جمهــوريــة الســودان
جامعة السودان للعلوم والتكنولوجيا
كليــة الدراســات العليـا

التعدد الخطي أسبابه تأثيراته والمعالجة بانحدار الحافة وانحدار المركبات الرئيسية مع التطبيق على بيانات افتراضية

Multicollinearity its Causes, Effects and Remedies by Ridge Regression and Principle Components Regression with Application to Hypothetical Data

بحث لنيل درجة دكتوراه الفلسفة في الإحصاء

الدارس: محمد سليمان محمد جبريل المشرف الرئيس: أ. د. عبيد محمود محسن الزوبعي

المشرف المعاون: د. عادل موسى يونس

1435 هـ - 2014م

بسم الله الرحمن الرحيم

Sudan University of Science & Technology

College of Graduate Studies

جامعة السودان للعلوم والتكنولوجيا

كلية الدراسات العليا

صفحـــة الموافقـــة

اسم الباحث : [illegible]

عنوان البحث : ....

....

....

....

....

موافق عليه من قبل :

الممتحن الخارجي

الاسم: [illegible]

التوقيع: [illegible] التاريخ: 17-08-2014

الممتحن الداخلي

الاسم: د. أحمد محمد [illegible] حمدان

التوقيع: [illegible] التاريخ: [illegible]

*** المشرف**

الاسم: [illegible]

التوقيع: [illegible] التاريخ: 25-8-2014

cgs @ sustech edu. البريد الالكتروني فاكس/ 83 769363 ص.ب 407

[illegible]

[illegible]

ا لإ هـد ا ء

إلى روح والدتي العزيزة ...

إلى أبي أمد الله في أيامه...

إلى كل من علمني حرفا...

إلى زوجتي، وأبنائي مصعب، ندى، نون

إليهم جميعا أهدي ثمرة هذا الجهد

الـشكـر و الــتقـديــر

أولا الشكر لله الذي أوجدني من عدم ومهد لي سبيل العلم، والذي بتوفيقه تم هذا الجهد، والصلاة والسلام على سيدنا محمد وعلى آله وصحبه أجمعين إلى يوم الدين، ثم أما بعد:-

فلابد لي أن أتقدم بأسمى آيات الشكر والتقدير لجامعة السودان للعلوم والتكنولوجيا التي أتاحت لي فرصة الالتحاق بالدراسات العليا فيها. ويمتد شكري وتقديري لأستاذي العلامة الأستاذ الدكتور عبيد محمود محسن الزوبعي الذي اقتطع من وقته السمين في توجيهي وإسداء النصح لي، فلم يبخل أو يضن علي بفكره الثاقب إلى أن وصل هذا الجهد إلى خواتيمه. ويمتد شكري وتقديري للدكتور عادل موسى يونس المشرف المشارك الذي دعمني دعما منقطع النظير، وقدم لي ثاقب فكره وتوجيهاته النيرة التي استفدت منها كثيراً؛ لهما مني الشكر والتقدير وجعلهما الله ذخرا للعلم وللأمة العربية والإسلامية.

ويمتد شكري إلى كل من دعمني وشجعني على الاجتهاد والمثابرة من أجل إتمام هذا البحث أولهم أبي الذي شجعني كثيراً والإخوة والأصدقاء والزملاء جميعا لهم مني الشكر والتقدير، وإلى زوجتي العزيزة التي لم ينقطع تشجيعها لي،لها ولأبنائي الشكر والتقدير على صبرهم علي أثناء إعداد تقرير هذا البحث.

الباحث

قـائـمـة الـمـحـتـويـات

الموضوع		رقم الصفحة
الاهـــــداء		II
الشكـــــر		III
قائمة المحتويات		IV
قائمة الجداول		VI
قائمة الأشكال		VII
مستخلص البحث		IX
Abstract		XI
الفصل الأول: المقدمة		1
1 – 1	تمهيد	2
2 – 1	أهمية البحث	3
3 – 1	مشكلة البحث	4
4 – 1	أهداف البحث	5
5 – 1	فروض البحث	6
6 – 1	حدود البحث	6
7 – 1	الدراسات السابقة	7
8 – 1	هيكلية البحث	19
الفصل الثاني: بعض المفاهيم الأساسيات		20
1 – 2	تمهيد	21
2 – 2	نموذج الانحدار Regression Model	21
3 – 2	تقدير المعلمات بطريقة المربعات الصغرى Least Square Method	23
4 – 2	مصفوفة تباين – تغاير معلمات الانحدار	28
5 – 2	خصائص البواقي Residuals Properties	29
6 – 2	الاستدلال الاحصائي Statistical Inference	34
7 – 2	تشخيص النموذج Model Diagnostic	41
8 – 2	بناء النموذج واختيار المتغيرات	54
9 – 2	**نظام التحليل الاحصائي** Statistical Analysis System (SAS)	58
الفصل الثالث: التعدد الخطي Multicollinearity		60
1 – 3	تمهيد	61
2 – 3	التعدد الخطي Multicollinearity Problem	61
3 – 3	تشخيص التعدد الخطي Multicollinearity Diagnosis	67
4 – 3	طرق معالجة مشكلة التعدد الخطي	72
5 – 3	انحدار الحافة Ridge Regression	73
6 – 3	انحدارالمكونات الرئيسية Principle Component Regression	89

الموضوع		رقم الصفحة
الفصل الرابع: الجانب التطبيقي		96
4 – 1	تمهيد	97
4 – 2	تصميم الدراسة Study Design	97
4 – 3	دراسة محاكاة مونت كارلو Monte Carlo Simulation Design	99
4 – 4	نتائج دراسة المحاكاة	103
4 – 5	مناقشة النتائج	117
الفصل الخامس: النتائج والتوصيات		126
5 – 1	تمهيد	127
5 – 2	النتائج	127
5 – 3	التوصيات	129
المراجع		130
الملاحق		141

قائمة الجداول

رقم الجدول	عنوان الجدول	رقم الصفحة
1 – 2	جدول تحليل التباين للانحدار المتعدد	39
1 – 4	بعض الطرق المقترحة لتحديد معلمة الحافة k.	98
2 – 4	المتغيرات التفسيرية وفقا لعدد المشاهدات ومستوى الارتباط.	102
3 – 4	قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = .5).	104
4 – 4	قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (σ = .5).	105
5 – 4	قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = .5).	106
6 – 4	قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = .3).	107
7 – 4	قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (σ = .3).	108
8 – 4	قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = .3).	109
9 – 4	قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = .1).	110
10 – 4	قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (σ = .1).	111
11 – 4	قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = .1).	112
12 – 4	قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = 5).	113
13 – 4	قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (σ = 5).	114
14 – 4	قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = 5).	115
15 – 4	طرق مقدرات الحافة الأفضل أداء عند مستويات الارتباط والتباين وأحجام العينات وعدد المتغيرات المحددة.	116

قــائــمــة الأشـكــال

رقم الصفحة	عنوان الشكل	رقم الشكل
30	نموذج الانحدار الخطي	2 – 1
42	انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة عدم خطية دالة الانحدار.	2 – 2
42	انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تزايد التباين عند قيم المتغير التفسيري.	2 – 3
43	الرسم الصندوقي للكشف عن القيم الشاذة والمتطرفة	2 – 4
43	رسم الانتشار لقيم متغير الاستجابة والمتغير التفسيري للكشف عن القيم الشاذة والمتطرفة.	2 – 5
43	رسم الجذع والورقة للكشف عن القيم الشاذة والمتطرفة.	2 – 6
46	انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، لحالة ملائمة النموذج.	2 – 7
47	انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة ثبات التباين.	2 – 8
47	انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تناقص تباين حدود الخطأ مع تزايد قيم المتغير التفسيري.	2 – 9
48	انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تزايد تباين حدود الخطأ مع تزايد قيم المتغير التفسيري.	2 – 10
48	انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تناقص التباين ليصل حده الأدنى مقابل القيم المتوسطة للمتغير التفسيري، ثم تزايدها ليصل حده الأعلى مقابل القيم الكبيرة للمتغير التفسيري.	2 – 11
49	انتشار البواقي المعيارية مقابل المتغير التفسيري للكشف عن وجود قيم شاذة أو متطرفة.	2 – 12
50	انتشار البواقي المعيارية مقابل الزمن، عدم استقلالية حدود الخطأ أو الارتباط الذاتي، حالة تنازل قيم البواقي المعيارية مع الزمن.	2 – 13
50	انتشار البواقي المعيارية مقابل الزمن، عدم استقلالية حدود الخطأ أو الارتباط الذاتي، حالة تزايد قيم البواقي المعيارية مع الزمن.	2 – 14
50	انتشار البواقي المعيارية مقابل الزمن، عدم استقلالية حدود الخطأ أو الارتباط الذاتي، نمط متقلب دوريا تتغير معه إشارة الارتباط الذاتي.	2 – 15
51	انتشار البواقي المعيارية مقابل الزمن، حالة وجود قيم موجبة تتبعها قيم موجبة.	2 – 16
51	انتشار البواقي المعيارية مقابل الزمن، حالة وجود قيم سالبة تتبعها قيم سالبة.	2 – 17
52	رسم الاحتمال الطبيعي للبواقي المعيارية للكشف عن التوزع الطبيعي للبواقي.	2 – 18
55	مخطط بناء نموذج الانحدار الخطي المتعدد.	2 – 19

رقم الشكل	عنوان الشكل	رقم الصفحة
2 – 20	مخطط لتحليل مبسط للبيانات في برنامج SAS	59
3 – 1	مستويات الارتباط الخطي غير التام بين المتغيرات التفسيرية.	63
3 – 2	مقارنة مقدر منحاز بتباين أصغر بآخر غير منحاز بتباين أكبر.	74
3 – 3	التقدير بالمربعات الصغرى المقيدة في بُعدين، ومحيط المنحرف لمجموع مربعات البواقي ودالة انحدار الحافة المقيدة.	77
3 – 4	دالة طريقة المربعات الصغرى ودالة انحدار الحافة بالنسبة إلى عامل التحيز k.	78
3 – 5	أثر الحافة لثلاث دوال لمعلمات انحدار الحافة.	83

