# مقارنة الطريقتين المقترحتين (MPRM) و (MPLSKURSD) مع طرائق المربعات الصغرى الجزئية (PLS) الحصينة

أ.م.د سجى محمد حسين*                    م.د. رباب عبدالرضا صالح**


## المستخلص

ان ظهور مشكلة التعدد الخطي التام في المتغيرات التوضيحية لنموذج الانحدارالخطي المتعدد المتغيرات يجعل من الصعوبة تطبيق الطرائق الكلاسيكية مثل طريقة (ols) لانها تعطي نتائج غير دقيقة ولمعالجة مثل هذه المشكلة تستعمل طرائق اخرى منها المربعات الصغرى الجزئية (pls) .الا ان هذه الطريقة تكون حساسة تجاه القيم الشاذه ان وجدت في مجموعة البيانات لذا فمن المستحسن اللجوء الى الطرائق الحصينة ومنها (PRM) و (PLSKURSD). في هذا البحث سيتم استعمال الطريقتين اعلاه فضلاً عن (pls) وسيتم مقارنتها مع الطريقتين المقترحتين (MPRM) و (MPLSKURSD) عن طريق المحاكاة من خلال تجربتين اعتمدت التجربة الاولى على عدة انواع من القيم الشاذة من البيانات وبنسب مختلفة من التلوث ولحجوم عينات وابعاد متغيرات مختلفة واعتمدت الثانية على المقارنة بين الطرائق عندما يتوزع الخطأ توزيعاً طبيعياً فضلاً عن توزيعات اخرى .

المصطلحات : المربعات الصغرى الجزئية ، الشواذ ،خوارزمية simpls ، وسيط متعدد المتغيرات، مصفوفة التباين المشترك الحصينة ، التفرطح،الاسقاطات ، تجزئة القيمة المفردة



## Abstract

*The emergence of multi-linear full problem in the explanatory variables to model multi Alanhaddaralkhti variables makes it difficult to apply the classical methods such as the method (ols) because they give inaccurate results and to address such a problem using other methods such as least-squares District (pls) .ala that this method be sensitive towards anomalous values, if any, in the data set, so it is advisable to resort to methods such as fortified (PRM) and (PLSKURSD). In this research will be to use the two methods above in addition to (pls) will be compared with the methods proposed (MPRM) and (MPLSKURSD) by simulation through two experiments first experiment relied on several types of anomalous values of data and different rates of pollution and volumes of samples and the dimensions of different variables and adopted a second on a comparison between the methods when the error is distributed naturally distributed in addition to other distributions.*

*Terminology: partial least squares, gay, simpls algorithm, broker multivariate, covariance matrix fortified, kurtosis, projections, segmentation singular value*


---

* جامعة بغداد / كلية الادارة والاقتصاد .

** جامعة بغداد / كلية الادارة والاقتصاد .

مقبول للنشر بتأريخ 2016/3/9

مستل من أطروحة دكتوراه

# 1-1 المقدمة

في نماذج الانحدار وخاصة التصاميم ذات الابعاد الكبيرة وجود مشكلة الاعتمادات الخطية بين الاعمدة لتلك التصاميم ، في هذه الحالة فان النتائج ممكن ان تبدو وهمية او متناقضة . ان الاعتماد على نتائج تقديرات المربعات الصغرى او الامكان الاعظم تصبح اسوأ عندما تزداد العلاقة الخطية بين اعمدة التصميم الناتجة عن التعدد الخطي . في الادبيات هناك طرائق مختلفة لمعالجة مشكلة التعدد الخطي بين المتغيرات التوضيحية ، واحد من هذه الحلول الممكنة لعدد كبير من المدخلات المرتبطة هو باستخدام طريقة المربعات الصغرى الجزئية (PLS) التي تنتج عدد اصغر من التوافيق الخطية المتعامدة للمدخلات الاصلية. الا ان هذه الطريقة عاجزة عن

معالجة الشواذ (outliers) اذ ان الشواذ بشكل عام يصعب اكتشافها في الابعاد العالية ولكنها بصورة عامة تؤثر على التقدير . الاساليب الحصينة اوجدت الحلول لهذه المشكلة لذا افترض الباحثون استبدال طريقة (pls) بطرائق (pls) الحصينه التي بعض منها تعتمد على الانحدار الحصين مثل طريقة (PRM) وبعض يعتمد على حصانة مصفوفة التباين والتباين المشترك مثل طريقة (PLSKURSD) . هناك الكثير من الدراسات والبحوث التي تخصصت بهذا المجال نذكر منها : في عام (1995م) قدم كل من (and Andrews,C.W.Cummins,D.J.)[4] بحثاً حول حصانة (PLS) يعتمد على طريقة المربعات الصغرى الجزئية التكرارية(IRPLS) وقد استعمل عدة دوال للوزن . في عام (1997م) قدم كل من Pena (and Prieto) [14] طريقةحصينة لتقدير مصفوفة التباين والتباين المشترك تعتمد على المعلومات الحاصل عليها من خلال الاسقاطات على مجموعة محددة من الاتجاهات . في عام (1998 م) قدم كل من(Juan,A.andRosario,R.)[12]تقنية حصينة جديدة الى المربعات الصغرى الجزئية (PLS) تعتمدعلى حصانةمصفوفة التباين الحصينة باستعمال المقدر (Stahel_donoho,SD) وقدأستعمل معيار(PRESS)لمقارنة بين الطريقة الجديدة وبقية الطرق( PLS,PLS2,PLSIR,IRPLS).فـي عام (2003م) قدما كل من(Hubert &Branden) [10] طرائق حصينة جديدة الى طريقة انحـدار المربعات الصغرى الجزئية تعتمد عـلـى خـوارزمية (SIMPLS) وهمـا خـوارزمية RSIMCD و RSIMPLS )

في عام (2007م) قدم (المندلاوي)[2] بحثاً حول حصانة (PLS) ذكرت فيه ان هناك خوارزميات حصينة لانحدار المربعات الصغرى الجزئية بالاعتماد على خوارزمية (SIMPLS) وهي خوارزمية (RSIMPLS) والتي تستعمل للابعاد العليا وخوارزمية (RSIMCD) والتي تستعمل للابعاد القليلة .في عام(2010م) قدم كل من (Lieebmann,B.,& et. al)[13]بحثاً حول حصانة (PLS) باستعمال انحدار(M الحصين الجزئي (PRM)) حيث ان هذه الطريقة تعتمد على حساب اوزان معينة

# 2-1 الهدف:

ان هدف البحث هو معالجة مشكلة التعدد الخطي بين المتغيرات التوضيحية بوجود القيم الشاذة باستعمال بعض الطرائق الحصينة للمربعات الصغرى الجزئية بالاضافة الى الطرائق المقترحة باستعمال خوارزمية (SIMPLS) والمقارنة بين دقة مقدرات المربعات الصغرى الجزئية والحصينة منها من خلال تجربتين للمحاكاة ،التجربة الاولى عند وجود عدة انواع من القيم الشاذه في البيانات. اما في تجربة المحاكاة الثانية فقد تمت المقارنة بين المقدرات عندما يتوزيع الخطأ توزيعا طبيعيا اضافة الى توزيعات اخرى وفي كلا التجربتين تم تغيير حجوم العينات وابعاد المتغيرات.

## 2 - أنحدار المربعات الصغرى الجزئية[10] (PLSR)

ان انحدار المربعات الصغرى الجزئية (PLS) يستخدم لربط مجموعتين من المتغيرات بواسطة نموذج خطي . ان المجموعة الاولى من المتغيرات تحتوي علی p من المتغيرات التوضيحية التي تتضمن ارتباطا عاليا او عندما يكون عدد المتغيرات التوضيحية يفوق عدد المشاهدات. والمجموعة الثانية تتضمن واحد او q من متغيرات الاستجابة . في هذه الطريقة تم وضع عدة خوارزميات لحل مشكلة تقليل الابعاد وبالتالي التخلص من مشكلة التعدد الخطي ومنها خوارزمية (SIMPLS).

افرض بان المصفوفة: $X_{n,p} = (\underline{X}_1, \underline{X}_2, ..., \underline{X}_n)'$ حيث $\underline{X}_i = (X_{i1}, X_{i2}, \cdots, X_{ip})$ الى $i^{th}$ من المشاهدات. والمصفوفة: $Y_{n,q} = (\underline{Y_1}, \underline{Y_2}, \cdots, \underline{Y_n})'$ فأنموذج الانحدار الخطي التالي :

$$y_i = \beta_0 + \beta'_{p,q}\ x_i + e_i \qquad \cdots (1)$$

أذ ان $e_i$: حد الخطأ الذي يشترط فيه: $\mathrm{E}(e_i) = 0$ , $\mathrm{cov}(e_i) = \Sigma_e$ لحجم q .

