دراسة تحليلية لدالة تأثير معامل الارتباط القويم بين صادرات النفط وعائداتها لبعض دول منظمة أوبك بإستعمال طريقة المتوسط ثنائي الوزن

الباحث/ زهراء خليل حمودي العميـــــــد [1]
وزارة النفط / بغداد ، شارع فلسطين
أ.د. لقاء علي محمد العلــــــوي [2]
جامعة بغداد / كلية الادارة والاقتصاد ، بغداد ، الاعظمية

المستخلص

إن أساليب التحليل الإحصائي متعددة المتغيرات تعتمد على الوصف الدقيق والتفسير الواضح للظواهر التي تعتمد على مجوعتين من المتغيرات ومنها أسلوب تحليل الارتباط القويم، إذ يعتمد في تحليله على نهجين إما بإستعمال مصفوفة التباين المشترك أو مصفوفة الارتباط لكل مجموعة من المتغيرات ، ولقد تناول العديد من الباحثون تقدير معامل الارتباط القويم بإستعمال عدد من الطرائق التي تكون بعض الاحيان متحيزة تجاه القيم الشاذة وتدعى الكلاسيكية إضافة الى طرائق حصينة مقاومة لوجود تلك القيم. ولتحديد مدى دقة وكفاءة المقدرات الحصينة تم إستعمال مقاييس توضح تأثير القيم المتطرفة على المقدر الحصين ومنها دالة التأثير.

تم في هذا البحث تسليط الضوء على طريقة المتوسط ثنائي الوزن الحصينة لتقدير معامل الإرتباط القويم لمجوعتين من المتغيرات، الاولى تمثل المعدلات الشهرية لكميات النفط الخام والثانية العائدات المالية المستحصلة من تصدير تلك الكميات وللسنوات (2015-2019) فضلا عن تطبيق معيار المقدرات المرجحة والمنقولة لدالة التاثير التجريبية ، وقد تبين أن اعلى تأثير على معامل الارتباط القويم كان عند الشهر الرابع والثلاثون وأقل تأثير عن الشهر السابع والعشرون .

__الكلمات المفتاحية :__ الارتباط القويم ، الطرق الحصينة ، القيم الشاذة ، معامل إرتباط المتوسط ثنائي الوزن ، دالة التأثير.

An analytical study of canonical correlation coefficient influence function between oil exports and Returns for OPEC countries using Biweight Midcorrelation method

Zahraa Khaleel Al-Ameed [1]
Ministry of Oi , Falastin St., Baghdad
Prof. dr. Lekaa Ali Al-Alway[2]
University Of Baghdad, Adhamiya,Baghdad, Iraq

Abstract

The multivariate statistical analysis methods depend on accurate description and clear explanation of the phenomena that depend on two sets of variables, including the method of canonical correlation analysis, as it depends in its analysis on two approaches either using the covariance or the correlation matrix for each group of variables. Many researchers have dealt with the estimation of canonical correlation by using a number of methods that are sometimes biased towards outliers and called classical, in addition robust methods that are resistant to the presence of these values. To determine the accuracy and efficiency of the robust estimators, we used measures that show the effect of outliers on these estimators, including influence function.

In this research, we highlighted Biweight Midcorrelation method of estimating canonical correlation coefficient for two sets of variables, the first representing the monthly average of quantities for crude oil and the second was financial returns obtained from the export of these quantities, for years (2015-2019), as well as the application of the weighted and transferred estimates standard for the empirical influence function. The result was found that the highest effect on the canonical correlation coefficient was at thirty-fourth month, and the lowest effect on the twenty-seventh month.

Key Words: Canonical correlation, robust method, Outliers, Biweight Midcorrelation coefficient, Influence function.

1-المقدمـــة

إن تحليل الارتباط القويم هو أسلوب من أساليب تحليل متعدد المتغيرات فهو يحتوي في تركيبته على متغير إستجابة أحادي يرتبط بمتغيرين تفسيريين أو اكثر ليتضح لنا مدى علاقته بها.

أما في حالة الارتباط القويم ، فيكون لدينا عدد q من متغيرات الاستجابة (Y's) والمطلوب معرفة شكل إرتباطاتها مع مجموعة المتغيرات التفسيرية (X's) أي أننا نعني بالارتباط القويم بأنه الارتباط بين مجموعتين من المتغيرات عن طريق إيجاد أكبر ترابط خطي للمتغيرات في إحدى المجموعتين مع التراكيب الخطية للمتغيرات في المجموعة الثانية و تكون لكلا المجموعتين توزيع مشترك .

إن البيانات التي يتم جمعها من المصادر الخارجية قد تحتوي على مشاهدة شاذة أو أكثر ، وهي قيمة أو مجموعة قيم تسلك سلوك مغاير عن باقي المشاهدات وتكون غير متسقة معها، وتتأثر فيها التقديرات الكلاسيكية كالمتوسط والتباين إضافة الى الارتباطات تأثراً سلبياً وبالتالي تفشل في توفير مقاييس ملائمة وجيدة للبيانات.

ولمعالجة وجود هذه القيم إستعمل الباحثين طرق حصينة مقاومة لتأثير الشواذ على المقدر الحصين وتكون ذي فاعلية أكبر من الطرق الكلاسيكية، ولمعرفة مدى حصانة هذه الطرق يتم إختبارها عن طريق مقاييس تكمن أهميتها في الكشف والفحص الموثوق به عن القيم المتطرفة ومن هذه المقاييس دالة التأثير والتي سوف يتم إستعمالها في بحثنا هذا لقياس تأثير المشاهدات الشاذة على معامل الارتباط القويم الحصين، إذ أن في عام (1992)م، قدم العالم (Mario Romanazzi) إشتقاق دالة التأثير لمربع معامل الارتباط القويم والمتعدد إضافة الى شرح ووصف مفصل لثلاثة أنواع من تحويلات العينة لدالة التأثير إضافة الى إيجاد دالة التأثير للمتجهات القويمة والمميزة والقيم المميزة بالاعتماد على دراسة العالم (Hample) في مقتبل السبعينات [5] وفي عام (2000)م ، قدمت الباحثة (أنعام عبد الوهاب عبد الجبار) دراسة تفصيلية عن دالة التأثير في تحليل الارتباط القويم ودراسة كل خصائصها وجميع المقاييس الحصينة المشتقة منها إضافة الى علاقتها بمتغيرات (Jackknife) ومن ثم إستعمال مقدرات M لتقدير المعالم، وقد جاءت بنتائج توضح مدى تحسس دوال التأثير للمتجهات القويمة المرتبطة بالارتباط القويم