مستخلص البحث

يعتبر الانحدار الخطي المتعدد من التقنيات الاحصائية الأكثر استخداما بين الباحثين في مختلف المجالات، وكثيرا ما يواجه الباحثين مشكلة التعدد الخطي Multicollinearity عند بناء نموذج الانحدار الخطي المتعدد، عند وجود علاقة ارتباط تامة بين متغيرين تفسيريين أو بين جميع المتغيرات التفسيرية المضمنة في النموذج، بحيث يصبح محدد مصفوفة النظام $X'X$ يساوي صفرا حيث يستحيل ايجاد معكوس المصفوفة $X'X$ وبالتالي عدم امكانية استخدام طريقة المربعات الصغرى الاعتيادية OLS، أو أن يكون محدد المصفوفة $X'X$ قريبا من الصفر في حالة الارتباط غير التام، الذي معه تضمحل قدرة OLS على عكس الخصائص الحقيقية لمعلمات النموذج ويكون النموذج ذا قدرة تنبؤية ضعيفة. لتخطي مشكلة التعدد الخطي تم اقتراح كثير من الحلول منها طريقتي انحدار المكونات الرئيسية PCR، وانحدار الحافة RR . هدف هذا البحث إلى دراسة مشكلة التعدد الخطي وطرق تشخيصها وتأثيراتها على النموذج، ومقارنة أداء OLS وPCR وRR في معالجة مشكلة التعدد الخطي، وذلك من خلال محاكاة مجموعة من المتغيرات التفسيرية المرتبطة بمستويات مختلفة من الارتباط (γ = 0.7 to 0.99) وباحجام عينات (n = 10 , 50 ,100)، وأعداداً مختلفة من المتغيرات التفسيرية (p = 2 , 4 , 8 , 12). تم استخدام متوسط مربعات الخطأ MSE كمعيار لتقييم أداء مُقدرات الطرقة المختلفة. توصل البحث إلى أن كل من طريقتي PCR وRR قدمتا أداءاً أفضل من أداء OLS في جميع الحالات. وكذلك توصل البحث إلى أن PCR كانت الأفضل أداءاً بين الطرق المختلفة في حالة التعدد الخطي شبه التام (γ = 0.99) باستثناء طريقة انحدار الحافة K_{KM12} حيث كانت أفضل أداءاً. بينما في حالة التعدد الخطي المرتفع (γ = 0.7 ,0.9) في أغلب الأحوال قدمت طرق RRأداءاً أفضل مقارنة بطريقة OLS وطريقة PCR. وأوضحت نتائج المقارنة بين طرق RR المختلفة أن الطريقة K_{KM12}كانت الأفضل أداءاً وتليها الطرق K_{KHB}وK_{LW}وK_{KM2}وK_{MED}وK_jعند مستوى تباين (σ^2 = .25). وكذلك توصل البحث إلى أن الطرق (K_{KM12}وK_{KM2}، K_DوK_{NHSL}وK_{AS}وK_jوK_{arith}^{KS}وK_{KS})قدمت أداءاً أفضل عند ($p \leq 4$) بينما قدمت الطرق (K_{HK}وK_{HKB}وK_{LW}وK_{HSL}وK_{MED}) أداءا أفضل عند ($p \geq 5$). أيضا توصل البحث إلى أنه كلما زادت شدة التعدد الخطي زادت قيمة MSE تبعا لها بغض النظر عن مستوى التباين أو حجم العينة وذلك مع جميع الطرق المستخدمة OLSوPCR وRR بطرقها

المختلفة. وكذلك توصل البحث إلى أن حجم العينة يلعب دوراً مؤثراً في أداء PCR وأداء RRبحيث كلما زاد عدد المشاهدات صغرت تبعا لذلك قيم MSE. كذلك أوضحت النتائج أن قيم MSE تتزايد تبعا لتزايد التباين في المتغيرات التفسيرية ومتغير الاستجابة معاً والعكس صحيحاً. خلص الباحث إلى عدة توصيات منها : ضرورة استخدام PCR في حال وقوع النموذج تحت تأثير التعدد الخطي التام أو شبه التام.ضرورة استخدام RR عندما يكون النموزج واقعا تحت تأثير مشكلة تعدد خطي بحيث$(\gamma \le 0.9)$.ضرورة إجراء بحوث أخرى لدراسة العلاقة بين أداء الطرق PCR وRR وOLS ، ومستوى التباين داخل المتغيرات التفسيرية. ضرورة إجراء بحوث أخرى لدراسة PCR وRR وOLS ، والعلاقة بين مستوى تباين المتغيرات التفسيرية وقيم MSE.

Abstract

The multiple regression is one of the wide statistical techniques applied by researchers in various fields. Often face in these and other areas the problem of multicollinearity when they attempt to build a multiple linear regression model. When there is a semi or perfect correlation between two variables or among all Independent Variables (IV), whenthesystemmatrix$X'X$ and its determinant equals zero, therefore the inverse of $X'X$ will be absent and we can't calculate the Ordinary Least Square (OLS) parameters. Also in the case of high correlation when the $X'X$ determinant will be near to zero, the OLS could not reflect the real properties of the model parameters and results in inaccurate parameters. To overcome the problem of multicollinearity many solutions are presented amongst them are Principal Component Regression(PCR) and Ridge Regression(RR). The goal of this research is to study the problem of multicollinearity, methods of diagnosis, and its effects on the properties of model parameters, as well as to compare the performance of the method of OLS, PCR, and RR. By simulating a set of IVs associated with different levels of correlation ($\gamma = 0.7$ to 0.99) and samples sizes (n = 10, 50, 100). The models included different numbers of IVs included in the pattern (p = 2, 4, 8, 12). The Mean Square Error (MSE) was used as criterion to evaluate the performance of the different estimators. The researcher found that PCR and RRhave provided better performance than that of the OLS in all cases. In addition, it was found that the PCRhas the best performers among the different methods in the case of semi or perfect multicollinearity ($\gamma = 0.99$ or 1.0),with the exception of the methodK_{KM12}which is the best performer. While in the case of high correlation ($\gamma = 0.7$, 0.9) in most cases the methods of RR provided better performance compared to OLS and PCR. The results of the comparison between the different methodsof RRshowed that K_{KM12} was the best, followed by $\mathrm{K_{KHB}}$, $\mathrm{K_{LW}}$, K_{KM2} , K_{MED} , and K_j at level of variation $(\sigma^2 = 0.25)$. Moreover, the researcher found that the methods K_{KM12},K_{KM2},K_D,K_{NHSL}, K_{AS},K_j,K_{arith}^{KS}, K_{KS} have provided better performance in the case of little number of IVs $(p \le 4)$. While with large numbers of IVs (p ≥ 5) those methods K_{HK} , $\mathrm{K_{KHB}}$, $\mathrm{K_{LW}}$, K_{HSL} , and K_{MED}

provided better performance.Moreover, the results showed that with all methods the value of MSE increases according to the severity of multicollinearity regardless of the sample size and variance. In addition, the results showed that the sample size affects the performance of PCR and RR. Also the value of MSE increases when the variances of IVs increase. Out of all these results,the researcher recommended that: It is better to use the PCR in the case of semi or perfect multicollinearity. It is better to useone of the methods of $RR$when there is high multicollinearity ($\gamma \leq 0.9$). Also researcher recommend other research to study the relation between performance of PCR, RR, and OLS and the different levels of variation among IVs. Additionally, researchers recommended to conduct a researches to study the relationship between PCR, RR, and OLS performance and the values of MSE.

الفصل الأول

المقدمـــــــــــــة

الفصل الأول

1 – 1: تمهيد

تستخدم النماذج الخطية بصورة واسعة في مختلف مجالات العلم، ويعتبر الانحدار الخطي المتعدد واحداً من النماذج الخطية المستخدمة بكثرة في تحليل بيانات العديد من البحوث في المجالات الاقتصادية، والإدارية، والاجتماعية، والصحية، والطبية، والعلوم التطبيقية الأخرى. نموذج الانحدار الخطي المتعدد كغيره من النماذج الاحصائية الخطيةيقوم على توافر مجموعة من الفروض الإحصائية،وعند تخلف واحداً منها أو أكثر يتعرض النموذج إلى عدة مشكلات، أهمها ظهور ما يسمى بمشكلة عدم ثبات الأخطاء أو مشكلة عدم تجانس التباينHetroscedasticity، أو تخلف فرض استقلال الأخطاء، الذي ينتج عنه ما يعرف بمشكلة الارتباط التسلسلي، أو الارتباط الذاتي للأخطاء Autocorrelation، ومنها أيضا مشكلة تخلف الفرض الخاص بعدم وجود إرتباط خطي بين المتغيرات المستقلة،أو ما يعرفبالتعدد الخطي Multicollinearity، ويعبر عنها أيضا بعدم التعامد Non-orthogonalبين أعمدة مصفوفة المتغيرات المستقلة[6]$X'X$.

تعتبر مشكلة التعدد الخطي (التسامت)، أو الارتباط المتعدد Multicollinearityبين المتغيرات التفسيرية،واحدة من أهم وأكثر المشكلات التي تقف عقبة أمام الباحثين عند استخدام تحليل الانحدار الخطي المتعدد،وكذلك الانحدار غير الخطي والانحدار اللوجستي المتعدد.وهي تنشأ عندما يتضمن نموذج الانحدار أكثر من متغير تفسيري وتكون هناك علاقة ارتباطتامة أو قوية جدا بين اثنين أو أكثر من هذه المتغيرات، أو بين جميع المتغيرات. تجدر الإشارة هنا إلى أن مشكلة التعدد الخطي لا تقل أهمية وخطورة عن مشكلتي عدم تجانس التباينHetroscedasticity،أو الارتباط الذاتي Autocorrelation، خاصة أنه قد يتزامن مع حدوث مشكلة التعدد الخطي، وجود مشكلة الارتباط الذاتي،أو وجود مشكلة عدم تجانس التباين، وأحيانا قد تؤدي عملية إزالة الارتباط الذاتي إلى بروز مشكلة التعدد الخطي خاصة في بيانات السلاسل الزمنية[6][2][1].