$\beta_0 = (\beta_{01} \ldots \beta_{0q})$، حد ثابت غير معلوم ببعد q .

$\beta_{pq}$ : تمثل المعلمات غير المعلومة وهي مصفوفة الميل ببعد pxq.

ان خوارزمية SIMPLS تفترض ان المتغيرات x و y لها علاقة من خلال النموذج الثنائي الاتي:

$$x_i = \bar{x} + P_{p,k}\tilde{t}_i + g_i \quad ...(2)$$

$$y_i = \bar{y} + A'_{q,k}\tilde{t}_i + f_i \quad ..(3)$$

$\bar{x}, \bar{y}$: الوسط الحسابي للمتغيرات y،x ; $\tilde{t}_i$ : القياسات او النقاط (scores) ببعد k حيث k ≤p

$p_{p,k}$ : مصفوفة التحميل (x- loading) ; $A'_{k,p}$ : مصفوفة الميل في انحدار $y_i$ على $\tilde{t}_i$ $g_i$ , $f_i$ : البواقي.

وفيما يأتي تلخيص لاهم الخطوات لخوارزمية SIMPLS

تفترض الخوارزمية اولا تشكيل او بناء المركبات حيث يتم الحصول على (h)من المركبات التي تكون كتوافيق خطية من x من المتغيرات والتي لها اكبر تباين مشترك مع التوافيق الخطية من المتغيرات (y) .

1- تبدأ هذه الخوارزمية مع $S^1_{xy} = S_{xy}$ التي هي مصفوفة التباين والتباين المشترك cross-covariance matrix.

2- تكرر الخطوات من 1-2 الى 4- 2 لكل h= 1 … a

2-1 حساب اول زوج من المتجهات($r_a$،$q_a$)والتي تضع الى يسار ويمين المصفوفة $S_{yx}$ حيث انالمتجة هو $\underline{r_a}$ المتجه المميز الى $S_{xy} * S_{yx}$ و ابعادها (p×p) والمتجة $\underline{q_a}$ هو المتجة المميز الى $S_{yx}*S_{xy}$ و ابعادها (q×q)

2-2 حساب $h^{th}$ من المركبات الخطية وكما يلي ويكون normalize

$$\tilde{T}_{n,h} = \tilde{X}_{n,p}\ R_{p,h}$$

أذ $R_{p,h}$ =(, $\underline{r_1}$ … $\underline{r_h}$ )

2-3 حساب $h^{th}$ من x-loading وذلك بانحدار x على $t_h$ وحسب المعادلة الاتية

$$P_j = \tilde{X}'\ \tilde{X}\ r_j\ /\ (r'_j\ \tilde{X}\tilde{X}\ r_j) = S_x\ r_j\ /\ (r'_j\ S_x r_j)$$

حيث: $S_x$ هي مصفوفة التباين للمتغيرات التوضيحية

2-4 اذا h=h+1 يتم حساب المعادلة الاتية

$$S^a_{x,y} = S^{a-1}_{x,y} - P_{a-1}(P'_{a-1}\ P_{a-1})^{-1} P'_{a-1}\ S^{a-1}_{x,y}$$

3- واخيرا يتم الحصول على مقدرات المعلمات لنموذج الانحدار الخطي كما في الصيغة الاتية

$$\hat{\beta} = R_{p,h}(R'_{h,p}\ S_x R_{p,h})^{-1} R'_{h,p}\ S_{xy}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}'_{q,p}\ \bar{x}$$

## 3 - طرائق المربعات الصغرى الجزئية الحصينة

اذا احتوت بيانات متعدد المتغيرات على قيم شاذه فان طريقة (pls) تكون غير قادرة على التعامل لذا فان افضل علاج هو بالاستناد الى الطرائق الحصينة ، اذ يوجد نوعين لحصانة المربعات الصغرى الجزئية النوع الاول يعتمد على حصانة مصفوفة البيانات مثل( PRM ) والنوع الثاني يعتمد على الانحدار الحصين مثل طريقة (PLSKURSD) .

### 1-3 تقدير M الحصين الجزئي ( Partial Robust M-estimators (PRM)) [17,13]

ان طريقة انحدار M الحصين الجزئي جائت تسميتها بالجزئية كما في انحدار (pls) الكلاسيكية وذلك لان فكرتها هي تخفيض الابعاد باستخدام متغيرات كامنه قليلة حيث يتم ابدال المتغيرات التوضيحية الاصلية بالمتغيرات الكامنة المتعامدة مع اكبر تباين مشترك مع Y . وان الستراتيجية الرئيسية للانحدار الحصين للمربعات الصغرى الجزئية هو تقليل وزن الشواذ اولا ثم يتم التقدير الحصين لمصفوفة التباين والتباين المشترك . للنموذج (1) يخفض الى نموذج انحدار المتغيرات الكامنة كالاتي :

$$y_i = t_i g + \delta_i \qquad \cdots (4)$$

حيث ان $g = (g_1, \cdots g_a)$ : معاملات الانحدار الجديدة و($\delta_i$) : وحد الخطأ .
ان بعد النموذج الجديد هو: a حيث ان (a< m) .
هناك نوعين من الشواذ تؤثر على تقديرات الانحدار هي نقاط الانعطاف (الرفع) points(leverage) وهي لشواذ متعدد المتغيرات في مجال متغيرات الانحدار و الشواذ العمودية (vertical outliers )التي هي ليست ضمن مجال الانحدار ولكن لها بواقي كبيرة . ان طريقة RM تقترح حصانه من هذين النوعين من الشواذ .
ان الاوزان الكلية التي ترجح لكل المشاهدات فتعرف بالصيغة التالية :

$$W_i = \sqrt{W_i^x W_i^r} \qquad \cdots (5)$$

$W_i^x$ : الاوزان المسؤلة عن معالجة نقاط الانعطاف (الرفــع)(leverage points)

$W_i^r$: الاوزان التي تعود الى الشواذ العمودية (vertical outliers)

اما خطوات هذه الخوارزمية فيمكن الرجوع *

## 2-3 طريقة PLSKURSD [16,15,14,7]

تعتمد خوارزمية PLSKURSD على تطبيق خوارزمية SIMPLS على مصفوفة التباين والتباين المشترك الحصين ،ان تقدير مصفوفة التباين يتم بواسطة اسقاط البيانات في بعض الاتجاهات وايجاد القيم الشاذة على هذه الاتجاهات وحذفها من العينة واستعمال البيانات النظيفة لحساب مصفوفة التباين والتباين المشترك. اما خطوات هذه الخوارزمية فانها تكون من ثلاث خطوات* وهي:

1- حساب الاتجاه الذي يعظم معامل التفلطح وكذلك الاتجاه الذي يصغرمعامل التفلطح للاسقاطات.
2- حساب الاتجاه العشوائي باسلوب المعاينة الطبقية ومن ثم البحث عن القيم الشاذه في هذه الاتجاهات
3- ازالة الشواذ من العينة ثم تكرار الخطوات لكي يتم التاكد من عدم وجود الشواذ.

## 3-3 الطريقة المقترحة (MPRM)

هذه الطريقة محورة من الطريقة الرئيسة (PRM) ويمكن تسميتها بتقدير(M الحصين الجزئية المعدلة MPRM) . حيث تم فيها أستعمال مبدأ الموازنة بين المشاهدات والذي يقلل من تاثير الشواذ (ان وجدت ) مما يجعل المقدرات غير شديدة الحساسية بوجود القيم الشاذة ويتم الاعتماد فيهاعلى ايجاد الاوزان الاولية الحصينة والمسؤلة عن معالجة نقاط الانعطاف من المركبات الرئيسية الحصينة بدلا من ايجاد الاوزان من بيانات X والسبب في استخدام المركبات الرئيسية الحصينة هو ان المركبات الرئيسية سوف تتضمن كل المعلومات عن المتغيرات التوضيحية اذ ان المركبة الاولى تعطي اعلى تباين ثم الثانية والثالثة وهكذا وتكون هذه المركبات متعامدة (مستقلة) بعضها عن البعض الاخر بينما البيانات الاصلية تتضمن جميع المعلومات لكنها تكون مرتبطة .
ويمكن تلخيص خطوات الطريقة المعدلة كالاتي:-

1- يتم ايجاد الوسيط متعدد الابعاد(Column-wise median [6]) أو L1-median(وهو مقدر حصين للموقع في متعدد المتغيرات وله خصائص حصينة جيدة تتمثل بخاصية التغاير المتساوي ويكون المقدرمتعامد ويسمى بالوسيط المكاني ا(spatial median) أيضاً والذي تم تعريفه من قبل العالم (Weber 1909 ) بانه اي نقطة لها اقل مجموع لمسافة أقليديس من كل النقاط في مجموعة البيانات ) [9,6].