للشواذ إضافة الى أن قيمها تزود الباحث بمعلومات مهمة عن المشاهدات التي تسبب تلوث البيانات قيد الدراسة.[3]

أما في عام (2012)م، قامت الباحثة (ذكاء يوسف عزيز) بدراسة العلاقة بين عمليات إدارة المعرفة والفساد الاداري على عينة من أساتذة كلية الادارة والاقتصاد عن طريق المقارنة بين الارتباط القويم الكلاسيكي والارتباط القويم الحصين لمعرفة مدى كفاءة التقدير بالاعتماد على مقياس متوسط مربع الخطأ (MSE) ، و أشارت النتائج الى ان الارتباط القويم الحصين بعد استعمال طريقة التقدير الحصينة (Minimum Coveriance Determinate)(MCD) أفضل من الارتباط التقويمي الاعتيادي [4]

وفي عام (2013) قدم كل من (Chun-Hou Zheng , Lin Yuan, Wen Sha , Zhan-Li Sun) دراسة عن التحليل التفاضلي الجيني(Coexpression) وذلك بإستعمال معامل إرتباط متوسط ثنائي الوزن (Biweight Midcorrelation) كونه معامل إرتباط حصين مقاوم أكثر للقيم المتطرفة من معامل إرتباط (Pearson) بمقاومته للشواذ ، وبعد تطبيق المحاكاة تبين أن إستعمال معامل ارتباط متوسط ثنائي الوزن أفضل من الطرق السابقة المستعملة في هذا التحليل .[6]

في عام (2016) ناقش كل من(Patrick Veenstra , Coline Cooper & Steve Phelps) تحليل العلاقة بين عائدات الأوراق المالية المختلفة بإستعمال معامل إرتباط متوسط ثنائي الوزن (Bicor) (Biweight Midcorrelation) بدلا من معامل إرتباط(Pearson) بإعتباره من الاجراءات الاكثر قوة لمعرفة العلاقة بين العوائد ، ووضحت النتائج أنه يمكن إستعمال طريقة (Bicor) لتحسين طريقة بناء المحفظة المالية عند التعامل مع مصفوفة الارتباط وبذلك الحصول على أداء أفضل [7]

2-الارتباط القويم (Canonical Correlation)

يدرس الارتباط القويم (Canonical Correlation) العلاقة بين مجموعتين من المتغيرات ، الاولى $(X_1, X_2, \ldots\ldots, X_p)$ والمجموعة الثانية $(Y_1, Y_2, \ldots\ldots, Y_q)$.

من خلال التراكيب الخطية ومن بين متطلبات الارتباط القويم ان كل مجموعة من المتغيرات تقلص الى متغير مفرد موصوف بتركيبة خطية من المعالم ثم ايجاد علاقة

الارتباط بين المجوعتين، و ان المتغيرات التي يتم الحصول عليها من خلال هذه التركيبة الخطية تعرف بالمتغيرات القويمة (Canonical Variables) والارتباط بينهما يعرف بالارتباط القويم (Canonical Correlation) [1].

3-التحليل الرياضي للارتباط القويم (Mathmatical Analysis for Canonical Correlation)

يهدف الارتباط القويم الى دراسة العلاقة بين مجموعة من المتغيرات التفسيرية X ومجموعة من متغيرات الاستجابة Y .

وبافتراض دراسة مجموعتان من المتغيرات :

X_{p*1} عبارة عن متجه بأبعاد $p*1$ لمتغيرات المجموعة الاولى $X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{bmatrix}$

Y_{q*1} عبارة عن متجه بأبعاد $q*1$ لمتغيرات المجموعة الثانية $Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_q \end{bmatrix}$

إذ ان p هو عدد المتغيرات في المجموعة الاولى (X) وq تمثل عدد المتغيرات في المجموعة الثانية (Y) . [2, pp. 378]

ومتغيرات كلا المجموعتين يتبعان التوزيع الطبيعي متعدد المتغيرات و أن كل مجموعة تمتلك المواصفات التالية :

$E(x) = \mu_x \qquad E(y) = \mu_y$

$Var(x) = \Sigma_{xx} \qquad Var(y) = \Sigma_{yy}$

$$\binom{X}{Y} \sim MVN\left[\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}\right]$$

ومصفوفة تغاير مابين المجموعتين من المتغيرات معرفة بالآتي :

$Cov(x,y) = \Sigma_{xy}$

وان $\Sigma_{xx} > 0$ ، $\Sigma_{yy} > 0$ وبإفتراض أن $p \leq q$ فإنه بالامكان تعريف عدد من التراكيب الخطية مساوية في العدد الى $Min_{(p,q)}$ وفق العلاقة :

$$u = \acute{a}x_1 \qquad \dots\dots \qquad (1)$$

إذ أن U ؛ متجه من درجة $n \times 1$

$$v = \acute{b}y_1 \qquad \dots\dots(2)$$

V ؛ متجه من درجة $n \times 1$

ان كل تركيبة خطية تختلف في قيم الاوزان المرافقة الى كل متغير لاختلاف أهمية المتغير ضمن المجموعة وتأثيره على المتغير القويم U_i أو V_i.

ولحساب معامل الارتباط القويم مابين المتغيرين أي ؛ Corr $\left(\frac{\underline{x}}{\underline{y}}\right)$

وبالاستناد الى أساس التباين لكل مجموعة من المتغيرات القويمة

$$\text{Var }(u) = \underline{\acute{a}}\,\Sigma_{xx}\,\underline{a} = 1 \dots\dots(3)$$

$$\text{Var }(v) = \underline{\acute{b}}\,\Sigma_{yy}\,\underline{b} = 1 \dots\dots\dots(4)$$

$$\underline{\acute{a}}\,\Sigma_{xx}\,\underline{a} = \underline{\acute{b}}\,\Sigma_{yy}\,\underline{b} = 1 \dots\dots\dots(5)$$

إذ أن ؛ $\underline{a}$ ، $\underline{b}$ تمثلان الاوزان القويمة المناظرة لأكبر معامل إرتباط قويم .