تأتي أهمية مشكلة التعدد الخطي في أنها تؤدي إلى عدة مشكلات،أهمها: زيادة قيم كل من الأخطاء المعيارية والتباينات بدرجة كبيرة،مما يؤدي إلى حدوث تأثيرات سالبة

عنـد اتخـاذ الانحـدار الخطـي كأسـاس لإجـراء اختبـارات الفـروض، أو دقـة التقـديرات والتنبؤ.وكذلك كبر قيم تباين الخطأ، وتؤدي كذلك إلى اتساع فترات الثقة بدرجة كبيرة، كما قد يعاني النموذج من عدم الاستقرار، حيث يؤدي الارتباط الطردي العالي بين متغيرين تفسيريين أو أكثر إلى ظهور إشارةمعاملاتبعض هذه المتغيرات بإشارات سالبة تخالف ما هو مستقر في الواقع النظري أو التطبيقي[26][6].

لأهمية مشكلة التعدد الخطي في تحليل الانحدار الخطي، تم اقتراح طرق عديدة من قبل الكثير من العلماء والباحثين لحلها، ويمكن تقسيم هذه الطرق إجمالا إلى قسمين: أولهما الطرق غير المتحيزة Unbiased Methodsلتخطي مشكلة التعدد الخطي،وأهمها طريقة المربعات الصغرى الجزئية Partial Least Squares (PLS)، والانحدار المتين أو القوي Robust Regression (ROR)،والانحدار المعنوي SignificanceRegression (SIR)،والانحدار المتدرج Stepwise Regression(STR)،وغيرها من الأساليب غير الرسمية لتخطي مشكلة التعدد الخطي. ثانيها الطرق المتحيزة Biased Methodsومن أهمها طريقة المكونات الرئيسية Principal Components (PC)، وانحدار الجذور الكامنة Latent Root Regression (LAT)، ومقدرات بيز Bayes Estimator (BYS)، ومقدرات شرنكن Shrunken Estimator (J & S)،وانحدار الحافة المعمم Generalized Ridge Regression (GRR)، وانحدار الحافة Ridge Regression (RR) وتندرج تحته عدة مقدرات تختلف باختلاف طرق تقدير ثابت التحيز في انحدار الحافة[28][2][1].

1 – 2: أهمية البحث

تنبع أهمية البحث من أن الانحدار الخطي المتعدد يعتبر من التقنيات الإحصائية الواسعة الاستخدام من قبل الباحثين في معظم ميادين البحث العلمي، لدراسة العلاقة بين متغير الاستجابة والمتغيرات التفسيرية، أو للتنبؤ بقيم المتغير التابع عند تحديد قيم معينة للمتغيرات التفسيرية.وعندما ينعدم وجود علاقة بين المتغيرات التفسيرية يقال إن المتغيرات التفسيرية متعامدة، إلا أن أغلبالبيانات في الواقع التطبيقي لا تخلو من وجود علاقة ارتباط بين المتغيرات التفسيرية في نموذج الانحدار الخطي المتعدد.وعندما تكون العلاقة بينبعض المتغيرات التفسيرية أو جميعهاعلاقة ارتباطتامةPerfectMulticollinearityبحيث تصبح مصفوفة المتغيرات المستقلة $X'X$ مصفوفة

وحيدة Singular، أو تكون المصفوفة غير تامة ولكنها قوية أو قريبة من المصفوفة الوحيدة، تشكل هذه الحالة ما يعرف بالتعدد الخطي، وهو يعتبر معضلة حقيقية أمام الوصول إلى نتائج تعكس طبيعة العلاقة بين المتغيرات التفسيرية ومتغير الاستجابةبصورة وافية. ولعل نسبة كبيرة من الباحثين غير المتخصصين يغفلون هذه المشكلة ومنهم من هو ملم بها ولكن يصعب عليه إيجاد معالجة عملية مقبولة لهذه المشكلة. لذا تأتي أهمية هذا البحث في أنه يحاول الوصول إلى فهم أعمق لمشكلة التعدد الخطي من خلال[78][67]:

- بحث طبيعة مشكلة التعدد الخطي وتشخيص مدى تأثيرها على النموذج.
- تناول طرق الكشف عن ظاهرة التعدد الخطي،وأهم المظاهر الدالة على وجودها في نموذج الانحدار الخطي.
- دراسةمقدرات انحدار الحافةRRالمختلفة التي تم اقتراحها لتخطي مشكلة التعدد الخطي،ودراسةطريقة انحدار المكونات الرئيسيةPCRكصيغة مقترحة لتخطي مشكلة التعدد الخطي.

1 – 3: مشكلة البحث

عند بناء نموذج الانحدار الخطي المتعدد$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$لبيانات تعبر عن ظاهرة حقيقية في الواقع العملي، يتوقع وجود قدر من الارتباط بين بعض أو جميع المتغيرات التفسيرية، ولكن إذا كان هذا الارتباط تاماً: أو غير تام ولكنه عال جداًفنجد أن النموذج يقع تحت تأثير علة، وهي ما يعرف بالارتباط الخطي المتعدد أو التعدد الخطي والتيلها تأثيرات على مقدرات معلمات الانحدار الخطي ($\boldsymbol{\beta}$s) عند استخدام طريقة المربعات الصغرى OLSوتتمثل في: كبر تباين معلمات النموذج مما يؤدي إلى ظهور المعلمة ليست ذات دلالة احصائية نتيجة لانخفاض قيمة t بالرغم من أهمية المتغير في النموذج. وتتزايد الأخطاء المعيارية للتقديرات تبعا لزيادة درجة العلاقة بين المتغيرات التفسيرية مما يؤدي إلى اتساع فترات الثقة لهذه المقدرات. والحصول على إشارات غير حقيقية لمعلمات مقدرات الانحدار، كأن نحصل على معلمة بإشارة سالبة بما يخالف الحقيقة النظرية والعملية لمتغير تفسيري معين. وصعوبة اختبار معنوية النموذج والمتغيرات المفسرة وذلك لارتفاع قيمة

معامل التحديد (R^2) وقيمة اختبار (F)، ويترتب على ذلك أيضا الحصول على قيم تنبؤيهخاطئة.عليه يمكن صياغة مشكلة البحث في عدة جوانب هي كما يلي[80][78][27]:

- ما أفضل الطرق المقترحة لاكتشاف وتشخيص مشكلة التعدد الخطي في نموذج الانحدار الخطي المتعدد؟
- هل انحدار الحافة يمثل البديل الأفضل لتخطي مشكلة التعدد الخطي غير التام في نموذج الانحدار الخطي المتعدد؟
- هل طريقة المكونات الرئيسية هي البديل الأفضل لحل مشكلة التعدد الخطي التام في نموذج الانحدار الخطي المتعدد؟

1 – 4: أهداف البحث

يسعى البحث إلى تحقيق عدة أهداف من خلال الدراسة النظرية والتطبيقية لمشكلة البحث تتمثل فيما يلي:

1. تقييم أداء مقدرات انحدار الحافة كبديل لمقدرات طريقة المربعات الصغرى في حال بروز مشكلة التعدد الخطي.
2. التعرف على طرق اكتشاف التعدد الخطي وتشخيصه، ومستوى تأثيره على نموذج الانحدار الخطي المتعدد.
3. دراسة انحدار المكونات الرئيسية كطريقة أساسية لتخطي مشكلة التعدد الخطي، في حال وجود متغيرات تفسيرية ترتبط مع بعضها ارتباطا تاما أو غير تام.
4. تقييم قوة انحدار الحافة في تخطي مشكلة التعدد الخطي، في حال وجود متغيرات تفسيرية ترتبط مع بعضها البعض ارتباطا غير تام بدرجات متفاوتة.
5. تحديد أفضل الطرق لتقدير قيمة معلمة الحافة أو ثابت التحيز kعند مستويات مختلفة لشدة التعدد الخطي.
6. التعرف على كيفية تأثير التعدد الخطي في عملية الاستدلال الإحصائي، والتنبؤ في نموذج الانحدار الخطي المتعدد.

7. التعرف على كيفية استخدام انحدار الحافة كطريقة لاختيار المتغيرات الداخلة في نموذج الانحدار الخطي.
8. تقييم أداء مقدرات انحدار الحافة عند مستويات مختلفة من الارتباط الخطي غير التام.
9. المقارنة بين مقدرات طريقة المربعات الصغرى وانحدار الحافة تحت تأثير مستويات مختلفة للتعدد الخطي.

1 – 5: فروض البحث

يحاول الباحث من خلال هذا البحثاختبار صحة عدة فرضيات تتعلق بمعالجة مشكلة التعدد الخطي تتمثل في:

1. تقدم مُقدرات RRأداءاًأفضل من OLS عند تأثر النموذجبوجود ارتباط غيرتام بين المتغيرات التفسيرية.
2. تعتبر طريقة PCR أفضلأداءاً من طريقتيOLSو RRفي حال الارتباط التام أو شبه التام بين المتغيرات التفسيرية.
3. يقترب أداء الطرق الثلاثOLS و PCR و RR كلما قل مستوى التباين داخل المتغيرات.
4. كلما زاد مستوى التباين داخل المتغيرات التفسيريةكبرت قيم MSE.
5. كلما كبر حجم العينة كلما صغرت قيم MSE .
6. كلما زادت شدة تأثير التعدد الخطي كبرت قيمة MSE.

1 – 6: حدود البحث

سوف يقتصر الجانب النظري والتطبيقي للدراسة على تناول الانحدار الخطي المتعدد في حال وجود متغيرات مفسرة ذات مستوى قياس مستمر، ليس من بينها قيم مفقودة، أو قيم قاصية، ولا تعاني من مشكلة الارتباط الذاتي Autocorrelation أو عدم

ثابت التباين Heteroscedasticity، وسوف تقتصر الدراسة التطبيقية على بيانات افتراضية يتم توليدها بمستويات مختلفة للتعدد الخطي.