الوسيط -L1 او (spatial median) الى مجموعة من البيانات X={$x_1, x_2, \cdots, x_n$ } لكل $x_i \in R^p$ هو متجه يرمز له بـ( $\hat{\mu}_{sm}$ ) والذي يحقق أقل مجموع وكالاتي :

$$\hat{\mu}_{sm} = \arg\min_{\eta \in R^P} \sum_{i=1}^{n} \| x_i - \eta \|$$

أذ ان $\|.\|$ يشير الى الاقليديس الطبيعي.

وتوجد عدة خوارزميات لإيجاد ( l1-median) وقد تم الاعتماد على خوارزمية [17] (Hossjer and Croux ) ، اذ يتم ايجاده الى متغيرات (X) ثم يتم طرح الوسيط -L1 من المتغيرات التوضيحية (X).

* لمزيد من الاطلاع انظر (1)

2- نجد المركبات الرئيسية باستعمال خوارزمية (PCA)للبيانات الجديدة.
3- نجد قيم المركبات الرئيسية الحصينة بواسطة تقدير الوسيط -(L1_median) للمركبات الحاصل عليها من الفقرة الثانية وباستعمال مسافة اقليديس الى المركز الحصين نحصل على المسافات المتوسطة gi [17,13] وكالاتي:-

$$gi = \frac{\| pc - \tilde{pc} \|}{median \| pc - \tilde{pc} \|} \quad \cdots (6)$$

أذ أن: pc :- المركبة الرئيسية

$\tilde{pc}$ :- الوسيط(L1_meadian) للمركبات الرئيسية

4- ايجاد اوزان نقاط الانعطاف الى المركبات الرئيسية هي:

$$W_i^{pc} = \frac{1}{(1+ \left| \frac{gi}{4} \right| )^2} \quad \ldots (7)$$

5- اوزان البواقي $Wi^r$ تحسب كما في خوارزمية (PRM) [17,13] حيث يحسب بالصيغة التالية :

$$w^r = f(\frac{r_i}{\hat{\sigma}}, c) = f(hi, c)$$

اذ ان: c قيمة القطع وتكون ثابتة c=4

أذ أن دالة الوزن f تسمى دالة وزن"Fair" وهي واحدة من عدة دوال للوزن[4] منها ( Huber ,Cauchy ,Fair ,Bisquare ,Talworth)

ثم يتم حساب المسافات المتوسطة hi [17,13] و كما يأتي :-

$$hi = \frac{yi - y_{median}}{median \left| yi - y_{median} \right|} \quad \cdots (8)$$

فان اوزان البواقي يمكن ان تحسب كما يلي:

$$W_i^r = \frac{1}{(1+ \left| \frac{hi}{4} \right| )^2} \quad \cdots (9)$$

6- الاوزان الكلية تحسب كالاتي :

$$W_i = \sqrt{W_i^r W_i^{pc}} \quad \ldots (10)$$

7- يتم ترجيح الاوزان الكلية لكل المشاهدات لكي تقلل من التأثير العكسي لوجود الشواذ في أنموذج الانحدار.

8- انجاز PLS الكلاسيكية بخوارزمية (SIMPLS) بأوزان المشاهدات الجديدة وهي ( $W_i x_i$ & $W_i yi$ ) . هذا التحليل ينتج اول تقدير الى معاملات الانحدار g و $t_i$ وهي ( المركبات)لـPLS ولكي تكون النتائج صحيحة يجب ان تقسم (المركبات لـPLS) على الوزن الكلي $W_i$ .

$$t_i = t_i / W_i$$

9- اعادة احتساب اوزان $W_i^r$ للبواقي و $W_i^{pc}$ من نقاط (scores(PLS))والحصول على الاوزان النهائية التي سترجح قيم المشاهدات لكي تقلل من التاثير العكسي للشواذ وكما موضحة بالخطوات التالية:

- حساب البواقي $r_i$ الجديدة كما في الصيغة التالية [17,13]:

$$r_i = y_i - t_i g$$

حساب المسافات المتوسطة hi بواسطة تبديل $y_i$ مع البواقي$r_i$وكالاتي :-

$$hi = \frac{\left| r_i - median\,(r_i) \right|}{MAD(r_1 \ldots r_n)}$$

$$MAD(r_{1\ldots} r_n) = median \left| (r_i - median\,(r_i)) \right|$$

اوزان البواقي $W_i^r$ الجديدة يتم الحصول عليها من تطبيق الصيغة (9)

ولتحديث اوزان الانعطاف ( leverage weights) $w_i^{pc}$ يتم استبدال قيم المركبات الرئيسية في المعادلة (6) بمجموعة النقاط الحالية للمتجه $t_i$ وكالاتي

$$gi = \frac{\left\| t_i - med_{L1}(T) \right\|}{median_i \left\| t_i - med_{L1}(T) \right\|}$$

وبالتالي تطبيق الصيغة (7) ،ثم نحسب الاوزان النهائية $W_i$ كما في الصيغة (10) و يعاد ترجيح البيانات الاصلية للمصفوفة X و Yبالاوزان النهائية المحدثة ثم تطبق خطوات الانحدار لخوارزمية ( Simple) حتى يتم التقارب بمعامل الانحدار g

10- نحصل على معاملات الانحدار( $b_{mprm}$ ) مباشرة من اخر خطوة الى خوارزمية(SIMPLS).

وتطبق هذه الطريقة في حالة عدد المشاهدات اكبر من عدد المتغيرات وبالعكس اي ان (p> n ) أذ يتم أستخدام خوارزمية $(SVD)$ على مصفوفة البيانات 'x، نفرض ان Xمصفوفة ببعد nxp ونفرض ان =t min {n,p}

أذ : $X' = V\ S\ U'$

S :- مصفوفة قطرية ذو بعد txt،عناصر القطرفيها هي t من القيم المفردة الى X

U :- مصفوفة متعامدةذات رتبة nxt ; V :- مصفوفة ذات رتبة pxt

فبدلا من تطبيق خطوات الخوارزمية على Xيتم تطبيقها على البيانات المختزلة ببعد(t x t)

$$\tilde{\chi} = US$$

ان مقدرات انحدار (Mالحصين الجزئي) ( $\hat{\beta}$ ) يجب ان ترجح وتحول الى

$$\hat{\beta} = V\ \tilde{\beta}$$

هذا التقدير هو مكافئ رياضيا الى التقدير المستحصل عليه من تطبيق الخوارزمية مباشرة على المصفوفة الكامله لـ X .

## 4-3 الطريقة المقترحة ( MPLSKURSD)

يتم في هذه الخوارزميه ايجاد مصفوفة التباين والتباين المشترك الحصينةوالتي توظف في خوارزمية (SIMPLS) ، أذ تعد هذه الطريقة كتعديل لطريقة المربعات الصغرى الجزئية الحصينةلخوارزمية(PLSKURSD) حيث ان هذه الطريقة تعتمد على اربع خطوات وكالاتي :

1- في الخطوة الاولى يتم ايجاد قيم $\tilde{X}$ الحصينه وذلك بحساب تقدير الوسيط متعدد الابعاد (L1-median)(كما تم ذكره في المبحث السابق) حيث يتم طرح الوسيط L1 من المتغيرات التوضيحية ( X) .ويتم إيجاد قيم $\tilde{y}$ الحصينة بأستعمال الوسيط الذي يعتبر من المقاييس الحصينة أيضاً حيث يتم طرح الوسيط من المتغير (y) ، ثم تحول البيانات باستعمال المعادلة التالية لمصفوفة البيانات $Z = [\tilde{X}, \tilde{y}]$

يتم أيجاد

$$\tilde{Z}_i = S_z^{-1/2}\ Z_i \qquad , i = 1,2,\ldots,n$$

ثم تكرر نفس خطوات خوارزمية (PLSKURSD) [16,15,14,7] وكما ياتي :

2- في هذه الخطوة يتم إيجادالاتجاه الذي يعظم ويصغر معامل التفرطح للبيانات الحصينة .اول اتجاه نحصل علية من حل الصيغة الآتية :-

$$D1 = \arg\max\ 1/n \sum_{i=1}^{n} (d'\tilde{Z}_i)^4$$
$$s.t \quad d'd = 1$$

حيث 'd الاتجاه الحاصل عليه من تعظيم معامل التفلطح. ونفس العملية تكرر لحساب الاتجاه الذي يصغر معامل التفلطح

$$D2 = \arg\min\ 1/n \sum_{i=1}^{n} (d'\tilde{Z}_i\ )^4$$
$$s.t \quad d'd = 1$$

حيث 'd الاتجاه الحاصل عليه من تصغير معامل التفلطح ، ثم أيجاد أسقاطات القيم الحصينة على هذين الاتجاهين وكالاتي:

$$P_i^{(j)} = d'_j \tilde{z}_i \qquad i = 1,2,\dots,n \quad j = 1,2$$

وثم تحسب قياس الشواذ للمتغير الاحادي او المسافات $r_i^{(j)}$ للمتغير الاحادي الطبيعي كالاتي

$$r_i^{\,j} = 1/B_p \frac{\left|P_i^{(j)} - median_i(P_i^{(j)})\right|}{MAD_i(P_i^{(j)})} \qquad j = 1,2 \qquad \dots(11)$$