والتغاير بين التراكيب الخطية

$$\text{Cov }(\underline{\acute{a}}\,\underline{x}\,,\underline{\acute{b}}\,\underline{y}) = \underline{\acute{a}}\,\Sigma_{xy}\,\underline{b} \dots\dots\dots\dots(6)$$

عند ذلك فإن الارتباط القويم يوصف ؛

$$\text{Corr (u,v)} = \frac{cov(u,v)}{\sqrt{var(u)}\sqrt{var\ (v)}} \qquad \dots(7)$$

ومنه

$$\text{Corr }(\underline{\acute{a}}\,\underline{x}\,,\underline{\acute{b}}\,\underline{y}) = \frac{\underline{\acute{a}}\,\Sigma_{xy}\,\underline{b}}{\sqrt{\underline{\acute{a}}\,\Sigma_{xx}\,\underline{a}}\ \sqrt{\underline{\acute{b}}\,\Sigma_{yy}\,\underline{b}}} \qquad \dots(8)$$

فالزوج الاول من المتغيرات القويمة (v_1, u_1) يتم اختيارها لأجل تعظيم التغاير بينهما ، فالتراكيب الخطية للزوج ؛

$$u_1 = \underline{a_1}\,\underline{x} \quad , \quad v_1 = \underline{b_1 y}$$

وبما أن تباين المتغيرات القويمة للزوج الاول مساوي للواحد فان الارتباط القويم بينهما ؛

$$\rho_{(u1,v1)} = max_{corr\underline{a},\underline{b}}(\acute{a}\,\underline{x}\,,\acute{b}\,\underline{y}) \dots \dots \dots .(9)$$

والارتباط الناتج يمثل معامل الارتباط القويم للزوج الاول .

اما الزوج الثاني من المتغيرات القويمة (u1,v1) يتم اختيارهما لأجل تعظيم التغاير cov(u,v) شرط أن التراكيب الخطية للزوج تكون متعامدة على الزوج الاول (u1,v1) بمعنى أن

$$\text{Cov}\ (\acute{a}\,\underline{x}\,,u1) = 0 \dots \dots \dots ..(10)$$

$$\text{Cov}\ (\,\acute{b}\,\underline{y}\,,v1) = 0 \dots \dots \dots ...(11)$$

$$\text{Var}\ (\,\acute{b}\,\underline{y}) = \text{Var}\ (\acute{a}\,\underline{x})=1..\ (12)$$

ان تعظيم الارتباط مابين $\underline{\acute{a}_2 x}$ و $\underline{\acute{b}_2 y}$ يدعى معامل الارتباط القويم الثاني وبصورة عامة ان الزوج (Uj,Vj) من المتغيرات القويمة يتم اختياره لأجل تعظيم التغاير Cov والمشروطة بأن التراكيب الخطية للزوج (j)تكون متعامدة على الزوج السابق (u_1, v_1) $(j-1)$ من التراكيب.

وبذلك فإن معاملات الارتباط القويم بدلالة المتغيرات القويمة والتباينات تقدر بالعلاقة:

$$r_c = \frac{\acute{U}S_{xy}V}{\sqrt{\acute{U}S_{xx}U}\sqrt{\acute{V}S_{yy}V}} \quad ...(13)$$

S_{xy} : مصفوفة التباين للمستويات في المجموعتين X,Y

S_{xx} : مصفوفة التباين للمستويات في المجموعة X

S_{yy} : مصفوفة التباين للمستويات في المجموعة Y

ويمكن حساب معامل الارتباط القويم عن طريق مصفوفة الارتباط :

إذا كان لدينا p من المستويات في المصفوفة X و q من المستويات في المصفوفة Y ، نفرض ان مصفوفة التحويل :

S= DRD

$$D_X = \begin{bmatrix} \sqrt{S_{X_1X_1}} & \dots . & 0 \\ 0 & \dots . & \sqrt{S_{X_pX_p}} \end{bmatrix} \qquad D_Y = \begin{bmatrix} \sqrt{S_{Y_1Y_1}} & \dots . & 0 \\ 0 & \dots . & \sqrt{S_{Y_qY_q}} \end{bmatrix}$$

إذ أن :

R مصفوفة الارتباط للمجموعة X أو Y أو التغاير بينهما وتساوي :

$$R_{XX}=\begin{bmatrix}1 & r_{X_1X_2}\ \ldots\ldots & r_{X_1X_p}\\ r_{X_2X_1} & 1 & :\\ r_{X_pX_1}1 & \ldots\ldots & 1\end{bmatrix}$$

$$R_{YY}=\begin{bmatrix}1 & r_{Y_1Y_2}\ \ldots\ldots & r_{Y_1Y_q}\\ r_{Y_2Y_1} & 1 & :\\ r_{Y_qy_1}1 & \ldots\ldots & 1\end{bmatrix}$$

$$R_{XY}=\begin{bmatrix}r_{X_1Y_1} & r_{X_1Y_2}\ \ldots\ldots & r_{X_1Y_q}\\ : & r_{X_2Y_2} & :\\ r_{X_pY_1} & \ldots\ldots & r_{X_pY_q}\end{bmatrix}$$

D : مصفوفة قطرية عناصرها تمثل جذر تباين كل متغير وتعرف:

D=diag$\left(\sqrt{S_{ij}}\right)$

وبذلك فإن صيغة معامل الارتباط القويم من العلاقة بمصفوفات الارتباط يوصف :

$$r_c=\frac{\acute{C}R_{xy}D}{\sqrt{\acute{C}R_{xx}C}\sqrt{\acute{D}R_{yy}D}}\quad \ldots\ldots(14)$$

إذ أن :

D&C : تمثل الاوزان القويمة لمعامل الارتباط القويم

ولأجل تقدير الاوزان القويمة التي تعظم الارتباط القويم ، يتم وصف الدالة :

$$g=\acute{C}R_{xy}D-\frac{\sqrt{\lambda_1}}{2}\acute{C}R_{xx}C-\frac{\sqrt{\lambda_2}}{2}\acute{D}R_{xx}D\ \ldots\ldots(15)$$

والسعي الى تعظيمها $max_{c,d}(g)$ من خلال :

$$\frac{\partial g}{\partial d}=0\ ,\frac{\partial g}{\partial c}=0$$

$$\frac{\partial g}{\partial c}=R_{xy}\underline{d}-\sqrt{\lambda_1}R_{xx}\acute{c}\ldots\ldots\ldots\ldots\ldots\ldots\ (16)$$

$$\frac{\partial g}{\partial d}=\underline{c}R_{xy}-\sqrt{\lambda_2}\underline{\acute{d}}R_{yy}\ldots\ldots\ldots\ldots\ldots\ldots\ (17)$$