1 – 7: الدراسات السابقة

في عام (1986م) قدم Roger W. Hoerl وآخرون دراسة تحت عنوان : "محاكاة مقدرات منحازة وتقنيات اختيار المتغيرات في الانحدار"، هدفت الدراسة إلى مقارنة ثلاثة تقديرات منحازة هي Ridge Regression using the Lawless-Wang k (RRLW) و Ridge Regression using the k (RRLW)، و Principal Components (PC)، وأربعة تقنيات انحدار للاختيار الفئوي مع طريقة المربعات الصغرى الاعتيادية OLS ، باستخدام المحاكاة. قام الباحثون بإدخال مجموعة من المعلمات المناسبة للمقارنة، وروعي بشكل منهجي أن تكون متباينة ضمن مدى واسع لكل منها، حيث استخدمت مستويات مختلفة للارتباط تراوحت بين ($\rho = 0, 0.2, 0.4, 0.6, and\ 0.8$)، وأحجام مختلفة للعينات ($n = 17, 36$)، مع تثبيت قيمة التباين لجميع التجارب ($\sigma^2 = 1$)، وتم استخدام متوسط مربعات الخطأ MSE كمعيار للمقارنة بين المقدرات وطرق الاختيار المختلفة. في البدء لم يتم استخدام المعلمات المهمة في محاكاة التقنيات الفرعية، وتم ادخال نسبة من المتغيرات التفسيرية غير الضرورية ضمن البيانات. وكانت أهم النتائج أنه لا المقدرات المنحازة ولا تقنيات الاختيار الفرعية أظهرت تفوقاً مستقراً على المُقدرات الأخرى، في حين أظهرت طريقة الاختيار المتدرج Stepwise Selection أداءًا ضعيفا جدا، كما أظهر انحدار المكونات الرئيسية أداء ضعيفاً. وأوصت الدراسة بعدم التغاضي عن استخدام المقدرات المنحازة التي جرى اختبارها في هذه الدراسة، كما اوصت الرسالة باستخدام خوارزميات الاختيار كدليل أو للمقارنة المرجعية، وأوصت الدراسة كذلك بعدم استخدام طرق الاختيار المختلفة كاستراتيجية رئيسة في تخطي مشكلة التعدد الخطي، أيضا أوصت الدراسة بأفضلية استخدام انحدار RR الحافة على طريقة المكونات الرئيسية PC في تخطي مشكلة التعدد الخطي[58].

في العام (2006م) قدم Norliza Adnan وآخروندراسة بعنوان : " دراسة مقارنة على بعض طرق معالجة مشكلة التعدد الخطي ". هدف الباحثون من خلال هذه الدراسة إلى بناء معادلة خطية تعكس العلاقة بين جميع المتغيرات التفسيرية،ومتغير الاستجابة، عند بروز مشكلة التعدد الخطي بين المتغيرات المستقلة، كذلك هدفت الدراسة إلى تقييم أداء ثلاث طرق مقترحة لتخطي

مشكلة التعدد الخطي هي: انحدار الحافة RR، انحدار المكونات الرئيسية PCR، وانحدار المربعات الصغرى الجزئي PLSR، وذلك باستخدام طريقة محاكاة مونت كارلو على البرنامج S-plus، ولأغراض مقارنة أداء هذه الطرق الثلاث اعتمد الباحثون متوسط مربعات الخطأ MSE كمعيار للمقارنة. تم توليد عدد $(p = 2, 4, 6 \; and \; 50)$ من المتغيرات التفسيرية بأحجام عينات تراوحت بين $(n = 20, 30, 40, 60, 80 \; and \; 100)$ حيث تم استخدام الخوارزمية التالية لتوليد المتغيرات التفسيرية ومتغير الاستجابة:

$$x_1 = N(0,1)$$
$$x_{p-1} = N(0,1) + x_1$$
$$Y = x_1 + \cdots + x_p + N(0,1)$$

توصلت الدراسة إلى أن انحدار الحافة قدم أداء يفوق كل من انحدار المكونات الرئيسية والانحدار الجزئي للمربعات الصغرى في حالتي إدخال عدد محدود أو كبير من المتغيرات إلى النموذج (p = 2 or 50)، بينما أظهر كل من انحدار الحافة والانحدار الجزئي للمربعات الصغرى كفاءة عالية مقارنة بانحدار المكونات الرئيسية في حال إدخال عدد متوسط من المتغيرات إلى النموذج (p = 4 or 6)، كما توصلت الدراسة إلى وجود قدر كبير من الاتساق بين النتائج لمختلف أحجام العينات الصغيرة والكبيرة[12].

في العام (2009م) قدم Yazid M. Al-Hassan و Mowafaq M. Al-Kassabدراسة بعنوان : " دراسة مقارنة بين طريقتي انحدار الحافة وانحدار المكونات الرئيسية باستخدام طريقة مونت كارلو".استخدمت الدراسة طريقة محاكاة مونت كارلو لتقييم أداءًا كل من طريقة انحدار المكونات الرئيسية (PCR) وطريقة انحدار الحافة (RR) في حال وجود مشكلة التعدد الخطي، وتم استخدام متوسط مربعات الخطأ OLSكمعيار لتحديد أفضل الطرق أداءًا. صممت التجربة لمقارنة طريقة انحدار المكونات الرئيسية (PCR) وخمس طرق مختلفة لانحدار الحافة بناءًا على طرق تقدير ثابت التحيز $(k_{HK}, k_{HKB}, k_{LW}, k_{HSL}, k_{KS})$. وبواسطة المعادلة المستخدمة من قبل McDonald and Galarneau (1975) وآخرون، $x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ij}$ لتوليد عدد $(p = 20)$من المتغيرات التفسيرية وحجم عينة $(n = 30)$، وفقاً لمستويات مختلفة من الارتباط تراوحت بين $(\gamma^2 = 0.35, 0.51, 0.67, 0.84 \; and \; 0.99)$، وتم توليد قيم متغير الاستجابة

باستخدام المعادلة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$. توصلت نتائج الدراسة إلى أن جميع طرق انحدار الحافة قدمت أداءًا أفضل من طريقة انحدار المكونات الرئيسية تحت جميع مستويات الارتباط، وأعداد المتغيرات التفسيرية المستخدمة، وكانت مقدرات الحافة $(k_{HKB}, k_{LW} and\ k_{KS})$ على ذات الترتيب الأفضل أداءًا بين مقدرات الحافة الأخرى، وأوصت الدراسة باستخدام طريقة المكونات الرئيسية وانحدار الحافة كمقدرات منحازة، كما أوصت الدراسة كذلك باستخدام الطريقة (k_{HKB}) لتقدير معلمة الحافة (k) باعتبارها أفضل[14].

قدم هيثم يعقوب يوسف وآخرونفي عام (2010م) دراسة بعنوان : " استخدام الأساليب الإحصائية في معالجة مشكلة التعدد الخطي"، هدفت الدراسة إلى اختبار ومعالجة مشكلة التعدد الخطي من خلال استخدام طريقة المربعات الصغرى الاعتيادية OLS، وطريقة المكونات الرئيسية PC، وانحدار الحافة RR، بالتطبيق على نموذج مؤشرات اقتصادية للشركة العامة لصناعة البطاريات في العراق للفترة من (1992 – 2002)، وتم استخدام معيار الكفاءة النسبية للمقارنة بين أداء الطرق الثلاثة، حيث أوضحت النتائج أن طريقة انحدار الحافة RR كان لها أداءًا أفضل من طريقة المربعات الصغرى الاعتيادية OLS، بينما أظهرت طريقة المكونات الرئيسية PC أداء أفضل من طريقتي انحدار الحافة والمربعات الصغرى الاعتيادية. ومن أهم ما أوصت به الدراسة استخدام طريقة المكونات الرئيسية عندما يكون التعدد الخطي ناتج من ارتباط فوق (0.9) بين المتغيرات التفسيرية، وعندما يكون الارتباط بين (0.9 – 0.7) يمكن استخدام انحدار الحافة بجانب المكونات الرئيسية، أما عندما تتراوح قيمة معامل الارتباط بين (0.7 – 0.5) فيمكن استخدام طريقة المكونات الرئيسية وانحدار الحافة وطريقة المربعات الصغرى الاعتيادية. كذلك أوصت الدراسة باستخدام طرائق أخرى بديلة عند ظهور مشكلة التعدد الخطي للوصول إلى مقدرات كفوءة والطرق هي: طريقة المربعات الصغرى المقيدة Restricted Least Square، طريقة مزج بيانات السلسلة الزمنية والمقاطع العرضية Pooling Time Series and Cross-Section Data، وطريقة المربعات الصغرى المتدرجة Stepwise Least Square، وطريقة استخدام المعلومات المسبقة، وطريقة التقدير المختلط Mixed Estimation[9].

في عام (1975م) قدم Gary C. McDonald و Diane I. Galarneau دراسة بعنوان "تقييم بعض أنواع مقدرات الحافة بمحاكاة مونت كارلو"، هدفت الدراسة إلى تقييم طريقتين $(R_2 and\ R_3)$ تم اقتراحهما بواسطة الباحثين. تم استخدام طريقة محاكاة مونت كارلو لاختبار أداء الطريقتين بالتطبيق على بينات افتراضية، حيث تم توليد $(p = 3)$ من المتغيرات التفسيرية باستخدام المعادلة $x_{ij} = (1 - \alpha^2)^{\frac{1}{2}} z_{ij} + \alpha z_{i4}$ عند مستويات مختلفة من الارتباط تراوحت بين $(\alpha = 0.8, 0.9, 0.95\ and\ 0.99)$ وحجم عينة $(n = 100)$، وسبعة مستويات للتباين تراوحت بين $(\sigma = 0.01, 0.1, to, 0.5\ and\ 1.0)$، وتم استخدام المعادلة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i$ لتوليد قيم متغير الاستجابة، وتم استخدام عدة معايير لتحديد طريقة تحديد معلمة الحافة ذات الأداء الأفضل، أهمها متوسط مربعات الخطأ MSE ومتوسط مربعات الخطأ الكلي TMSE، ومربع بُعد متجه المعلمات المقدرة، ومجموع مربعات البواقي، ومربع معامل ارتباط القيم الفعلية والموفقة لمتغير الاستجابة Y. توصلت الدراسة إلى عدة نتائج تم عرضها وفقا لعدة معايير، فمنها النتائج المتعلقة بالأداء كدالة في (σ) حيث كان أداء (R_2) و (R_3) في حده الأدنى مماثلا لأداء OLS، وعند $(\sigma > \sigma(\alpha))$ يكون أداء الطريقتين المقترحتين (R_2) و(R_3) أفضل من طريقة المربعات الصغرى. وعند النظر إلى أداء المقدرات كدالة في الارتباط $(\alpha > \alpha(\sigma))$ فأظهرت الطريقتان (R_2) و (R_3) أداء أفضل مقارنة بطريقة المربعات الصغرى[76].