وان

$$MAD_i(p_i^{(j)}) = median_i\left|p_i^{(j)} - median_i(p_i^{(j)})\right| \qquad \dots(12)$$

وان MAD :وسيط الانحراف المطلق (Median Absolute Deviation)

و ان قيمة القطع $B_p$ تعتمد على عدد المتغيرات (p) ويتم الحصول عليها من تجارب محاكاة لضمان في حالة غياب الشواذ بان نسبة المشاهدات الصحيحة التي صنفت بانها شواذ تقريبا هي 0.05 . في هذه الخطوة ستعرف القيم الشاذة على انها القيم التي لها $r_i^{(j)} > 1$

3-أيجاد الاتجاه العشوائي وذلك من اسلوب المعاينة الطبقية ومن ثم البحث عن القيم الشاذة في هذه الاتجاهات العشوائية كل اتجاه يولد بخطوتين وهي :-

**الخطوة الاولى:**

يتم اختيار مشاهدتين عشوائيا من العينة ثم تحسب الاتجاه المعرف لهاتين المشاهدتين ( اما ان يكون على اساس المسافة او الفرق بين المشاهدتين) ومن ثم اسقاط المشاهدات على الاتجاه العشوائي وتكرر الى h من المرات حيث h= 1,2, ...,10* p حيث p هو( عدد المتغيرات) .

**الخطوة الثانية:**

في هذه الخطوه يتم بناء مجموعة من k من العينات الطبقية حيث يتم ترتيب الاسقاطات وتقسم الى k من الفترات حيث k تؤخذ اما 3 او 5 تحدد مسبقا وكل فترة سيكون حجمها n/k لكل k من الفترات، 1<k<K، سيتم اختيار عينة جزئية تتكون من p من المشاهدات تنتخب وبدون الارجاع ويحسب الاتجاة dj الذي يستعمل لحساب الشواذ كما في الخطوة الاولى حيث ان الاسقاطات $\tilde{p}_i^{(j)}$ هي:-

$$\tilde{P}_i^{(j)} = \tilde{d}_j \tilde{Z}$$

ومن خلالها سوف تحسب المسافات للمتغير الاحادي الطبيعي $\tilde{r}_i$

$$\tilde{r}_i^{\,j} = 1/B_p \frac{\left|\tilde{P}_i^{(j)} - median_i(\tilde{P}_i^{(j)})\right|}{MAD_i(\tilde{P}_i^{(j)})} \qquad j = 1,2,\dots,h*K$$

4- في هذه الخطوة يتم التدقيق لكل مشاهدة (i) بعد ايجادقياس الشواذ للمتغير الاحادي $r_i$

$$r_i = \max\{\, r_i^1, r_i^2, \tilde{r}_1 \quad \cdots \tilde{r}_i^{\,j} \,\}$$
$$j = 1,2, \quad \cdots h*K$$

والتي تم الحصول عليها بموجب المعادلات (16) ، (17).

فاذا كانت ri > 1 ،فان المشاهدة i تصنف على انها قيمة شاذة وتزال من العينة اذا كان عددهم اصغر من [2/ [n – [ n+ p + 1 عدا ذلك فان المشاهدات التي عددها [ n- [n + p +1]/2] والتي لها اكبر قيم الى ri فهي تعتبر شواذ وتزال من البيانات

واخيرا نجد المجموعة (U) والتي تمثل المشاهدات التي لاتتضمن القيم الشاذه بعدها يتم حساب مسافة مهالنوبس بالاعتماد على المشاهدات الجيدة وكما يلي:-

$$\tilde{m} = \frac{1}{|U|}$$

$$\tilde{S}_z = \frac{1}{|U|-1}\sum_{i\in U}(Z_i - \tilde{m})(Z_i - \tilde{m})'$$

$$V_i = (Z_i - \tilde{m})' \tilde{S}_Z^{-1} (Z_i - \tilde{m})$$

حيث ان $|U|$ تمثل عدد المشاهدات في المجموعة (U). فاذا كانت $V_i < \chi^2_{P-1,0.99}$ فانها لا تصنف على انها قيم شاذه وتوضع في مجموعة U ، وفي حالة $\tilde{S}_z^{-1}$ غير موجودة سوف نستعمل المعكوس العام لها.وهذة الخطوات تكرر حتى لاتبقى قيم شاذه (او تصبح U مجموعة كل البيانات) .
ونستطيع ان نستخدم هذه الطريقة في حالة عدد المتغيرات أكبر من المشاهدات أيضاً بأستعمال خوارزمية(SVD).

## 4- تحديد العدد الامثل للمركبات في PLS [11,10]

ان قرار تحديد العدد الامثل من المركبات (h) في بناء انموذج المربعات الصغرى الجزئية PLS يعتبر قرار مهم وهنالك طرائق متعددة لتحديد العدد الامثل والاكثر شيوعا منها جذر متوسط مربعات الخطا الشرعي والتي تسمى((leave-one-out cross-validation (LOO-cv)). والاسلوب الاخر هو (m-fold cross-validation ) و يكون هذا الاسلوب شائع الاستعمال عندما يكون حجم البيانات كبير اخذين بنظر الاعتبار عدد المشاهدات في مجموعة التدريب لاتقل عن (6-10) مرات بقدر عدد المتغيرات و ان البيانات الاصلية تقسم الى مجموعتين ، مجموعة الاختبار ( test set) والتي تكون مستقلة عن مجموعة التدريب (training set) والمستعمل في تقدير معلمات أنموذج المربعات الصغرى الجزئية.*

## 5- الجانب التجريبي

من خلال اسلوب المحاكاة تمت مقارنة الطرائق الحصينة فضلاً عن طريقة (PLS)مع الطريقتين المقترحتين عن طريق تجربتين

### التجربة الاولى :

بوجود عدة انواع من القيم الشاذه في البيانات وتمت المقارنة بين سلوك الطرق الحصينة من خلال تغيير حجوم العينات وابعاد المتغيرات حيث تم توليد العينات بحجوم (n=30,50,200) وعدد المتغيرات التوضيحية(p=5) في حالة n>P اما في حالة p>nفقد تم توليد العينات بالحجم n=20 وعدد المتغيرات التوضيحية (p=40,60, 150). وكان عدد المركبات ( 2) في الحالتين اما مصفوفة التباين الى t (مصفوفة القياس) فهي قطرية diag[6,4] وتم استعمال نسب تلوث ϵ مساوية الى (10%, 20%, 30%).في هذة التجربة سنقارن بين اربعة انواع من الشواذ بالاضافة الى المشاهدات الاعتيادية. بالاعتماد على النموذج الابتدائي(2, 3 ) وعلى فرض ان الوسط الحسابي يساوي صفر حيث النموذج يتبع التوزيع الطبيعي وغير المرتبط تم توليد البيانات وفقا لمايلي: [10,7]

$$t \propto N_a(0_a, \Sigma_t)$$
$$x = I_{p,a} t + N_p(0_p, 0.1\, I_p)\ , \quad p > a$$
$$y = q'\ t + N(0,1)$$

و $I_{p,a}=1$ عندما i=j و $I_{p,a}=0$ ماعدا ذلك .

q متجه من الواحدات ذو بعد ax1 : $I_p$مصفوفة الوحدة ذو بعد p x p

اما بالنسبة للبيانات الملوثة فقد تم توليد مجموعة تتكون ]100*(ϵ -1) ]من المشاهدات وفقا للأنموذج اعلاه وتم اضافة [100*ϵ] من المشاهدات الملوثة التي تولد من خلال نماذج التلوث ادناه لتكون انواع التلوث التالية :

1- نقاط الرفع (انعطاف)غير جيدة (bad Leverage points)

$$x_\varepsilon = I_{p,a} t_\varepsilon + N_p(0_p, 0.001 I_p)$$
$$t_\varepsilon \propto N_a(10_a, \Sigma_t)$$

-1

2- الشواذ العمودية ( Vertical outliers)

$$y_\varepsilon = q't + N(30, 0.3)$$

3- نقاط انعطاف جيدة (good Leverage points)

$$x_\varepsilon = t_\varepsilon I_{a,p} + N_p((0_a, 3_{p-a}), 0.01 I_p)$$
$$t_\varepsilon \propto N_a(3_a, \Sigma_t)$$
$$y_\varepsilon = q't + N(30, 0.3)$$

* لمزيد من الاطلاع انظر(1)

## 4- الشواذ المركزة (concentrated outliers)

$$x_{\varepsilon} = I_{a,p} t_{\varepsilon} + N_p (1_p , 0.001 I_p)$$
$$t_{\varepsilon} \propto N_a (1_a , \Sigma_t)$$