بضرب المعادلة (17) في (c) والمعادلة (2-20) في (d) على التوالي نحصل على :

$$c\ R_{xy}\ d-\sqrt{\lambda_1}\ \acute{c}\ \ R_{xx}\ c=0$$

$$c\ R_{xy}\ d-\sqrt{\lambda_2}\ \acute{d}\ \ R_{yy}\ d=0$$

وبحل المعادلتين اعلاه والاستفادة من الشرط $\acute{c}\, R_{xy}\, c = \acute{d}\ R_{yy}\, d = 1$ فإن :

$$c\, R_{xy}\, d = \sqrt{\lambda_1} = \sqrt{\lambda_2}$$

وبتعويض $\sqrt{\lambda_2}$, $\sqrt{\lambda_1}$ أحدهما بدل الآخر في المعادلات نحصل :

$$R_{xy}\, c = \sqrt{\lambda}\ \ R_{yy}\, d$$

$$\underline{C} = \frac{1}{\sqrt{\lambda_1}}\ R_{xx}^{-1}\ R_{xy}\ \underline{d} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (18)$$

وبتعويض C في $cR_{xy} = \sqrt{\lambda}\ \ R_{yy}\, d$ نحصل على العلاقة :

$$(R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy} - \lambda I)\ \underline{d} = \underline{0} \qquad \ldots\ldots\ldots\ldots (19)$$

وهي تمثل المعادلة المميزة (eigen equation) للمصفوفة $R_{yy}^{-1}R_{yx}R_{xx}^{-1}R_{xy}$ وان الجذور λ_i التي لاتساوي صفر المتحققة من حل هذه المعادلة تساوي q وتدعى بالجذور الذاتية وان مربع معامل الارتباط القويم بين كل زوج من المتغيرات القويمة يساوي قيمة الجذر المميز وفق الصيغة التالية : [2]

$$r_c^2 = \lambda$$

4- المشاهدات الشاذة في بيانات متعدد المتغيرات

(Outlier Observations in Multivariate Data)

تحتوي العينات الاحصائية في متعدد المتغيرات على مشاهدة شاذة والتي تنتج بسبب الاخطاء الحاصلة عند تسجيل البيانات ويعتبر أكثر تعقيد مما عليه في حالة المتغير الواحد.

ولأجل إختبار وجود تلك المشاهدات نلجأ الى إستعمال مسافات مهلنوبس التربيعية الكلاسيكية :

$$\mathrm{d}_i^2 = (\underline{X}i - \overline{X})' S^{-1} (\underline{X}i - \overline{X}) \quad \ldots\ldots\ldots\ldots\ldots (20)$$

ثم إستبعادها وتطبيق التقديرات على المشاهدات الجيدة، وكذلك يمكن إستعمال مسافات مهلنوبس التربيعية الحصينة :

$$RD_i^2 = (\underline{X}i - \underline{\mu}_n)' \textstyle\sum_n^{-1} (\underline{X}i - \underline{\mu}_n) \quad \ldots\ldots (21)$$

و يتم إحتسابها لكل مشاهدة ثم تقارن مع قيمة $\chi^2_{(P+q,0.05)}$

الجدولية فأذا كانت المسافة المحسوبة للمشاهدة أكبر من القيمة الجدولية فتعتبر تلك المشاهدة شاذة.[8]

5-معامل ارتباط متوسط ثنائي الوزن (The Biweight Midcorrelation Correlation Coefficient)

إن من عيوب معامل إرتباط بيرسون هو سهولة تعرضه لتأثيرات القيم الشاذة إرتباط المتوسط ثنائي الوزن .

نفرض أن ψ أي دالة فردية وان μ_x و μ_y أي مقياس للموضع للمتغير العشوائي X و Y على التوالي ، ونفرض أن ℓ_x و ℓ_y أي مقياس للقياس للمتغير العشوائي X و Y على التوالي فإذا كان K عبارة عن مقدار ثابت وعرف المتغيرات بدلالة المعالم السابقة بالصيغة :

$$U = \frac{(X-\mu_x)}{K\ell_x} \quad , \quad V = \frac{(Y-\mu_y)}{K\ell_y}$$

عند ذلك فان مقياس التغاير مابين المتغيرين X و Y يوصف بالعلاقة :

$$\gamma_{xy} = \frac{nk^2.\tau_x.\tau_y\, E(\psi(u).\psi(v))}{E(\psi(u).\psi(v))} \qquad \ldots\ldots\ldots\ldots\ldots (22)$$

إذ أن مقياس الارتباط ρ_b يحسب بالعلاقة التالية :

$$\rho_b = \frac{\gamma_{xy}}{\sqrt{\gamma_{xx}.\gamma_{yy}}} \qquad -1 \le \rho_b \le 1 \qquad \ldots\ldots\ldots\ldots (23)$$

وبإختيار K=9 والدالة ψ والتي تمثل دالة الوزن الثنائي (Biweight Function) والتي تعرف بالعلاقة التالية؛[9, pp. 697]

$$\psi(x) = \begin{cases} x(1-x^2)^2 & if\ |x| < 1 \\ 0 & if\ |x| \ge 1 \end{cases} \qquad \ldots\ldots(24)$$

وبإفتراض أن med_x و med_y عبارة عن وسيط المتغير X و Y على التوالي محسوبة من عينة

عشوائية للازواج المرتبة للمشاهدات (X_1,Y_1)و(X_2,Y_2)،،،، (X_n,Y_n) ومنها ينتج تعريف المتغيرات [9]

$$U_i = \frac{(X_i - med_x)}{9.MAD_x} \; , V_i = \frac{(Y_i - med_y)}{9.MAD_y}$$

نلاحظ أن المقدار U_i متناسب مع المسافة بين X_i والوسيط لـ X .[7]

إذ أن (Median Absolute Deviation) $(MAD_y و MAD_x)$ تمثل متوسط الانحرافات المطلقة عن الوسيط للمتغير العشوائي X و Y على التوالي :

$$MAD_x = med_i|x - med_{xi}| = med|x - med_x|$$

واذا ماتم تعريف المتغيرات a_i و b_i عن علاقتهما بالمتغيرات U_i و V_i ,

$$a_i = \begin{cases} 1 & -1 \le U_i \le 1 \\ 0 & O.W \end{cases}$$

$$b_i = \begin{cases} 1 & -1 \le V_i \le 1 \\ 0 & O.W \end{cases}$$

ومنه نحصل على وسط التغاير ثنائي الوزن (Biweight Midcoveriance) مابين المتغير X و Y بالعلاقة :

$$Bicov(x,y) = \frac{n\Sigma a_i(X_i - med_x)(1-U_i^2)^2 b_i(Y_i - med_y)(1-V_i^2)^2}{[\Sigma a_i(1-U_i^2)(1-5U_i^2)][\Sigma b_i(1-V_i^2)(1-5V_i^2)]} \quad ....(25)$$