قدم G. R. Pasha و Shah, M. A. A.في العام (2004م) دراسة بعنوان " تطبيق انحدار الحافة على بيانات واقعة تحت تأثير التعدد الخطي". كان من أهم أهدافها استكشاف انحدار الحافة في حال وجود بيانات لها تعدد خطي، حيث ناقشت الدراسة كل من خصائص انحدار الحافة، معامل تضخم التباين VIF، القيم المميزة Eigen Values، وأيضا استعرضت الدراسة عدة طرق مقترحة لاختيار معلمة الحافة أو ثابت التحيز K هي(Ridge trace, k_{HKB}, k_{LW}, k_{PZ}, and Plotting C_k)، وتناول البحث مشكلة القيم الكامنة ومعايرة Standardization البيانات من خلال دراسة تطبيقية مقارنة بين طريقة المربعات الصغرى OLS وانحدار الحافة RR. وتوصلت الدراسة إلى أن طريقتي k_{HKB}, k_{PZ} هما الأفضل أداءًا بين طرق تقدير معلمة الحافة k، ومن جهة أخرى فإن طريقة انحدار الحافة بمختلف تقنيات

تقدير ثابت التحيز k المستخدمة أعطت مقدرات لها متوسط مربعات خطأ MSE أقل من متوسط مربعات الخطأ لمقدرات طريقة المربعات الصغرى[82].

في العام (2005م) قدم John Zhang و Mahmud Ibrahim دراسة بعنوان : "دراسة محاكاة انحدار الحافة وطرق المربعات الصغرى الاعتيادية على بيانات ذات تعدد خطي باستخدام البرنامج SPSS" هدفت الدراسة إلى مقارنة انحدار الحافة وطريقة المربعات الصغرى، مع تقييم الأداء للطريقتين في ظل وجود مشكلة التعدد الخطي، وتقييم الأداء باستخدام محاكاة مونت كارلو على البرنامج SPSS، كما هدفت الدراسة إلى تقديم نموذج جيد يعكس تأثير المتغيرات المفسرة على متغير الاستجابة. وتم استخدام متوسط مربع الخطأ MSE كمعيار لمقارنة أداء الطريقتين، كذلك تم تطبيق المعادلة المستخدمة من قبل McDonal and Galarneau (1975) وكل من Gibbons (1981) و Kibria (2003) $x_{ij} = (1-\rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{ij}$ لتوليد متغيرات تفسيرية واقعة تحت تأثير التعدد الخطي، حيث تم توليد عدة عينات بحجم $(n =)$ وفقا لمستويات مختلفة من الارتباط تراوحت بين $(\rho = 0.8, 0.9, 0.95, and\ 0.99)$، وكذلك تم استخدام قيم مختلفة للخطأ المعياري $(\sigma = 0.01, 0.1, 0.2, 0.3, 0.4, 0.5\ and\ 1.0)$، واستخدمت المعادلة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ لتوليد قيم متغير الاستجابة. توصل الباحثان إلى أن انحدار الحافة على برنامج SPSS لم يقدم تقديرات أفضل من تقديرات طريقة المربعات الصغرى، وفي حالات درجة التعدد الخطي العالية قدمت طريقة المربعات الصغرى تقديرات أفضل، عليه أوصت الدراسة الباحثين بعدم استخدام انحدار الحافة على برنامج SPSS الا بعد الدراسة المسبقة له[105].

قدم Mahdi A. Alkhamisi و Ghazi Shukur في عام (2007م) الباحثان دراسة بعنوان " دراسة بطريقة مونت كارلو لمعلمات انحدار الحافة الحديثة "، هدفت الدراسة إلى تقديم طريقة جديدة لتقدير معلمة الحافة ثابت التحيز في انحدار الحافة (K)، وإجراء دراسة مقارنة بواسطة طريقة مونت كارلو بين الطرق الجديدة المقترحة (K_{NHKB} ، K_{NAS} ، K_{ARITH} ، K_{NMED} ، K_{NLW}) وعدة طرق أخرى مقترحة لتقدير معلمة الحافة. وتم التطبيق على بيانات ذات مستويات مختلفة لشدة التعدد الخطي جرى توليدها باستخدام البرنامج S-plus حيث تم تطبيق المعادلة: $N_2(0, \Sigma), where\ \Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \rho =$

كمـــا تـــم اســـتخدام معادلــــة McDonald and Galarneau (1975) كبـديل لتوليـد المتغيـرات التفسـيرية، وتـم توليـد قـيم متغيـر الاسـتجابة باســــتخدام المعادلـــــة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ ، وبأحجـــــام عينـــات تراوحـــت بـــين $(n = 10, 20, 30, 40, 50\ and\ 100)$ ، وتوزيعـــات أخطـــاء مختلفـة. اسـتخدم الباحثـان أصـغر متوسـط مربـع خطـأ MSE كمعيـار لتحديـد الطريقـة الأفضـل. وكشــفت نتــائج البحــث أن واحــدة مــن هــذه الطــرق $(K_{NLW}$ ، K_{NMED} ، K_{ARITH} ، $K_{NAS})$ علـى الأقـل لهـا متوسـط مربـع خطـأ أقـل مـن متوسـط مربـع الخطـأ لكـل مـن طريقتـي OLS وK_{HK} ، كـذلك كشـفت نتـائج البحـث أن حجـم العينـة وقـوة الارتبـاط بـين المتغيـرات التفسـيرية وتوزيع الأخطاء لها تأثير مهم على خصائص مقدرات انحدار الحافة[19].

في عـام (2008م) قـدم Yazid M. Al-Hassan دراسـة بعنـوان : " تقيـيم لـبعض مقـدرات الحـرف باسـتخدام طريقـة مونـت كـارلو" هـدفت إلـى تقيـيم بعـض مقـدرات الحافـة وفقـا لاختيـار معلمـة الحافـة k ، حيـث تعرضـت الدراسـة إلـى سـبعة مـن الطـرق لتقـدير معلمـة الحافـة هـي: الطريقـة (k_{HK})، (k_{HKB})، (k_{LW})، (k_{HSL})، (k_{AM})، (k_{GM})، (k_{KS})، تـم توليـد قـيم عـدد مـن المتغيـرات التفسـيرية $(p = 5, 10\ and\ 20)$ عنـد مسـتويات مختلفـة مـن الارتبـاط الخطـي $(\gamma = 0.7, 0.8, 0.9\ and\ 0.99)$ باستخدام معادلة McDonald and Galarneau (1975) $x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ij}$ وعينات باحجام مختلفة $(n = 15, 25, and\ 30)$، وتم توليد قيم متغير الاسـتجابة باسـتخدام المعادلـة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$، ومـن ثـم استخدمت طريقـة مونـت كـارلو لتقيـيم أداء الطـرق المسـتخدمة، حيـث كـررت التجربـة لعـدد (5000) مرة، واعتمد الباحث متوسط مربعـات الخطـأ MSE كمعيـار لتحديـد أفضـل الطـرق أداءًا. وأوضـحت نتـائج الدراسـة أن تقيـيم الأداء كدالـة فـي مسـتوى الارتبـاط (γ) و(λ_1/λ_p) ووفقـا للمعطيـات $(n\ and\ p)$ لـوحظ أن المقـدر (k_{GM}) كـان الأفضـل أداءًا عنـد القـيم المتوسـطة لمعامـل ارتبـاط المتغيرات التفسيرية، ولكـن عنـد القـيم العليـا معامـل الارتبـاط قـدم المقـدر (k_{HKB}) أداءًا أفضـل مـن المقـدر (k_{GM}) وعنـد القـيم العاليـة جداً للارتبـاط $(\gamma = 0.99)$ جميـع المقـدرات باسـتثناء المقـدر (k_{AM})، كان لها أداءًا اما أفضل أو مثل أداء المقدر (k_{GM})، ولذات المعطيات وجد أنه عند تناقص قيم MSE لجميع المقدرات تتزايد النسبة (λ_1/λ_p). أما بالنسبة للأداء كدالة في معلمـة الحافـة (k)،

فان للمعطيات $(n \ and \ p)$، وجد أن قيمة معلمة الحافة (k) تتناقص لمعظم المقدرات عندما تتزايد النسبة (λ_1/λ_p) وكان أفضل المقدرات أداءًا (k_{AM}) و (k_{GM}) وهي التي تمتلك أكبر قيم لمعلمة الحافة (k). في حين توضح نتائج الأداء كدالة في $(n \ and \ p)$ أنه مع إعطاء قيم (γ) فان تزايد حجم العينة وعدد المتغيرات التفسيرية يصحبه تناقص في MSE لجميع المقدرات وللمقدر (k_{AM}) عند $(\gamma = 0.99)$، ولوحظ كذلك مع الأحجام الكبيرة للعينات وارتفاع الارتباط يكون أداء المقدر (k_{GM}) ضعيفا جدا مقارنة بالمقدرات الأخرى[15].