**لاجل المقارنة بين متجه المعلمات β ومتجه المعلمات المقدر $\hat{\beta}_a^{(l)}$ تم أستعمال المقياس الاتي [7] :**

$$MSE_a(\hat{\beta}) = 1/m \sum_{L=1}^{m} \left\| \hat{\beta}_a^{(L)} - \beta \right\|^2 = Norm(\hat{\beta})$$

**حيث || || يشيرالى اقليديس الطبيعي و m يشير الى عدد التكرارات فقد تم تكرار التجربة 1000 .**

## التجربة الثانية :

**تم توليد الخطأ على اساس توزيعات اخرى بالاضافة الى التوزيع الطبيعي وبحجوم عينات وابعاد متغيرات مختلفة حيث تم توليد العينات,p=20,40,60,100,200 (n=10) و(p=100,n=25) و (p=200,n=50) وبأستعمال الأنموذج الابتدائي الأتي: [7]**

$$y_m = X\beta + \varepsilon_m$$

**$\beta$ : متجه معاملات الانحدار الذي يولد على أساس التوزيع الطبيعي القياسي**

**$\varepsilon_m$ حد الخطأ العشوائي الذي يولد حسب التوزيعات التالية : التوزيع الطبيعي القياسي (standard normal)، توزيع (tstudents t) بدرجة حرية 5، كوشي (Cauchy)، لابلاس (laplace). بوجود تعدد خطي بين المتغيرات التوضيحية والذي يمثل كما يلي:**

$$X = TR'$$

**حيث T مصفوفة القياس ذو بعد (nxa).**
**R مصفوفة ذو بعد (pxa)والاثنان يولدان من التوزيع الطبيعي القياسي.**
**وقد كررت التجربة 1000 مرة للحصول على الدقة العالية، و قد نفذت المحاكاة من خلال برامج كتبت بلغة ماتلاب (version 7.12.0.635)(R2011a)**

## 5-1 نتائج تجربة المحاكاة الاولى في حالة n> p

**من ملاحظة الجداول (1) و(2) و(3) نستطيع ان نلخص النسب المئوية لمتوسط مربعات الخطأ (MSE) للمعلمات المقدرة للخوارزميات المذكورة اعلاه في الجدول كالاتي :**

**جدول يوضح النسب المئوية لمتوسط مربعات الخطأ (MSE) لمتجهات المعلمات المقدرة للخوارزميات عند انواع القيم الشاذة المختلفة**

| الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|
| لاتوجد شواذ | 0 | 0.67 | .33 | 0 | 0 |
| الشواذ العمودية | 0 | 0.45 | 0.45 | 0.1 | 0 |
| نقاط الرفع الجيدة | 0 | 0 | 0 | 0.56 | 0.44 |
| الشواذ المركزة | 0 | 0.33 | 0.67 | 0 | 0 |
| نقاط الرفع السيئة | 0 | 0.78 | 0.22 | 0 | 0 |

**يتضح من الجدول اعلاه ان الطريقة (PRM)هي الافضل في حالة عدم وجود الشواذ ونقاط الرفع السيئة اما الطريقة المقترحة(MPRM)فقدكانت متساوية مع الطريقة (PRM)في حالة الشواذ العمودية والافضل في الشواذ المركزة اما طريقة(PLSKUR)هي الافضل عند نقاط الرفع الجيدة ثم تليها الطريقة المقترحة (MPLSLUR)حيث كان التفاوت قليل.**

## 5- 2 نتائج تجربة المحاكاة الاولى في حالة p>n

**من ملاحظة الجداول (4) و(5) و(6) نستطيع ان نلخص النسب المئوية لمتوسط مربعات الخطأ (MSE) للمعلمات المقدرة للخوارزميات المذكورة اعلاه في الجدول كالاتي :**

**جدول يوضح النسب المئوية لمتوسط مربعات الخطأ (MSE) لمتجه المعلمات المقدرة للخوارزميات**

| الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|
| لاتوجد شواذ | 0 | 0 | 0 | 0 | 1 |
| نقاط الرفع السيئة | 0 | 0.33 | 0.33 | 0 | 0.33 |
| الشواذ العمودية | 0 | 0.22 | 0.44 | 0 | 0.33 |
| نقاط الرفع الجيدة | 0 | 0 | 0 | 0 | 1 |
| الشواذ المركزة | 0 | 0.11 | 0.33 | 0 | 0.56 |

- يتضح من الجدول اعلاه في حالة عدم وجود الشواذ ونقاط الرفع الجيدة والشواذ المركزة فان الطريقة المقترحة (MPLSKUR) هي الافضل اما في حالة الشواذ العموديه فنجد بان الطريقة المقترحة (MPRM) كانت الافضل تليها (MPLSKUR) ثم طريقة PRMاما في حالة نقاط الرفع السيئة فنلاحظ بان الطرائق (MPRM,PRM و MPLSKUR) كانت متساوية

## 3-5 نتائج تجربة المحاكاة الثانية :

يتضح من الجداول من(7) الى(13) بان متوسط مربعات الخطأ (MSE) ولجميع الطرائق ولكافة توزيعات الاخطاء عند المركبات (h=1,2 ,3) ولجميع الحجوم والابعاد متساوي تقريباً وهذا دليل على ان الدقة الاحصائية لكافة الطرائق متساوية تقريباً في حالة ان الاخطاء تتوزع توزيع معين.وبهذا من جميع الجداول السابقة نستيطع ان نلخص النسب المئوية لمتوسط مربعات الخطأ MSEلمتجه المعلمات المقدرةللخوارزميات المذكورة اعلاه في الجدول الاتي :

جدول يوضح النسب المئوية لمتوسط مربعات الخطأ (MSE) لمتجه المعلمات المقدرة للخوارزميات وتوزيعات مختلفة للاخطاء

| الطرائق / التوزيعات | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|
| Normal | 0.275 | 0.225 | 0.225 | 0.15 | 0.125 |
| Laplace | 0.19 | 0.29 | 0.26 | 0.12 | 0.14 |
| $t_5$ | 0.136 | 0.295 | 0.25 | 0.159 | 0.159 |
| Cauchy | 0.026 | 0.5 | 0.474 | 0 | 0 |

من الجدول يتضح ان طريقة (PLS) هي الافضل في حالة ان الاخطاء تتوزع توزيع (normal) وان طريقة (PRM) هي الافضل عندما الاخطاء تتوزع توزيع (laplaceو $t_5$ وCauchy) وان الطريقة المقترحة (MPRM) كانت مقاربة لنتائجها ثم جاءت الطريقتان MPLSKUR و PLSKUR على التوالي.

## 1-6الاستنتاجات

1- تم التحقق من كفاءة الصيغ المقترحة والمتعلقة بحساب تقدير المعلمات ولبعض انواع الشواذ حيث اثبتت الطريقة المقترحه(MPRM)كفائتها وذلك لحصولها على اقل ( $MSE_a$ ( $\hat{\beta}$ عند استعمال اسلوب المحاكاة الاول ولابعاد متغيرات وحجوم عينات مختلفة وفي حالة عدد المشاهدات أكبر من عدد المتغيرات لبعض انواع الشواذ وخصوصا عند الشواذ المركزه اما عند الشواذ العمودية فهي كانت متساويه مع طريقة (PRM)عندما ابعاد المتغيرات وحجوم العينات مختلفة ، وكانت هي الافضل عندما عدد المتغيرات اكبر من عدد المشاهدات اما في حالة عدم وجود الشواذ ونقاط الرفع السيئة كانت (PRM) الافضل،وعند نقاط الرفع الجيدةكانت(PLSKURSD) هي الافضل عندما ابعاد المتغيرات وحجوم العينات مختلفة.
2- - اما في حالة عدد المتغيرات اكبر من المشاهدات فقد اثبتت الطريقة المقترحه الحصينة الثانية (MPLSKURSD)كفائتها في تقدير معالم النموذج الخطي حيث حصلت على اقل( $MSE_a$ ( $\hat{\beta}$مقارنة ببعض الطرق الحصينة الاخرى فهي كانت الافضل في حالة عدم وجود الشواذ ونقاط الرفع الجيدة والشواذ المركزة، وكانت متساوية مع الطرائق(MPRM، PRM) عند نقاط الرفع السيئة والشواذ العمودية .
3- في حالة تجربة المحاكاة الثانية اتضح ان طريقة (PLS) هي الافضل عندما تتوزع الاخطاء توزيع (normal)وان طريقة (PRM) هي الافضل عندما تتوزع الاخطاء توزيع(laplace و t5وcauchy)واوضحت النتائج بان الطريقة المقترحة (MPRM) تكافئ تقريبا الطريقة الحصينة(PRM)وكذلك الطريقة المقترحة الثانية (MPLSKURSD) تكافئ تقريبا الطريقة الحصينة(PLSKURSD)ولجميع انواع توزيعات الاخطاء وعند جميع ابعاد المتغيرات وحجوم العينات المختلفة.
4- أظهرت النتائج في حالة استعمال اسلوب المحاكاة الثاني وعندما حجم العينة صغير وعدد المتغيرات اكبر من المشاهدات بان الفروقات بين الخوارزميات تكاد تكون متقاربه.
5- أظهرت النتائج بأن قيم متوسط مربعات الخطأ( $MSE_a$ ( $\hat{\beta}$تتناقص بزيادة حجوم العينات وتتزايد في حالة عدد المتغيرات أكبر من عدد المشاهدات.