وبتطبيق صيغة الارتباط فإن تقدير معامل وسط الارتباط ثنائي الوزن (Biweight Midcorrelation) يحسب بالعلاقة التالية :

$$r_{bi} = \frac{bicov(x,y)}{\sqrt{bicov(x,x).bicov(y,y)}} \quad ....(26)$$

ولأجل اختبار معامل وسط الارتباط ثنائي الوزن وسيتم اختبار الفرضية

$$H_0 : \rho_b = 0$$

والتي تدل على أن X و Y متغيران مستقلان ولحساب إحصاءة الاختبار

$$T_b = r_b . \sqrt{\frac{n-2}{1-r_b^2}}$$

ويتم رفض فرضية العدم اذا تحقق

$$|T_b| > t_{1-\frac{\alpha}{2}}$$

وان $t_{1-\frac{\alpha}{2}}$ قيمة جدولية في توزيع t بدرجات حرية V=n−2 وخطأ من النوع الاول مساوي الى α.[10]

6-مقاييس المقدرات الحصينة (Measurement of Robust Estimators)

إن مقاييس المقدرات الحصينة تعتبر أدوات في غاية الاهمية لوصف وفهم سلوك هذه المقدرات ومن هذه المقاييس هي دالة التأثير والتي سيتم إستعمالها في هذا البحث .

إن هذه الدالة تزودنا بمدى تأثر المقدر عندما يتم تغير مشاهدة ، فتكون قيمتها مفيدة عند دراسة حصانة المقدر وكذلك في حالة حساب مصفوفات التباين لأنواع معينة من المقدرات ومصفوفات الارتباط لأنواع أخرى من المقدرات خاصة عندما يكون من الصعوبة إستعمال الطرق التقليدية .[11]

7- دالة التأثير (Influence Function) (IF)

تعتبر دالة التأثير أداة تحليلية في الاساس يمكن إستعمالها لتقييم تأثير مشاهدة على المقدر ويرمز لها (IF) والى المقدر T_n عند دالة التوزيع F بالعلاقة: [9, pp. 55]

$$IF_{T_n,F(x)} = \lim_{\omega \to 0} \frac{[T_n(F_\omega) - T_n(F)]}{\omega} \quad \ldots\ \ldots \ldots \ldots \ldots (27)$$

إذ أن :

F : دالة التوزيع التجميعية .

$T_n(F)$: دالة لدالة التوزيع ويطلق عليها رياضياً (Functional).

$IF_{T_{n,F(x)}}$ ، مشتقة $T_n(F)$ وهي التي تمثل دالة التأثير.

وإن :

$$F_\omega = (1-\omega)F + \omega\delta_x \quad \ldots.(28)$$

إذ أن :

ω : تمثل نسبة التلوث وتتراوح بين $0 < \omega < 1$

δ_x : تمثل مقياس إحتمالي بحيث :

$$\delta_{x0}(x) = \begin{cases} 1 & \text{for } x \geq x_0 \\ 0 & \text{for } x < x_0 \end{cases} = CDF\ for\ a\ point\ mass\ 1\ at\ \ x_0$$

بعد التعرف على دالة التأثير، أصبح من اللازم التطرق الى شي من التفصيل الى مقدرات مشابهة ومقاربة في الخصائص لدالة التأثير وذو علاقة مباشرة أكثر بالعينات المستعملة والتي تحقق الاهداف التالية :

1- الحصول على فكرة مباشرة عن تأثير القيم الشاذة على المقدر.

2- لقياس تأثير المشاهدة على المقدر $T(F_n)$.

3- لتقدير التباين المحاذي للمقدر $T(F_n)$.[11]

وهذه المقدرات يمكن وصفها كالتالي :

8-المقدرات غير المقيَسة وغير المنقولة

(Unscaled and Untranslated Estimators)

افرض أن $x_1, x_2, \dots . x_{n-1}$ عينة ذات حجم n −1 فإنه بالامكان أن نعرف المقدر :

$$I^1_{(x)} = T_n(x_1, x_2, \dots . x_{n-1}, x) \qquad \dots\dots\dots(29)$$

إذ أن $I^1_{(x)}$ تمثل قيمة المقدر عندما يتم إضافة مشاهدة جديدة ولتكن x الى العينة وتدعى هذه الحالة (addition – corruption).

أما بإفتراض أن $x_1, x_2, \dots . x_n$ عينة ذات حجم n فإنه بالامكان تعريف المقدر التالي :

$$I^2_{(x)} = T_n(x_1, x_2, \dots . x_n, x) \qquad \dots\dots\dots\dots (30)$$

و أن $I^2_{(x)}$ تمثل قيمة المقدر بعد إحلال مشاهدة ولتكن x_n بالمشاهدة x وتدعى هذه الحالة (replacement – corruption).[12]

9-المقدرات غير المقيَسة والمنقولة (Unscaled and Translated Estimators)

ان المقدرات التالية تعتبر مقدرات منقولة (translated) من المقدر $I^1_{(x)}$ و $I^2_{(x)}$ [12]

$$I^3_{(x)} = T\ (x, x_1, x_2, \dots . x_n) - T\ (x_1, x_2, \dots . x_n) \dots\dots\dots\dots (31)$$

والمقدر

$$I^4_{(x)} = \sum_{i=1}^{n} [T(x \text{ replacing } x_j) - T(x_1, x_2, \dots . x_n) \dots \dots . (32)$$

10- المقدرات المقيَسة والمنقولة (Scaled and Translated Estimators)

ان دالة التأثير التجريبة (Empirical Influence Function) تعرف بالاعتماد على المقدرات السابقة بالصيغة

$$EIF(x, F_n) = IF(x, F_n)$$
$$= U_{\omega\to 0} \frac{T(F_n + \omega(\delta_x - F_n)) - T(F_n)}{\omega} \dots\dots\dots \quad (33)$$

إذ أن :

F_n ، دالة التوزيع

$(\delta_x - F_n)$: الفرق بين توزيع المشاهدة الملوثة وتوزيع المشاهدة الاصلية

لذلك فإن المقدار $T\left(F_n + \omega(\delta_x - F_n)\right)$ يتم الحصول من خلاله على مقدر (T) ذي توزيعين أغلب مشاهداته تتبع التوزيع الطبيعي (التوزيع الاصلي) ولكن يحتوي على مشاهدات قليلة تتبع التوزيع الملوث(الناتج من إضافة أو إحلال مشاهدة ملوثة).