قدم Kristofer Mansson وآخرون في عام (2010م) دراسة بعنوان : " بعض مقدرات انحدار الحافة: دراسة باستخدام محاكاة مونت كارلو تحت قيود مختلفة لتباينات الخطأ" اهتمت الدراسة بعدة مُقدرات لمعلمة الحافة Ridge Parameter مقترحة من قبل الباحثين $(k_2, k_3, k_4, k_9, and \ k_{12})$ جميعها تعتمد على مقدرات كل من Khalaf and Shukur (2005) و Alkhamisi et al. (2006) و Muniz et al. (2010)، كما تمت مقارنتها بمجموعة أخرى من مقدرات انحدار الحافة المعروفة.تميزت هذه الدراسة بأنها طُبقت على متغيرات تفسيرية يتراوح عددها بين (4 – 12) متغير على خلاف كثير من الدراسات الأخرى التي استخدمت فيها ما بين (2 – 4) متغيرات تفسيرية. وتم توليد المتغيرات التفسيرية بواسطة المعادلة المستخدمة من قبل McDonald and Galarneau (1975) وآخرون $x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip}$، واستخدمت المعادلة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ لتوليد قيم متغير الاستجابة. استخدم كل من متوسط مربعات الخطأ MSE ومجموع مربعات التنبؤ $PRESS$ كمعايير للأداء، واستخدمت قيم مختلفة لتباينات الخطأ تتراوح بين (0.5 – 5) لمقارنة أداء المقدرات من خلال تطبيق طريقة المحاكاة. أوضحت نتائج الدراسة أن زيادة قوة الارتباط بين المتغيرات التفسيرية يؤدي إلى تأثيرات سالبة على متوسط مربعات الخطأ MSE ومجموع مربعات التنبؤ $PRESS$، في حين أن زيادة عدد المتغيرات المستقلة يكون له تأثير موجبا على متوسط مربعات الخطأ MSE ومجموع مربعات التنبؤ $PRESS$. أيضا أوضحت الدراسة أنه مع زيادة حجم العينة تتناقص قيمة MSE وإن كان الارتباط بين المتغيرات التفسيرية كبيراً. عليه فإن أداء المقدرات يعتمد بصفة أساسية على حجم خطأ التباين عندما يكون حجم العينة صغيراً. ومما توصلت إليه الدراسة أيضا أن المقدرات المقترحة

$(k_2, k_3, k_4, k_9, and\ k_{12})$ كان لها أداءًا أفضل من المقدرات الأخرى لمعلمة الحافة، واتضح أن المقدر k_{12}هو الأفضل أداءًا بين المقدرات المقترحة في جميع الأحوال[72].

قدم A. V. Dorugade وD. N. Kashid في العام (2010م) دراسة بعنوان : "طرق بديلة لاختيار مقدرات انحدار الحافة "، في هذه الدراسة اقترح الباحثان طريقة جديدة لاختيار معلمة الحافة (k_D)، وقام الباحثان بمقارنة أدائها بأداء مجموعة من طرق تحديد معلمة الحافة المعروفة باستخدام طريقة محاكاة مونت كارلو، وتم استخدام متوسط مربعات الخطأ MSE كمعيار لتحديد أفضل الطرق أداءًا، كما تمت مقارنة أداء جميع طرق اختيار معلمة الحافة المقترحة بواسطة *Hoerl and Kennard* و*Khalaf and Shukur* وآخرين وطريقة المربعات الصغرى. قام الباحثان بتوليد متغيرين تفسيريين إثنين مرتبطان ببعضهما بمستويات ارتباط $(\rho =$ 0.999, and 0.9999)، وبعينات ذات أحجام تراوحت بين $(n = 20, 50, 75, and\ 100)$، وتم استخدام تباينات الخطأ $(\sigma^2 = 5, 10, 25, and\ 100)$،وتم تكرار التجربة (1500) مرة للحصول على متوسط متوسطات مربعات الخطأ $AMSE$ لكل من طرق اختيار معلمة الحافة المستخدمة.كما تم أولاً حساب نسبة$AMSE$ لمقدر OLS لكل المقدرات، وثانيا تم حساب نسبة $AMSE$ لمقدر الحافة (k_{HKB})إلى مقدر OLS والمقدرات الأخرى. أوضحت النتائج أن طريقة اختيار معلمة الحافة المقترحة (k_D) لها أفضل أداء بين الطرق الأخرى لجميع مستويات معامل الارتباط بين المتغيرات التفسيرية، ولجميع قيم تباين الخطأ، ولجميع العينات بمختلف أحجامها[33].

في عام (2010م)قدمYazid M. Al-Hassan دراسة تحت عنوان " أداء مقدرات جديدةلانحدار الحافة "، حيث هدفت إلى دراسة أداء المقدر الجديد (k_{NHSL}) ومقدر طريقة المربعات الصغرى الاعتيادية OLS ومقدرات الطرق المقترحة بواسطة Horel and Kennard (k_{HK}) و(k_{HSL}) Hockin et al. حيث استخدم عدد من المتغيرات التفسيرية تراوحت بين قليلة ومتوسطة وكبيرة، قام الباحث بتوليد هذه المتغيرات مرتبطة خطيا بمستويات مختلفة من الارتباط $(\rho = 0.7, 0.8, 0.9, and\ , 0.99)$، باستخدام معادلة McDonald and Galarneau (1975)وآخرون$x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip}$،وبأعداد مختلفة للمتغيرات التفسيرية هي (p=5, 10, and 20)، وكانت العينات ذات أحجام مختلفة تتراوح بين (n = 15, 25, and 30)، ولتوليد قيم متغير الاستجابة تم استخدام المعادلة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$.

تم استخدام متوسط مربعات الخطأ MSE كمعيار للمفاضلة بين الطرق المختلفة، وتم تطبيق طريقة محاكاة مونت كارلو حيث نفذت التجربة لعدد (1000) مرة لقيم مختلفة لـ $(p, n, and\ \rho)$. وتوصلت الدراسة إلى سيطرة أداء المقدر المقترح على أداء الطرق الأخرى[16].

في عام (2012م) قدم GhadbanKhalaf دراسة بعنوان " معلمات حافة مقترحة لتحسين مقدر المربعات الصغرى"، هدفت الدراسة إلى تقييم أداء مُقدري معلمة الحافة (k_{HK}) و (k_{HKB}) مقارنة بأداء طريقة المربعات الصغرى الاعتيادية OLS، وتم توليد عينات للمتغيرات التفسيرية بأحجام ($n = 10, 100, and\ 1000$)، عند مستويات مختلفة للتعدد الخطي تراوحت بين ($\rho =$ 0.9, 0.99, 0.999, and 0.9999) وباستخدام المعادلة $x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip}$ التي إقترحها كل من McDonal and Galarneau (1975) وGibbons (1981) وKibria (2003)، واستخدم الباحث المعادلة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ لتوليد قيم متغير الاستجابة. تم إستخدام طريقة محاكاة مونت كارلو لتنفيذ التجربة لعدد (10.000) مرات، واستخدم الباحث متوسط مربعات الخطأ لاختيار أفضل الطرق أداءًا. أوضحت نتائج الدراسة أن الطرق المقترحة (k_{HK}) و (k_{HKB}) أفضل أداءًا من طريقة المربعات الصغرى OLS[62].

قدم Gisela Muniz وآخرون في عام (2012م) دراسة بعنوان "حول تطوير معلمات انحدار الحافة: استقصاء رسومي"، في هذه الدراسة تم استعراض تسعة مقدرات قديمة لمعلمة الحافة، وقُدم (12) مقدراً جديداً لمعلمة الحافة تتبع مقدرات كل من *Hoerl and Kennard (1970)*، *Kibria (2003)*، و*Khalaf and Shukur (2005)* هي $(k_{KM1}, k_{KM2}, \ldots k_{KM12})$، وقام الباحثون باستخدام طريقة محاكاة مونت كارلو لمقارنة أداء جميع هذه المقدرات، تحت تأثير مجموعة من العوامل المتوقع أن تؤثر على خصائص المقدرات التي يجري تقيمها، حيث جرى استخدام مستويات مختلفة للتعدد الخطي بين المتغيرات التفسيرية تراوحت بين ($\rho =$ 0.7, 0.8, and 0.9)، وأستخدمت أحجام مختلفة للعينات هي ($n =$ 10, 20, 30, 40, 50, and 100)، ومستويات مختلفة للانحراف المعياري للخطأ تراوحت بين ($\sigma =$ 0.01, 0.5, 1, 3, and 5) وكان عدد المتغيرات التفسيرية التي تم توليدها ($p =$ 2 and 4) . لتوليد قيم المتغيرات التفسيرية اتبع الباحثون ذات المعادلة المقترحة من قل كل من من *McDonal and Galarneau (1975)* و*Gibbons (1981)* و*Kibria (2003)* $x_{ij} =$

واســـــتخدمت المعادلـــــة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots +$ ، $(1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip}$ $\beta_p x_{ip} + e_i$ لتوليد قيم متغير الاستجابة،وتم استخدام متوسط مربعات الخطأ MSE كمعيار لتحديد أفضل المقدرات أداءً. أظهرت النتائج أن جميع مقدرات معلمة الحافة كان لها أداءً أفضل من أداء طريقة المربعات الصغرى الاعتيادية OLS،وخمس من هذه المقدرات كان لها أداءً أفضل هي $(k_{KM4}, k_{KM5}, k_{KM8}, k_{KM10}, k_{KM12})$وكذلك المقدرين$(k_{KM1}, k_{KM2})$ كان أداءها أفضل بالنظر إلى صغر متوسط مربعات الخطأMSE، ولوحظ كذلك أن المقدرين (k_{KM8}, k_{KM12}) لهما أداء أفضل في الحالات التي يكون فيها حجم تباين البواقي للنموذج كبيرا. وأوصت الدراسة باستخدام $(k_{KM4}, k_{KM5}, k_{KM8}, k_{KM10}, k_{KM12})$ مـن قبـل المسـتخدمين، كمـا أوصـت باسـتخدام (k_{KM8}, k_{KM12}) مع النماذج التي يكون لها تباين كبير للبواقي[79].