## 2-6 التوصيات

بناءً على ماتم التوصل اليه من استنتاجات للجانب النظري نوصي بما ياتي:

1- دراسة نماذج متعدد المتغيرات (في حالة q>1 ) للطرق الحصينة الى(PLS).
2- دراسة الخوارزميات اخرى الى(PLS) غير خوارزمية (SIMPLS) واجراء مقارنات بين هذه الخوارزميات من خلال استعمالها في الطرق الحصينة الاخرى.
3- استعمال الطرق الحصينة المقترحة لطريقة المربعات الصغرى الجزئية في حالة احتواء البيانات على قيم شاذة وذلك للحصول على تقديرات مناسبة.

## 7- المصادر


1- البكري ، رباب عبد الرضا صالح ( 2013) " بعض الطرائق الحصينة لمعالجة مشكلة التعدد الخطي في النماذج الخطية مع تطبيق عملي"اطروحة دكتوراه ، كلية الادارة والاقتصاد ،جامعة بغداد.
2- المندلاوي ،ميسون علي رحمن 2007"مقارنة بعض الطرائق الحصينة للمربعات الصغرى الجزئية " رسالة ماجستير ،كلية الادارة والاقتصاد ،جامعة بغداد.

3- Branden ,K.V.&Hubert,M., (2003) "Robustness Properties of arobust PLS regression method"December10.J.AnalyticaChimica p229-41.
4- Cummins,D.JandAndrews,C.W., (1995) "Iteratively Reweighted Partial Least Squares: Aperformance Analysis by Monte Carlo simulation" journal of chemometrics, vol.9, 489-507
5- DeJong, S.,1993."SIMPLS:an alternative approach to partial least squares regression"Chemometrics and Intelligent Laboratory Systems, 18: 251–26 3.
6- Fritz,H., Filzmoser,P. , Croux,C., 2010" A comparison of algorithms for the multivariate l1_median"http://www.statistik.tuwien .ac.at
7- Gonzales, J., & et al. (2009) "A robust partial least squares regression method with application" J . Chemometrics; 23:78-90
8- GriepM.I., Wakeling,I.N., Vankeerberghen,P., Massart,.D.L.,(1995) "Comparison of semirobust and robust partial least squaresprocedures"chemometrics and intelligent laboratory systems 29 - 37-50 .
9- Hossjer and Croux (1995) "Generalizing Univariate Signed Rank Statistics for Testing and Estimating a Multivariate Location Parameter",Non-parametric Statistics, 4, 293.
10- Hubert ,M. &Branden ,K.V., (2003) "Robust Methods for Partial Least Squares Regression "9th ofOctober J. chemometrics p 537-549.
11- Hubert, M. and Engelen, S.(2006)"Fast Cross - Validation Of High-Breakdown Resampling Methods For PCA "J.ComputationalStatistics & Data Analysis
12- Juan,A.G. and Rosario,R., (1998) "On Robust Partial Least Squares (PLS) Methods " j. chemometrics 12,365-378
13- Lieebmann ,B., Filzmoser,P. and Varmuza ,K.,(2010) "Robust and classical PLS regression compared", J.chemometrics
14- Pena,D. and Prieto,F.J., (1997) "Robust Covariance Matrix Estimation and Multivariate Outlier Detection" working paper 97-08, statistics and Econometrics series 04
15- Pena,D. and Prieto,F.J.,(2001) "Multivariate Outlier Detection and Robust Covariance Matrix Estimation" Technometrics,vol.43,no.3
16- Pena,D., and Prieto,F.J., (2007)"Combining Random and Specific Direction for Outlier Detection and Robust Estimation in High-Dimensional Multivariate Data" journal of computational and graphical statistics, ,vol.16, number 1,pages 228-254
17- Serneels,s., Croux, C. , Filzmoser, P., Espen,PJ.V. , (2005) "Partial Robust M Regression" ,J.Chemometrics and Intelligent Laboratory Systems.
18- Yuditskaya,s., (2010)" An Overview of Methods in Linear Least-Squares Regression"pattern recognition and Analysis ,November 4 .

جدول (1)

يوضح قيم MSE لمتجه المعلمات المقدره عند ( n=30) و ( p=5) عند نسب التلوث (0.3,0.2,0.1)

| نسب التلوث | الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|---|
| 0% | لاتوجد شواذ | 3.93198 | 3.128095 | 3.191438 | 6.705675 | 6.29413 |
| 10% | نقاط الرفع السيئة | 2.604226 | 2.410823 | 2.425544 | 5.464185 | 6.312738 |
| | الشواذ العمودية | 17.85582 | 3.138295 | 3.261658 | 6.683227 | 7.938724 |
| | نقاط الرفع الجيدة | 26.62007 | 9.639245 | 9.626159 | 6.15471 | 6.179438 |
| | الشواذ المركزة | 3.536876 | 2.965567 | 2.99684 | 5.502633 | 5.878988 |
| 20% | نقاط الرفع السيئة | 2.64198 | 2.377776 | 2.395004 | 4.516184 | 4.702131 |
| | الشواذ العمودية | 30.993430.993 | 4.14967 | 4.070025 | 11.32486 | 9.957953 |
| | نقاط الرفع الجيدة | 28.49926 | 25.91649 | 25.93971 | 6.120336 | 5.882657 |
| | الشواذ المركزة | 2.977922 | 2.813003 | 2.817864 | 5.223898 | 5.066862 |
| 30% | نقاط الرفع السيئة | 2.52037 | 2.323835 | 2.325562 | 3.10733 | 2.977724 |
| | الشواذ العمودية | 52.44824 | 8.975385 | 9.178658 | 28.51931 | 26.12804 |
| | نقاط الرفع الجيدة | 29.05623 | 27.27656 | 27.3146 | 6.703405 | 7.145314 |
| | الشواذ المركزة | 2.961806 | 2.721444 | 2.70841 | 3.968517 | 4.030898 |

جدول (2)

يوضح قيم MSE لمتجه المعلمات المقدرة عند ( n=50) و ( p=5) عند نسب التلوث0.3,0.2,0.1

| نسب التلوث | الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|---|
| 0% | لاتوجد شواذ | 3.143267 | 2.760466 | 2.730027 | 4.455928 | 4.145776 |
| 10% | نقاط الرفع السيئة | 2.484038 | 2.287489 | 2.295501 | 4.48765 | 4.572959 |
| | الشواذ العمودية | 12.74727 | 2.871288 | 2.865137 | 4.238654 | 4.396087 |
| | نقاط الرفع الجيدة | 27.27451 | 11.72146 | 11.75841 | 4.212243 | 4.217679 |
| | الشواذ المركزة | 2.885578 | 2.698759 | 2.668627 | 4.211664 | 4.135587 |
| 20% | نقاط الرفع السيئة | 2.442899 | 2.305609 | 2.304643 | 3.819094 | 4.005598 |
| | الشواذ العمودية | 22.78322 | 3.483506 | 3.41411 | 4.551482 | 4.438952 |
| | نقاط الرفع الجيدة | 29.07668 | 26.89553 | 26.88556 | 4.707521 | 4.964694 |
| | الشواذ المركزة | 2.736025 | 2.628254 | 2.619497 | 4.290445 | 4.301218 |
| 30% | نقاط الرفع السيئة | 2.383217 | 2.297833 | 2.314675 | 2.729035 | 2.687254 |
| | الشواذ العمودية | 30.0315 | 6.204013 | 6.660494 | 10.47879 | 11.1865 |
| | نقاط الرفع الجيدة | 29.26588 | 27.7416 | 27.74081 | 4.551511 | 4.448104 |
| | الشواذ المركزة | 2.605533 | 2.500286 | 2.505283 | 3.579481 | 3.295105 |

جدول (3)