أما المقدار $T(F_n)$ يمثل المقدر الاصلي الناتج من دالة التوزيع الاصلية F_n بحجم عينة (n).

ويكون من الافضل تقدير دالة التأثير التجريبية (دالة التأثير) بالعلاقة :

$$EIF_e(x, F_n) = IF_e(x, F_n)$$

$$= \frac{T\left(F_n + \frac{1}{100\,n}(\delta_x - F_n)\right) - T(F_n)}{\frac{1}{100\,n}} \qquad \ldots\ldots\ldots\ldots(2\text{-}70)$$

إذ أن :

$\frac{1}{100\,n}$: تمثل النسبة التي يم أخذها لتلويث البيانات .

من ذلك فإن قيم دالة التأثير التجريبية يمكن أن تعرف

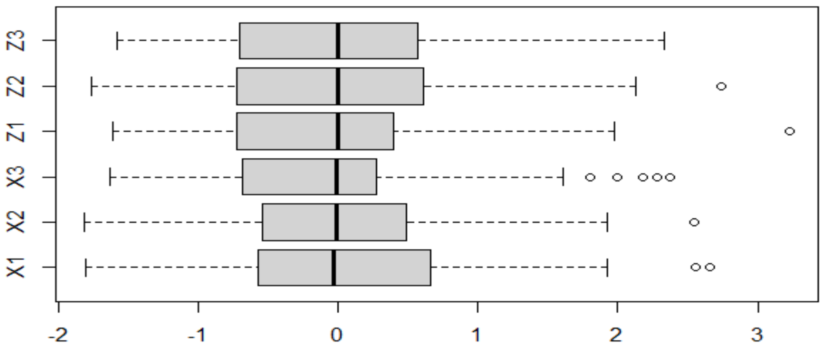
$$I_j = IF_{(xj)} = EIF\left(x_j, F_n\right) \quad \ldots\ldots\ldots\ldots\ldots(34)$$

والتي يمكن أن تقرب بإختيار قيم مختلفة الى ω (نسبة التلويث) مثل ($\frac{1}{n}$ ، $\frac{1}{\sqrt{n}}$ ، $\frac{1}{n+1}$ ، $\frac{1}{n-1}$) وغيرها من القيم دون أخذ النهاية (Limit) للمقدار .[12]

11- الجانب التطبيقي

تتألف بيانات الدراسة من مجموعتين المجموعة الأولى تتضمن كميات النفط الشهرية المصدرة (مليون برميل) لثلاثة دول منتجة للنفط ضمن منظمة أوبك وهي السعودية (x_1)، العراق (x_2) والكويت (x_3) مسجلة لفترة زمنية بطول ستون شهر ضمن السنوات بدءا من كانون الثاني (2015) ولغاية (2019) كانون الثاني ، اما مجموعة المتغيرات الثانية هي $\left(z_1, z_2, z_3\right)$ فتمثل الواردات المتحققة من تلك الكميات (مليون دولار) لنفس الدول وعلى التوالي.

وقد تم تطبيق طريقة الرسم الصندوقي (Boxplot) للكشف الاولي عن وجود القيم الشاذة كما مبين في الشكل (1) ادناه، وقد وضح وجود قيم شاذة في بيانات كميات النفط وعائداتها.



الشكل (1)

الرسم الصندوقي لبيانات كميات وعائدات النفط

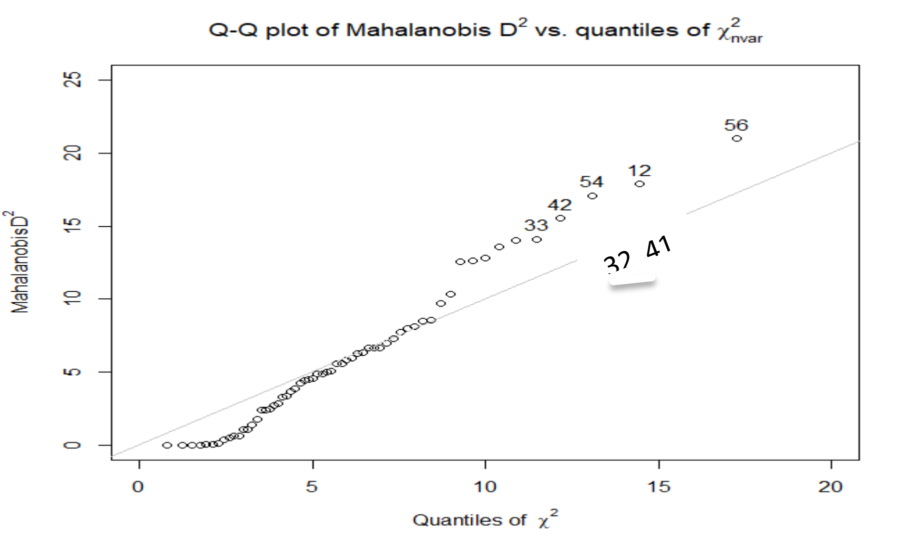
الجدول (1)

قيم (Mahalanobis Distance) لتحديد القيم الشاذة

obs	MD	Obs	MD	obs	MD	obs	MD
1	6.641	16	2.422	31	2.390	46	9.722
2	0.600	17	0.001	32	14.102	47	9.722
3	5.059	18	3.688	33	6.944	48	12.834
4	7.316	19	6.279	34	1.385	49	8.131
5	3.322	20	0.594	35	2.454	50	6.358
6	0.386	21	4.558	36	4.458	51	0.084
7	4.870	22	5.801	37	5.579	52	7.725
8	1.094	23	4.893	38	0.001	53	8.503
9	13.545	24	3.866	39	2.842	54	17.068
10	6.636	25	4.992	40	0.033	55	5.561
11	17.877	26	4.509	41	15.510	56	20.972
12	14.009	27	2.699	42	0.001	57	1.741
13	0.136	28	7.971	43	12.575	58	0.001
14	10.347	29	0.477	44	5.971	59	1.081
15	4.273	30	3.304	45	6.628	60	12.642

وقد تم إستعمال طريقة (Mahalanobis Distance)(MD) وذلك لتحديد الشواذ بشكل دقيق والنتائج في الجدول (1) أعلاه والشكل (2) ادناه تشير الى ان

المشاهدات الشاذة كانت (9، 11، 12، 32، 41، 48، 54، 56، 60) وذلك لان قيم (MD) المحسوبة لتلك المشاهدات كانت اكبر من قيمة $\chi^2_{(6,0.05)}$ الجدولية والبالغة (12.59).