قـدم الباحثـان Fitrianto Anwar وYik Ceng Lee فـي عـام (2014م) تحـت عنـوان "أداء طـرق مُقـدرات انحـدار الحافـة فـي حالـة العينـات الصـغيرة مـع اخـتلاف مسـتويات معامـل الارتبـاط" هـدفت الدراسـة إلـى مقارنـة أداء طريقـة المربعـات الصـغرى الاعتياديـة OLS بـأداء ثـلاث مـن طـرق تقـدير معلمـة انحـدار الحافـة RR وهـذه الطـرق هـي K_{HK} و K_{HSL} و K_{NHSL} وذلـك باسـتخدام متوسـط مربعـات الخطـأ MSE معيـارا للمقارنـة. قـام الباحثـان بتوليـد قـيم المتغيـرات التفسـيرية وفقـا للصـيغة $x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip}$ عنـد مسـتويات مختلفـة لمعامـل الارتبـاط (ρ = .5 , .7 , and .9) ومسـتويات مختلفـة للتبـاين تمثلهـا الانحرافـات المعياريـة (σ = .1 , .5 , 1 , 5, 10)، ولقـيم متغيـر الاسـتجابة اسـتخدم الباحثـان الصـيغة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ باحجـام عينـات تسـاوي (20) مشـاهدة، ولحسـاب متوسـط مربعـات الأخطـاء تـم تكـرار التجربـة لعـدد (1000) مـرة . توصـل الباحثـان إلـى أن قـيم MSE لطريقـة المربعـات الصـغرى الاعتياديـة OLS كـان أصـغر منهـا لطريقـة K_{NHSL} عنـد (σ = .1)، بينمـا قـدمت الطريقـة K_{NHSL} أداء أفضـل مـن OLS عنـد (σ > .1) بينمـا قـدم K_{HK} أداء أفضـل مـن OLS ومقـدرات K_{HSL} و K_{NHSL} . كـذلك توصـلت الدراسـة إلـى أن قـيم MSE لمقـدرات انحـدار الحافـة تكـون متجانسـة إلـى حـد كبيـر عنـدما يكـون حجـم الانحـراف المعيـاري صـغيرا (σ = .1 , .5)، وتزداد قيم MSE كلما زاد حجم الانحراف المعياري[41].

التعليق على الدراسات السابقة:

جميع الدراسات والبحوث السابقة التي تم استعراضها قدمت دراسة مقارنة بين مقدر طريقة OLS ومقدرات RR وطريقة PCR، وطرق أخرى منحازة لتقدير معلمات الانحدار الخطي المتعدد في حالة وجود مشكلة التعدد الخطي. ونجد أن معظم هذه البحوث قامت على إجراء تجارب لاختبار المقدرات الأفضل أداءًا باستخدام طريقة محاكاة مونت كارلو، واستخدم معظمها MSE معيارا لتحديد المقدرات ذات الأداء الأفضل، كما استخدمت معظم هذه البحوث بيانات ذات درجات مختلفة من التعدد الخطي جرى توليدها باستخدام دوال مختلفة، حيث استخدم أغلب الباحثين الدالة $x_{ij} = (1-\gamma^2)^{\frac{1}{2}} z_{ij} + \gamma z_{ip}$ لتوليد قيم المتغيرات التفسيرية المرتبطة، بينما استخدمت الدالة $y_i = \beta_o + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip} + e_i$ لتوليد قيم متغير الاستجابة.

اهتمت بعض الدراسات بمقارنة الطرق الرئيسية الثلاث OLSوRRوPCR بالإضافة إلى بعض الطرق المنحازة الأخرى المقترحة لمعالجة مشكلة التعدد الخطي. فنجد بعض هذه الدراسات هدفت إلى مقارنة طريقة المربعات الصغرى OLS أو طريقة المربعات الصغرى الجزئية PLSR وطريقة انحدار الحافة RR وطريقة المكونات الرئيسية PC، وتصلت هذه الدراسات إلى أن مقدر انحدار الحافة كان الأفضل أداءًا مقارنةً بين الطرق الثلاث مثل دراسة Norliza Adnan وآخرون، ودراسة Yazid M. Al-Hassan و Mowafaq M. Al-Kassab، بينما دراسة هيثم يعقوب يوسف وآخرون، توصلت إلى أن كل من RR و PC أظهرت أداء أفضل من OLS في حين توصلت إلى أن PCR أفضل أداء من RR على خلاف دراسة Norliza Adnan و Yazid Al-Hassan. وقد أوصت بعض هذه الدراسات باستخدام RR كطريقة بديلة لمعالجة مشكلة التعدد الخطي، وكذلك أوصت بعضها باستخدام طريقة PCR في معالجة مشكلة التعدد الخطي وفقاً لمستويات محددة لشدة التعدد الخطي.

فيما قارنت دراسات أخرى بين الطرق الثلاث OLS و RR و PCRبالإضافة إلى طرق أخرى منحازة مقترحة لمعالجة مشكلة التعدد الخطي. حيث توصلت دراسة Hoerl

وآخرون إلى تقديم طريقة RRأداءًا أفضل من طريقة PCR، بينما أظهرت طريقة الاختيار المتدرج أداءًا ضعيفاً جدا. كذلك ركزت الدراسات على مقارنة OLS بمجموعة مختلفة من مقدرات RRوفقاً لطرق تحديد معلمة الحافة (k)، تحت شروط محددة تمثلت في حالة وجود مستويات مختلفة لشدة التعدد الخطي، ومستويات مختلفة للتباين، وعدد مختلف للمتغيرات التفسيرية واحجام مختلفة للعينات. توصلت بعض الدراسات إلى تقديم انحدار الحافة أداءًا أفضل من طريقة المربعات الصغرى الاعتيادية مثل دراسة McDonaldحيث أوصت باستخدام انحدار الحافة عند بروز مشكلة التعدد الخطي.

بينما قارنت بعض الدراسات بين أداء بعض مقدرات طرق تحديد معلمة الحافة kفيما بينها من جهة، ومن جهة أخرى بينها وبين طريقة OLS، معظم هذه الدراسات توصلت إلى تفوق طرق انحدار الحافة المختلفة على طريقة المربعات الصغرى، وفيما يتعلق بالمقارنة بين طرق تحديد معلمة الحافة فتوصلت دراسة Pasha وShah إلى تفوق مقدر (k_{PZ}) و (k_{HKB}). أما دراسة Alkhamisi و Shukur فتوصلت إلى أن المقدرات $(K_{NLW}$ ، K_{NMED} ، K_{ARITH} ، $K_{NAS})$ لها متوسط مربع خطأ أقل من متوسط مربع الخطأ لكل من طريقتي OLS وK_{HK}، وأوضحت دراسة Al-Hassan أنه عندما يكون الأداء دالة في مستوى الارتباط أظهر المقدر (k_{GM}) أداء أفضل بين المقدرات عند المستويات المتوسطة للارتباط، وعند القيم العليا للارتباط قدم المقدر (k_{HKB}) أداء أفضل من المقدر (k_{GM})، وعند القيم العالية جداً قدمت معظم المقدرات أداء أفضل من (k_{GM})، وعندما كان الأداء دالة في (k) فإن (k_{GM}) كان الأفضل أداءًا. وتوصلت دراسة Mansson تقديم المقدرات المقترحة $(k_2, k_3, k_4, k_9, and\ k_{12})$ أداءًا أفضل من المقدرات الأخرى. وأوضحت دراسة Dorugade و Kashid تفوق المقدر المقترح (k_D) على مقدرات معلمة الحافة الأخرى. وكذلك توصلت دراسة أخرى لـ Al-Hassan إلى سيطرة المقدر المقترح k_{NHSL} على مقدرات معلمة الحافة الداخلة في التجربة. وبينت دراسة Khalaf تفوق أداء مقدري معلمة الحافة (k_{HK}) و (k_{HKB}) على أداء طريقة المربعات الصغرى. بينما أوضحت دراسة Muniz وآخرون أن المقدرات $(k_{KM4}, k_{KM5}, k_{KM8}, k_{KM10}, k_{KM12})$ لها أداء أفضل من بقية المقدرات التي جرى

اختبارها. وأوصت معظم هذه الدراسات باستخدام مقدرات معلمة الحافة التي حازت على مستويات جيدة من الأداء في تخطي مشكلة التعدد الخطي.

1 – 8: هيكلية البحث

يتكون هذا البحث من خمس فصول نستعرضها هنا كما يلي:

الفصل الأول: ويشمل أهمية ومشكلة وأهداف وفروض وهيكلية وحدود البحث إضافة إلى الدراسات السابقة ذات العلاقة بموضوع البحث.

ويتكون الفصل الثاني:من بعض أساسيات الانحدار الخطي،وتقدير معلمات نموذج الانحدار الخطيباستخدام طريقة المربعات الصغرى، والمشاكل التي تواجه تقدير معلمات نموذج الانحدار الخطي المتعددباستخدام طريقة المربعات الصغرى، ليكون مدخلا لما يليه من الجانب النظري.

ويقدم الفصل الثالث:تعريفاًللتعدد الخطي، وتناول طرق التعرف على وجود التعدد الخطي وتشخيصه، وتعريف طريقة انحدار الحافة وطرق تقدير معلمة الحافة، وطريقة انحدار المكونات الرئيسية.

ويشمل الفصل الرابع:استعراضا لأهم طرق توليد البيانات وتحديد الطريقة المناسبة لتوليد المتغيرات التفسيرية ومتغير الاستجابة، وفقا لمستويات محددة من الارتباط الخطي، وأحجام مختلفة للعينات، وكذلك يتم تناول تصميم تجارب المحاكاة.

وفي الفصل الخامس:سيتم عرض أهم الاستنتاجات التي تم التوصل إليها من خلال تطبيق التجارب ومناقشتها، ومن ثم سيتم استعراض أهم التوصيات التي توصل إليها الباحث.

الفصل الثاني

بعض المفاهيم الاساسية

الفصل الثاني

بعض المفاهيم الاساسية

2 – 1:تمهيد

يتضمن هذا الفصل عرض نموذج الانحدار، وبعض أساسيات نموذج الانحدار الخطي المتعدد، والمشاكل التي تواجه نموذج الانحدار الخطي المتعدد ليكون مدخلا لما يليه من الجانب النظري والتطبيقي.

