SAS Institute Inc., Cary, NC " Robust Regression and Outlier Detection with the ROBUSTREG Procedure ", Statistics and Data Analysis , Paper 265-27.
5. Cook, R. Dennis (Feb 1977). "Detection of Influential Observations in Linear Regression". Technometrics (American Statistical Association) 19 (1): 15–18.
6. Cook, R. Dennis (Mar 1979). "Influential Observations in Linear Regression". Journal of the American Statistical Association (American Statistical Association) 74 (365): 169–174
7. Habshah, M., Norazan, R. and Imon, A.H.M.R. (2009) The Performance of Diagnostic-Robust Generalized Potentials for the Identification of Multiple High Leverage Points in Linear Regression, (Accepted for Publication) Journal of Applied Statistics.
8. Hadi, A.S., Imon, A.H.M.R. and Werner, M. (2009). Outlier Identification in "Wiley Interdisciplinary Reviews: Computational Statistics" (In Press).
9. Hampel , F. R. ,Ronchtti, EM . Rousseeuw P.J. and W.A. Stahel, W. A. (1986) "Robust statistics the approach based on influence functions , Willy , New york.
10. Imon, A. H. M. R. (2009). Deletion Residuals in the Detection of Heterogeneity of Variances in Linear Regression, Journal of Applied Statistics, Vol. 36, No. 3, pp. 347-358.
11. Imon, A. H. M. R. and Davies, P. (2007). A New Graphical Display for Locating Multiple Influential Observations, High Leverage Points and Outliers in Linear Regression, International Journal of Statistical Sciences, Vol. 5 (Special issue in honour of Professor Mir Masoom Ali) pp. 315-330.
12. Imon, A.H.M.R. and Hadi, A.S. (2008). Identification of Multiple Outliers in Logistic Regression, Communications in Statistics-Theory and Methods, Vol. 37, No. 11, pp. 1697-1709.
13. Khan, M. A. I. and Imon, A. H. M. R. (2000) On Performance of Some Measures of Central Values as Simulated Mean in the Identification of High Leverage Points, Rajshahi University Studies, Part-B, Journal of Science, Vol. 28, pp. 113-123.
14. Mahalanobis ,p.C. (1936) " on the Generalized Distance in statistics", Proceedings of the national instituted of since of india 12:49-55.
15. Nurunabi, A.A.M., Imon, A.H.M.R. and Nasser, M. (2008) Applications of Robust Regression in Business, Econometrics and Social Science, North South Economic Reviews, Vol. 2, No. 1, pp. 1-14.
16. Rana, M.S., Habshah, M. and Imon, A.H.M.R. (2009). A Robust Rescaled Moments Test for Normality in Regression, (Accepted for Publication) Journal of Mathematics and Statistics.