الشكل (2)

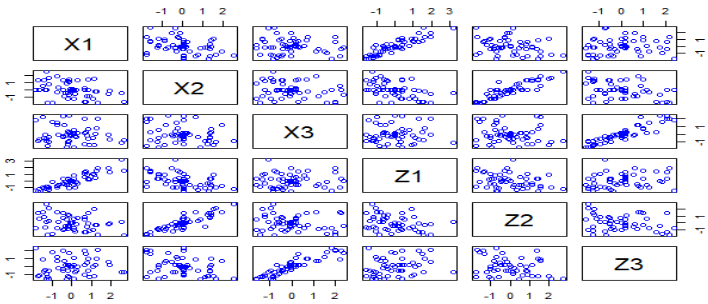
تحديد المشاهدات الشاذة بواسطة (Mahalanobis Distance)

كما تم حساب مصفوفة الارتباطات البسيطة بين المجموعتين ادناه وذلك باستعمال معامل ارتباط بيرسون، فضلا رسم طبيعة الارتباط بين كافة المتغيرات المدروسة كما هو موضح في الشكل (3-30).

$$\rho = \begin{bmatrix} 1 & -0.35 & -0.12 & 0.88 & -0.28 & 0.08 \\ -0.35 & 1 & -0.21 & -0.48 & 0.89 & -0.36 \\ -0.12 & -0.21 & 1 & 0.02 & -0.13 & 0.92 \\ 0.88 & -0.48 & 0.02 & 1 & -0.29 & 0.24 \\ -0.28 & 0.89 & -0.13 & -0.29 & 1 & -0.27 \\ 0.08 & -0.36 & 0.92 & 0.24 & -0.27 & 1 \end{bmatrix}$$

المصفوفة أعلاه واشكال الانتشار لمعاملات الارتباط الموضحة في الشكل (3) أدناه بينت ان معاملات الارتباط بين متغيرات كميات النفط للدول الثلاثة السعودية، العراق والكويت ومايقابلها من عائدات قد بلغت وعلى التوالي (0.88، 0.89، 0.92)، وان قيم

معاملات الارتباط بين المتغيرات غير المتناظرة تشير الى ان الارتباط بين هذه المتغيرات هو ارتباط ضعيف أي دون المتوسط.



الشكل (3)

طبيعة الارتباطات بين كميات وعائدات النفط

12- تقدير معامل الارتباط القويم والاوزان القويمة:

تم تقدير معاملات الارتباط القويم باستعمال طريقة B الحصينة، إضافة إلى تقدير الاوزان القويمة المناظرة له ، والجدول (4) ادناه يتضمن الجذور المميزة لمعامل الارتباط القويم والاوزان القويمة المناظرة له لبيانات كميات النفط المصدرة وايراداتها الملوثة وغير الملوثة.

جدول (2)

الجذور المميزة والاوزان القويمة لمعامل الارتباط B للبيانات الملوثة وغير الملوثة

	البيانات الملوثة			البيانات غير الملوثة		
Eigenvalues	0.9028	0.8185	0.7602	0.9517	0.8302	0.5909
$\hat{a}$	-0.0988	0.7082	-0.6710	0.5146	-0.5012	0.6802
$\hat{b}$	-0.9926	0.1473	0.1848	-0.9482	0.0831	-0.5083

إن الجدول (2) أعلاه يوضح ان قيم معامل الارتباط القويم المقدرة بطريقة B قد بلغت (0.9501) في حالة البيانات الملوثة، و(0.9755) في حالة البيانات غير الملوثة، وان هنالك تباين بين قيم متجهات الاوزان القويمة $\hat{a}$ و $\hat{b}$ في حالة البيانات الملوثة عن قيم في البيانات غير الملوثة من حيث القيم والاشارة.

13- تقدير دالة التأثير:

من خلال ايجاد قيم دالة التأثير التجريبية وفق أسلوب المقدرات المقيَسة والمنقولة لتوضيح تأثير مشاهدات البيانات المدروسة على معامل الارتباط القويم بين كميات النفط (المجموعة الأولى) والعائدات المناظرة لها (المجموعة الثانية).

وقد تم استخراج قيم الدالة لمعامل الارتباط القويم المقدر بطريقة (معامل إرتباط متوسط ثنائي الوزن) الحصينة، والنتائج في الجدولين (3) و (4) ادناه تبين قيم دالة التأثيرالتجريبية في تقدير معامل الارتباط القويم بطريقة B الحصينة في الحالتين أي وجود الشواذ وفي حالة إستبدالها ، ويتضح من خلالهما ان أعلى قيمة لدالة التأثير بلغت (0.7188) وهي القيمة التي تمتلكها المشاهدة (56)، في حين ان اقل قيمة لدالة التأثير هي القيمة التي تمتلكها المشاهدة (39) وبلغت (0.0766)، وان اعلى قيمة لمقدر دالة التأثير لمعامل الارتباط القويم بطريقة معامل إرتباط متوسط ثنائي الوزن الحصينة بعد استبدال المشاهدات الملوثة بلغت (0.4027) عند استبدال المشاهدة (34)، اي ان المشاهدة (34) هي صاحبة التأثير الاعلى في القيمة المقدرة لمعامل الارتباط القويم، في حين جاءت أقل قيمة لدالة التأثير (0.0039) في حالة استبدال المشاهدة (27)، وهذا يعني ان تأثير المشاهدة (27) ضعيف جدا على القيم المقدرة لمعامل الارتباط القويم، كذلك فان قيم دالة التأثير التجريبية المقدرة عند البيانات الملوثة اكبر من قيمها في حالة استبعاد المشاهدات الملوثة واستبدالها بغير ملوثة.