يوضح قيم MSE لمتجه المعلمات المقدره عند ( n=200) و ( p=5) عند نسب التلوث0.3,0.2,0.1

| نسب التلوث | الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|---|
| 0% | لاتوجد شواذ | 2.468142 | 2.332279 | 2.337385 | 2.498064 | 2.504395 |
| 10% | نقاط الرفع السيئة | 2.2636 | 2.223658 | 2.225356 | 2.587775 | 2.586279 |
| | الشواذ العمودية | 4.565388 | 2.378945 | 2.407979 | 2.555479 | 2.553579 |
| | نقاط الرفع الجيدة | 27.33878 | 13.28781 | 13.37544 | 2.570506 | 2.567473 |
| | الشواذ المركزة | 2.334942 | 2.297701 | 2.291472 | 2.570749 | 2.573741 |
| 20% | نقاط الرفع السيئة | 2.248788 | 2.224955 | 2.219996 | 2.553724 | 2.551361 |
| | الشواذ العمودية | 7.58146 | 2.568577 | 2.540077 | 2.602076 | 2.605444 |
| | نقاط الرفع الجيدة | 29.00699 | 26.79536 | 26.79755 | 2.623537 | 2.621256 |
| | الشواذ المركزة | 2.31701 | 2.291171 | 2.287811 | 2.607924 | 2.600909 |
| 30% | نقاط الرفع السيئة | 2.266568 | 2.215074 | 2.216326 | 2.258193 | 2.341524 |
| | الشواذ العمودية | 9.250026 | 3.045639 | 3.016626 | 2.688 | 2.748775 |
| | نقاط الرفع الجيدة | 29.69553 | 28.13204 | 28.13435 | 3.007342 | 3.29134 |
| | الشواذ المركزة | 2.318153 | 2.274656 | 2.274513 | 2.356786 | 2.365914 |

جدول (4)

يوضح قيم MSE لمتجه المعلمات المقدرة عند ( n=20) و ( p=40) عند نسب التلوث0.1,0.2 ,0.3

| نسب التلوث | الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|---|
| 0% | لاتوجد شواذ | 90.76676 | 83.99906 | 83.82995 | 83.67835 | 83.62133 |
| 10% | نقاط الرفع السيئة | 83.28603 | 80.40058 | 80.42986 | 80.7259 | 80.72827 |
| | الشواذ العمودية | 203.7097 | 86.25143 | 85.81304 | 130.7868 | 129.3845 |
| | نقاط الرفع الجيدة | 76.85447 | 77.59814 | 77.48883 | 76.39718 | 76.36648 |
| | الشواذ المركزة | 84.45805 | 81.77212 | 81.75398 | 81.15005 | 81.07931 |
| 20% | نقاط الرفع السيئة | 38.84241 | 36.40377 | 36.47896 | 37.1319 | 37.08395 |
| | الشواذ العمودية | 255.2324 | 50.69596 | 50.49213 | 121.5665 | 119.8704 |
| | نقاط الرفع الجيدة | 43.28528 | 42.9403 | 42.90585 | 42.79096 | 42.77652' |
| | الشواذ المركزة | 38.18807 | 36.89767 | 36.90983 | 36.98811 | 36.95709 |
| 30% | نقاط الرفع السيئة | 50.01483 | 47.60157 | 47.55851 | 48.6485 | 48.54654 |
| | الشواذ العمودية | 399.5773 | 174.342 | 175.4223 | 163.3382 | 159.6765 |
| | نقاط الرفع الجيدة | 51.60886 | 51.35346 | 51.36904 | 51.29083 | 51.2872 |
| | الشواذ المركزة | 48.3837 | 47.52974 | 47.50163 | 48.0536 | 47.9966 |

جدول (5)

يوضح قيم MSE لمتجه المعلمات المقدرة عند ( n=20) و ( p=60) عند نسب التلوث0.1,0.2 ,0.3

| نسب التلوث | الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|---|
| 0% | لاتوجد شواذ | 66.45336 | 58.97368 | 58.92541 | 58.18977 | 58.18797 |
| 10% | نقاط الرفع السيئة | 59.18198 | 56.12499 | 56.08824 | 55.7096 | 55.62378 |
| | الشواذ العمودية | 174.4606 | 60.80002 | 60.67636 | 95.6495 | 95.23426 |
| | نقاط الرفع الجيدة | 58.34411 | 59.85868 | 59.83663 | 57.55186 | 57.5446 |
| | الشواذ المركزة | 59.7887 | 56.97345 | 56.90954 | 55.82944 | 55.73347 |
| 20% | نقاط الرفع السيئة | 50.58159 | 47.94009 | 47.96561 | 48.2731 | 48.1673 |
| | الشواذ العمودية | 281.1992 | 59.44168 | 60.11083 | 123.3543 | 122.2643 |
| | نقاط الرفع الجيدة | 52.75271 | 52.3188 | 52.37399 | 51.99841 | 51.98885 |
| | الشواذ المركزة | 49.30511 | 48.11332 | 48.09967 | 47.91679 | 47.84368 |
| 30% | نقاط الرفع السيئة | 74.22322 | 71.43377 | 71.41488 | 72.46761 | 72.39801 |
| | الشواذ العمودية | 421.7863 | 207.2473 | 208.7589 | 175.031 | 173.1078 |
| | نقاط الرفع الجيدة | 74.58289 | 74.18248 | 74.1554 | 74.13526 | 74.10261 |
| | الشواذ المركزة | 72.52221 | 71.45205 | 71.42188 | 71.75499 | 71.66955 |

جدول (6)

يوضح قيم MSEلمتجه المعلمات المقدرة عند( n=20) و( p=150) عند نسب التلوث0.1,0.2 ,0.3

| نسب التلوث | الطرائق / الشواذ | PLS | PRM | MPRM | PLSKUR | MPLSKUR |
|---|---|---|---|---|---|---|
| 0% | لاتوجد شواذ | 147.6072 | 142.1276 | 142.0823 | 141.1587 | 141.1408 |
| 10% | نقاط الرفع السيئة | 143.8799 | 140.8549 | 140.887 | 140.1985 | 140.1638 |
| | الشواذ العمودية | 230.5816 | 142.828 | 142.8537 | 169.0253 | 168.9759 |
| | نقاط الرفع الجيدة | 143.02 | 143.0405 | 143.1292 | 141.3591 | 141.3234 |
| | الشواذ المركزة | 142.4376 | 141.1316 | 141.1782 | 139.6625 | 139.6232 |
| 20% | نقاط الرفع السيئة | 169.9366 | 167.3754 | 167.4046 | 167.4479 | 167.3653 |
| | الشواذ العمودية | 339.5476 | 172.7895 | 172.3558 | 215.9492 | 215.4868 |
| | نقاط الرفع الجيدة | 167.7448 | 166.7224 | 166.7445 | 166.3475 | 166.3294 |
| | الشواذ المركزة | 167.9935 | 166.9792 | 166.972 | 166.5497 | 166.5092 |
| 30% | نقاط الرفع السيئة | 158.4073 | 156.0714 | 156.0699 | 156.3259 | 156.265 |
| | الشواذ العمودية | 407.2903 | 260.7705 | 257.9959 | 218.4858 | 218.4827 |
| | نقاط الرفع الجيدة | 155.5631 | 154.491 | 154.4779 | 154.501 | 154.4384 |
| | الشواذ المركزة | 156.4661 | 155.3434 | 155.3383 | 155.3627 | 155.3524 |

جدول (7 )

يوضح قيم (MSE) لمتجه المعلمات المقدرة عند( n=10)و( p=20) ولطرائق مختلفةو لتوزيعات مختلفة للاخطاء وعند اعداد المركبات h=1,2,3

| H | Noise models | Pls | Prm | mprm | Plskusd | Mplskursd |
|---|---|---|---|---|---|---|
| 1 | Normal | 11.188 | 11.183 | 11.183 | 11.192 | 11.192 |
| | Laplace | 11.179 | 11.184 | 11.184 | 11.179 | 11.179 |
| | $T_5$ | 11.193 | 11.174 | 11.174 | 11.212 | 11.212 |
| | Cauchy | 11.173 | 11.208 | 11.208 | 11.177 | 11.177 |
| 2 | Normal | 11.077 | 11.075 | 11.075 | 11.077 | 11.077 |
| | Laplace | 11.074 | 11.075 | 11.752 | 11.074 | 11.074 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | $T_5$ | 11.084 | 11.079 | 11.079 | 11.083 | 11.083 |
| | Cauchy | 11.134 | 11.078 | 11.078 | 11.125 | 11.125 |
| 3 | Normal | 9.783 | 9.791 | 9.791 | 9.793 | 9.793 |
| | Laplace | 9.800 | 9.762 | 9.762 | 9.813 | 9.813 |
| | $T_5$ | 9.850 | 9.829 | 9.829 | 9.866 | 9.866 |
| | Cauchy | 11.045 | 9.831 | 9.831 | 11.009 | 11.009 |