جدول (3)

دالة التأثير لمعامل الارتباط القويم B المقدر بعد إستبدال المشاهدات الملوثة

Obs.	EIFST	Obs.	EIFST	Obs.	EIFST	Obs.	EIFST
1	0.133	16	0.0148	31	0.0196	46	0.2043
2	0.0797	17	0.0072	32	0.0164	47	0.049
3	0.0254	18	0.0124	33	0.0011	48	0.0665
4	0.0126	19	0.0618	34	0.4027	49	0.0105
5	0.0623	20	0.0798	35	0.0032	50	0.0964
6	0.0168	21	0.0048	36	0.3808	51	0.0055
7	0.0768	22	0.3808	37	0.0053	52	0.0092
8	0.0191	23	0.0004	38	0.1862	53	0.0623
9	0.2166	24	0.1043	39	0.0309	54	0.0012

10	0.0151	25	0.0042	40	0.0352	55	0.0102
11	0.0623	26	0.0389	41	0.0201	56	0.0115
12	0.0301	27	0.0039	42	0.1655	57	0.0213
13	0.0124	28	0.0221	43	0.029	58	0.017
14	0.0123	29	0.0044	44	0.0993	59	0.0077
15	0.0623	30	0.3808	45	0.0623	60	0.3808

جدول (4)

دالة التأثير لمعامل الارتباط القويم B المقدر للبايانات الملوثة

Obs.	EIFST	Obs.	EIFST	Obs.	EIFST	Obs.	EIFST
1	0.1807	16	0.0826	31	0.1723	46	0.3616
2	0.0956	17	0.0777	32	0.0985	47	0.1874
3	0.118	18	0.0766	33	0.1036	48	0.2136
4	0.3042	19	0.0784	34	0.1493	49	0.2825
5	0.0875	20	0.1394	35	0.0776	50	0.1443
6	0.0853	21	0.0783	36	0.0884	51	0.0796
7	0.0904	22	0.079	37	0.0937	52	0.3893
8	0.1026	23	0.1044	38	0.1837	53	0.1723
9	0.0774	24	0.2706	39	0.0766	54	0.094
10	0.1211	25	0.077	40	0.1376	55	0.135
11	0.0924	26	0.0862	14	0.0769	56	0.7188
12	0.3856	27	0.113	42	0.1902	57	0.1113
13	0.2218	28	0.2238	43	0.0766	58	0.0766
14	0.0812	29	0.1178	44	0.1896	59	0.0888
15	0.1007	30	0.0925	45	0.1542	60	0.1019

14-الاستنتاجات:

1. دالة التأثير التجريبية (EIFST) مقياس مهم لتحديد أهمية كل مشاهدة من مشاهدات البيانات قيد الدراسة في عملية التقدير، إضافة إلى دورها في تحديد أثر الشواذ على المقدَر، إذ تمكنا من خلال قيم دالة التأثير لمعامل الارتباط القويم تحديد المشاهدات ذي التأثير الكبير في تقدير معامل الارتباط القويم في حالتي البيانات الملوثة وغير الملوثة.
2. طريقة التقدير الحصينة معامل ارتباط متوسط ثنائي الوزن (B) كفوءة في تقدير معامل الارتباط القويم في حالة البيانات الملوثة.
3. من خلال قيم دالة التأثير التجريبية (EIFST) لمعامل الارتباط القويم المقدر بطريقة معامل إرتباط متوسط ثنائي الوزن (B) لكميات النفط وعائداته للدول الثلاثة في الحالتين قبل وبعد استبدال البيانات الملوثة، نجد أن قيم الدالة التجريبية انخفضت بعد استبدال المشاهدات الملوثة بالمتوسط لكل متغير من المتغيرات.
4. إن اعلى قيمة لدالة التأثير التجريبية (EIFST) لمعامل الارتباط القويم المقدرة بطريقة معامل ارتباط متوسط ثنائي الوزن (B) بين الكميات والعائدات في حالة قبل استبدال القيم الملوثة كان عند المشاهدة (56)، في حين جاءت أقل قيمة عند المشاهدة (39)، اما الاعلى والاقل لقيمة الدالة بعد استبدال المشاهدات الملوثة ، قد ظهرت عند القيم (34) و (27) على التوالي، وهذا يوضح ان اعلى تأثير على طبيعة العلاقة بين الكميات والعائدات هو في شهر اب من عام (2019) م وان أقل تأثير كان في شهر اذار من عام (2018) م في حين اعلى واقل تأثيرقد جاء على التوالي في شهري تشرين الأول واذارمن عام (2017) م، وهذا سيؤثر بشكل سلبي على في حال تفسير العلاقة بين الكميات وعائداتها.

المصادر

1- الراوي ، زياد رشاد ، " طرق التحليل الاحصائي متعدد المتغيرات "(2017) ، المملكة الاردنية الهاشمية ، المعهد العربي للتدريب والبحوث الاحصائية ،ص(7-8 و 126)

2- العلي ، إبراهيم محمد ، " أسس التحليل الاحصائي متعدد المتغيرات " (2020)، سورية ، كلية الاقتصاد – جامعة تشرين ، ص 378 .

3- عبد الجبار ، أنعام عبد الوهاب ،(2000)، " دراسة دالة التأثير في تحليل الأرتباط القويم" ، أطروحة دكتوراه ، كلية الأدارة والأقتصاد ، جامعة بغداد.

4-عزيز ، ذكاء يوسف ، (2012) ، " دراسة مقارنة بين الارتباط التقويمي والتقويمي الحصين " ، مجلة بحوث مستقبلية, 2012, المجلد 3, العدد 1, الصفحات 127-147،عدد39 .

5- Romanazzi, M, (1992) "Influence Function in Canonical Correlation Analysis" Pychometrika, Vol.57, No.2.

6- Zheng, Ch. , Yuan, L. , Sha, W. & Li-Sun, Z.,(2014) " Gene differential coexpression analysis based on biweight correlation and maximum clique " BMC Bioinformatics, 15(Suppl 15):S3

7- Veenstra, P. , Cooper, C. & Phelps, S. ,(2016)" The use of Biweight Mid Correlation to improve graph based portfolio construction " Computer Science and Electronic Engineering Conference, CEEC 2016 - Conference Proceedings ,pp. 101-106.

8- Croux, C. and Dehon, C, (2002), " Analyse Canonique basée sor des estimateurs robustes de La matrice de Covariance ", Revuede statistique Appliquee 2, PP. 5-26.

9- Song, Lin, (2012) "Comparison of co-expression measures: mutual information, correlation, and model based indices". BMC Bioinformatics. 13 (328).

10- Shevloykov, GL. and Oja, H., (2016) Robust Correlation: Theory and Applications (Wiley Series in Probability and Statistics) 1“st edition.

11- F. Hampel, E. Ronchetti, P. Rousseeuw, W. Stahel, (2011) Robust statistics: The approach based on influence functions.

12- Nasser, M. and Mesbahul, A. Md, (2006) "Estimators of Influence Function” Communications in Statistics—Theory and Methods, 35: pp, 21–32.