**جدول (8)**

**يوضح قيم(MSE) لمتجه المعلمات المقدرة عند( n=10) و( p=40) ولطرائق مختلفةو لتوزيعات مختلفة للاخطاء وعند اعداد المركبات h=1,2,3**

| H | Noise models | Pls | Prm | mprm | Plskusd | Mplskursd |
|---|---|---|---|---|---|---|
| 1 | Normal | 34.362 | 34.363 | 34.363 | 34.363 | 34.363 |
| | Laplace | 34.385 | 34.377 | 34.377 | 34.384 | 34.384 |
| | $T_5$ | 34.362 | 34.363 | 34.363 | 34.362 | 34.362 |
| | Cauchy | 34.706 | 34.397 | 34.397 | 34.745 | 34.745 |
| 2 | Normal | 38.135 | 38.136 | 38.136 | 38.136 | 38.136 |
| | Laplace | 37.994 | 37.993 | 37.993 | 37.997 | 37.997 |
| | $T_5$ | 37.928 | 37.929 | 37.929 | 37.930 | 37.930 |
| | Cauchy | 34.314 | 34.140 | 34.140 | 34.477 | 34.477 |
| 3 | Normal | 34.089 | 34.091 | 34.091 | 34.089 | 34.089 |
| | Laplace | 34.118 | 34.127 | 34.127 | 34.123 | 34.123 |
| | $T_5$ | 34.083 | 34.089 | 34.089 | 34.083 | 34.083 |
| | Cauchy | 34.894 | 34.149 | 34.150 | 34.887 | 34.887 |

**جدول (9)**

**يوضح قيم(MSE) لمتجه المعلمات المقدرة عند( n=10) و( p=60) ولطرائق مختلفةو لتوزيعات مختلفة للاخطاء وعند اعداد المركبات h=1,2,3**

| H | Noise models | Pls | Prm | mprm | Plskusd | Mplskursd |
|---|---|---|---|---|---|---|
| 1 | Normal | 53.171 | 53.177 | 53.177 | 53.170 | 53.170 |
| | Laplace | 53.178 | 53.168 | 53.168 | 53.179 | 53.179 |
| | $T_5$ | 53.175 | 53.180 | 53.180 | 53.174 | 53.174 |
| | Cauchy | 53.346 | 53.170 | 53.170 | 53.493 | 53.493 |
| 2 | Normal | 53.087 | 53.087 | 53.087 | 53.087 | 53.087 |
| | Laplace | 53.094 | 53.094 | 53.094 | 53.095 | 53.095 |
| | $T_5$ | 53.087 | 53.087 | 53.087 | 53.087 | 53.087 |
| | Cauchy | 53.133 | 53.105 | 53.105 | 53.181 | 53.181 |
| 3 | Normal | 53.087 | 53.091 | 53.091 | 53.093 | 53.093 |
| | Laplace | 53.105 | 53.102 | 53.102 | 53.108 | 53.108 |
| | $T_5$ | 53.089 | 53.094 | 53.094 | 53.112 | 53.112 |
| | Cauchy | 53.493 | 53.125 | 53.125 | 53.457 | 53.457 |

**جدول (10)**

**يوضح قيم (MSE) لمتجه المعلمات المقدرة عند( n=10) و( p=80) ولطرائق مختلفةو لتوزيعات مختلفة للاخطاء وعند اعداد المركبات h=1,2,3**

| H | Noise models | Pls | Prm | mprm | Plskusd | Mplskursd |
|---|---|---|---|---|---|---|
| 1 | Normal | 65.492 | 65.492 | 65.492 | 65.492 | 65.492 |
| | Laplace | 65.513 | 65.523 | 65.523 | 65.513 | 65.513 |
| | $T_5$ | 65.492 | 65.492 | 65.492 | 65.491 | 65.491 |
| | Cauchy | 65.873 | 65.564 | 65.564 | 65.869 | 65.869 |
| 2 | Normal | 65.000 | 64.999 | 64.999 | 65.001 | 65.001 |
| | Laplace | 65.008 | 65.002 | 65.002 | 65.008 | 65.008 |
| | $T_5$ | 65.001 | 65.000 | 65.000 | 65.001 | 65.001 |
| | Cauchy | 65.208 | 65.005 | 65.005 | 65.205 | 65.205 |
| 3 | Normal | 64.409 | 64.406 | 64.406 | 64.409 | 64.409 |
| | Laplace | 64.404 | 64.404 | 64.404 | 64.403 | 64.403 |
| | $T_5$ | 64.423 | 64.415 | 64.415 | 64.423 | 64.423 |
| | Cauchy | 64.455 | 64.408 | 64.408 | 64.452 | 64.452 |

جدول (11)

يوضح قيم(MSE) لمتجه المعلمات المقدرة عند( n=10) و( p=100) ولطرائق مختلفةو لتوزيعات مختلفة للاخطاء وعند اعداد المركبات h=1,2,3

| H | Noise models | Pls | Prm | mprm | Plskusd | Mplskursd |
|---|---|---|---|---|---|---|
| 1 | Normal | 74.920 | 74.921 | 74.921 | 74.921 | 75.140 |
| | Laplace | 74.858 | 74.858 | 74.858 | 74.858 | 75.041 |
| | $T_5$ | 74.861 | 74.862 | 74.862 | 74.863 | 75.076 |
| | Cauchy | 74.863 | 74.858 | 74.858 | 74.865 | 74.941 |
| 2 | Normal | 74.853 | 74.853 | 74.853 | 74.853 | 74.874 |
| | Laplace | 74.853 | 74.862 | 74.862 | 74.854 | 74.870 |
| | $T_5$ | 74.857 | 74.853 | 74.853 | 74.859 | 74.877 |
| | Cauchy | 74.896 | 74.876 | 74.876 | 74.891 | 74.883 |
| 3 | Normal | 74.271 | 74.272 | 74.272 | 74.272 | 74.562 |
| | Laplace | 74.270 | 74.272 | 74.272 | 74.271 | 74.700 |
| | $T_5$ | 74.288 | 74.283 | 74.283 | 74.295 | 74.509 |
| | Cauchy | 74.281 | 74.277 | 74.277 | 74.288 | 76.116 |

جدول (12)

يوضح قيم(MSE) لمتجه المعلمات المقدرة عند( n=25)و( p=100) ولطرائق مختلفةو لتوزيعات مختلفة للاخطاء وعند اعداد المركبات h=1,2,3

| H | Noise models | pls | Prm | mprm | Plskusd | Mplskursd |
|---|---|---|---|---|---|---|
| 1 | Normal | 74.918 | 74.919 | 74.919 | 74.918 | 74.919 |
| | Laplace | 74.918 | 74.918 | 74.918 | 74.918 | 74.918 |
| | $T_5$ | 74.860 | 74.859 | 74.859 | 74.860 | 74.860 |
| | Cauchy | 74.859 | 74.859 | 74.859 | 74.861 | 74.861 |
| 2 | Normal | 74.853 | 74.853 | 74.853 | 74.853 | 74.853 |
| | Laplace | 74.852 | 74.853 | 74.853 | 74.853 | 74.852 |
| | $T_5$ | 74.853 | 74.853 | 74.853 | 74.853 | 74.853 |
| | Cauchy | 74.876 | 74.854 | 74.854 | 74.875 | 74.880 |
| 3 | Normal | 74.272 | 74.270 | 74.270 | 74.272 | 74.272 |
| | Laplace | 74.270 | 74.270 | 74.270 | 74.271 | 74.271 |
| | $T_5$ | 74.273 | 74.271 | 74.271 | 74.273 | 74.273 |
| | Cauchy | 74.465 | 74.274 | 74.274 | 74.472 | 74.484 |

جدول (13)

يوضح قيم(MSE) لمتجه المعلمات المقدرة عند( n=50) و( p=200) ولطرائق مختلفةو لتوزيعات مختلفة للاخطاء وعند اعداد المركبات h=1,2,3

| H | Noise models | pls | Prm | mprm | Plskusd | Mplskursd |
|---|---|---|---|---|---|---|
| 1 | Normal | 164.868 | 164.868 | 164.868 | 164.868 | 164.868 |
| | Laplace | 164.797 | 164.797 | 164.797 | 164.797 | 164.797 |
| | $T_5$ | 164.797 | 164.797 | 164.797 | 164.797 | 164.797 |
| | Cauchy | 164.812 | 164.797 | 164.797 | 164.612 | 164.612 |
| 2 | Normal | 164.546 | 164.546 | 164.546 | 164.546 | 164.546 |
| | Laplace | 164.547 | 164.546 | 164.546 | 164.547 | 164.547 |
| | $T_5$ | 164.546 | 164.546 | 164.546 | 164.546 | 164.546 |
| | Cauchy | 164.567 | 164.547 | 164.547 | 164.564 | 164.564 |
| 3 | Normal | 164.544 | 164.544 | 164.544 | 164.544 | 164.544 |
| | Laplace | 164.545 | 164.545 | 164.545 | 164.545 | 164.545 |
| | $T_5$ | 164.545 | 164.545 | 164.545 | 164.545 | 164.545 |
| | Cauchy | 164.565 | 164.546 | 164.546 | 164.574 | 164.574 |

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

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

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