2 – 2: نموذج الانحدار Regression Model

إذا اعتبرنا الحرف Y يرمز لمتغير الاستجابة أو متغير الاستجابة الذي يرتبط خطيا بمجموعة p من المتغيرات التفسيرية أو المفسرة $X_1, X_2, \dots\ X_p$ وفقا للمعادلة (1.2):

$$\boldsymbol{Y = \beta_o + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p + \varepsilon} \qquad 1.2$$

تمثل المعادلة (1.2) ما يعرف بنموذج الانحدار الخطي المتعدد، وعندما يكون لدينا n مشاهدة لقيم Y ومجموعة من المتغيرات المفسرة X_i's ، يمكن صياغة المعادلة (1.2) بدلالة المصفوفات كما على المعادلة (2.2):

$$\boldsymbol{Y = X\beta +} \varepsilon \qquad 2.2$$

حيث:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_n \end{bmatrix} \qquad 3.2$$

أي أن**Y**تمثـل متجـه عمـود مـن الدرجـة nX1 يتكـون مـن n مشـاهدة لمتغيـر الاسـتجابة أو متغيــر الاســتجابة، وتفتــرض شــروط نمــوذج الانحــدار المتعــدد أن تكــون قــيم متغيــر الاسـتجابةYعلــى علاقــة خطيــة بــالمتغيرات التفســيريةالتي يتوقــع أن يكــون لكــل منهــا إســهام مقــدر فــي تبــاين متغيــر الاســتجابة، وأن يــتم قياســقيم متغيــر الاســتجابة بدقــة وأن تكــون صحيحة[6][5][2].

أما الرمز $\boldsymbol{X}$ فيمثل مصفوفة المتغيرات التفسيرية وهي كما يلي :

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{12} & \dots X_{1,p} \\ 1 & X_{21} & X_{22} & \dots X_{2,p} \\ 1 & X_{31} & X_{32} & \cdots X_{3,p} \\ \vdots & \vdots & \vdots & \vdots \quad \vdots \\ 1 & X_{n1} & X_{n2} & \cdots X_{n,p} \end{bmatrix} 4.2$$

المصــفوفة **X**مــن الدرجــة n × p تســمى مصــفوفة المتغيــرات التفســيرية أو مصــفوفة الثوابــت، يحتــوي العمــود الأول منهــا علــى القيمــة (واحــد) التــي تعبــر عــن المعامــل الثابــت، وحسـب فـروض نمـوذج الانحـدار الخطـي المتعـدد، يشـترط أن تكـون قـيم المتغيـرات التفسـيرية قــد قيســت بدقــة وصــحيحة، وأن يكــون تبــاين أي متغيــر مســتقل أكبــر مــن الصــفر $\sigma^2_{X_j} >$ 0, $\forall j = 1,2, \dots p$، وأن تكــون المتغيــرات التفســيرية عــن بعضــها الــبعض، أي أن تكــون أعمـدة مصـفوفة المتغيـرات التفسـيرية متعامـدة، وتكـون المتغيـرات التفسـيرية أيضـا مسـتقلة عن حد الخطأ العشوائي[4].

ويرمز لمتجه معلمات نموذج الانحدار المتعدد بـ $\boldsymbol{\beta}$:

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \qquad 5.2$$

يمثل العنصر الأول β_0 القيمة المتوقعة لمتغير الاستجابة **Y** عندما تكون قيم المتغيرات التفسيرية مساوية للصفر، بينما عناصر المتجه الأخرى من β_1 إلى β_p فكل منها يعبر عن التغير في القيمة المتوقعة لمتغير الاستجابة نتيجة للتغير في المتغير التفسيري المقابل لها بوحدة واحدة عند ثبيت أثر المتغيرات التفسيرية الأخرى. يتم تقدير قيم هذه المعلمات في نموذج الانحدار المتعدد باستخدام طريقة المربعات الصغرى بحساب ما يعرف بمقدرات المربعات الصغرى، وهي تتميز بأنها خطية بالنسبة إلى قيم متغير الاستجابة، وتعتبر غير متحيزة حيث تساوي القيمة المتوقعة لكل منها القيمة المقابلة لمعلمة المجتمع، وكذلك تتميز بأنها ذات أقل تباين (BLUE) بين جميع المقدرات الخطية غير المتحيزة[79][49].

ويرمز لمتجه حد الخطأ أو المتغيرات العشوائية بـ $\boldsymbol{\varepsilon}$:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ويفترض نموذج الانحدار الخطي أن ε عبارة عن متجه عمودي للمتغيرات العشوائية يتبع التوزيع الطبيعي، توقعه عبارة عن متجه عمودي صفري حيث $E\{\varepsilon\} = 0$، وتباينه عبارة عن مصفوفة التباين والتغاير $\sigma^2\{\varepsilon\} = \sigma^2 I$ وهو عبارة عن تباين متغير الاستجابة Y وهو ما يعرف بتجانس الخطأ Homoscedasticity ، وهو يتبع التوزيع الطبيعي $\varepsilon \sim N\,(0\,, \sigma^2 I)$.

2 – 3: تقدير المعلمات بطريقة المربعات الصغرى Least Square Method

يحتوي نموذج الانحدار الخطي المتعدد (2.2) على (p + 1) من المعلمات يمثلها المتجه $\boldsymbol{b} = (b_o, b_1, \dots, b_p)'$ ، وتعتبر طريقة المربعات الصغرى بأنها الأفضل بين الطرق المستخدمة في تقدير معلمات نموذج الانحدار الخطي، لأنها تمكن من الوصول إلى

أفضـل نمـوذج يجعـل مجمـوع مربعـات الأخطـاء أقـل مـا يمكـن، وبسـهولة يمكـن ايجـاد تقـديرات قيم **b** وفقا للإشتقاقات التالية[6][2] :

$$Y = Xb + e$$

$$\therefore e = Y - Xb$$

ومجموع مربعات الخطأ هو

$$Q = e'e =$$

$$=$$

عليه فإن

$$Q = e'e =$$

ولإيجـاد قـيم عناصـر المتجـه **b**التـي تجعـل مجمـوع مربعـات الأخطـاء $e'e$أصـغر مـا يمكن يتم أخذ المشتقة الجزئية بالنسبة لكل b_i من عناصر المتجه ومساواتها للصفر:

$$\frac{\partial Q}{\partial b} = \begin{bmatrix} \frac{\partial Q}{\partial b_o} \\ \frac{\partial Q}{\partial b_1} \\ \vdots \\ \frac{\partial Q}{\partial b_p} \end{bmatrix} = -2X'Y + 2X'Xb = 0 \qquad 8.2$$

وبعد التبسيط فإن

$$X'Xb = X'Y \qquad 9.2$$

وبضرب المصفوفة $(X'X)^{-1}$ في طرفي المعادلة (9.2) ضرب قبلي نحصل على :

$$b = (X'X)^{-1}X'Y \qquad 10.2$$

وللحصـــول علـــى مقــدرات معلمــات النمــوذج يشــترط أن تكــون أعمــدة المصـــفوفة $(X'X)$ غيـر مرتبطـة خطيـا بمعنـى عـدم وجـود ارتبـاط خطـي تـام بـين المتغيـرات التفسـيرية ويمكن إيجاد $(X'X)$ كما يلي:

$$X'X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & & \vdots \\ X_{1p} & X_{2p} & \cdots & X_{np} \end{bmatrix}$$

$$= \begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 & \cdots & \Sigma X_p \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & \cdots & \Sigma X_1 X_p \\ \Sigma X_2 & \Sigma X_2 X_1 & \Sigma X_2^2 & \cdots & \Sigma X_2 X_p \\ \vdots & \vdots & \vdots & & \vdots \\ \Sigma X_p & \Sigma X_p X_1 & \Sigma X_p X_2 & \cdots & \Sigma X_p^2 \end{bmatrix} \qquad 11.2$$

ويجـب أن تكـون أعمـدة المصـفوفة $(X'X)$ غيـر متعامـدة بحيـث تكـون مصـفوفة غيـر شـــاذة non-singular وبالتـــالي يمكـــن حســـاب معكوســـها $(X'X)^{-1}$ وإيجـــاد حـــل وحيـــد للمتجه **b** . أما المصفوفة $X'Y$ يمكن إيجادها وفقا لما يلي:

$$X'Y =$$

ويمكـن كتابــة المعادلــة (9.2) التــي تعبــر عــن المعــادلات الطبيعيــة بالمصــفوفات كمــا يلي:

$$\begin{bmatrix} n & \Sigma X_1 & \Sigma X_2 & \cdots & \Sigma X_p \\ \Sigma X_1 & \Sigma X_1^2 & \Sigma X_1 X_2 & \cdots & \Sigma X_1 X_p \\ \Sigma X_2 & \Sigma X_2 X_1 & \Sigma X_2^2 & \cdots & \Sigma X_2 X_p \\ \vdots & \vdots & \vdots & & \vdots \\ \Sigma X_p & \Sigma X_p X_1 & \Sigma X_p X_2 & \cdots & \Sigma X_p^2 \end{bmatrix} \begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \vdots \\ \Sigma X_p Y \end{bmatrix} \qquad 13.2$$

وإذا افترضـــنا أن $C = (X'X)^{-1}$ والمصـــفوفة $C_{(p+1)(p+1)}$ عبـــارة عـــن مصـــفوفة ثوابـــت ، فيمكن التعبير عن المعادلة (10.2) بالمصفوفات التالية[83][6][2] :

$$\begin{bmatrix} b_o \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} C_{00} C_{01} & C_{02} & \cdots & C_{0p} \\ C_{10} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{20} C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots \quad \vdots & \vdots & & \vdots \\ C_{p0} C_{p2} & C_{p2} & \cdots & C_{pp} \end{bmatrix} \begin{bmatrix} \Sigma Y \\ \Sigma X_1 Y \\ \Sigma X_2 Y \\ \vdots \\ \Sigma X_p Y \end{bmatrix} \qquad 14.2$$

2 – 3 – 1: خصائص المُقدرات β

إذا تـوفرت جميـع فـروض نمـوذج الانحـدار الخطـي المتعـدد حسـبما تشـير إليـه نظريـة جـاوس – مـاركوف فـإن مقـدرات طريقـة المربعـات الصـغرى **b** تكـون أفضـل المقـدرات، حيـث تكـون خطيـة وغيـر متحيـزة بالنسـبة إلـى $\boldsymbol{\beta}$ ويكـون لهـا أقـل تبـاين ويعبـر عنـه بـ (BLUE) وفيما يلي استعراضاً لهذه الخصائص[47][46][6][2]:

الخطية Linearity

تفتـرض أن تكونمقـدرات طريقـة المربعـات الصـغرى دوال خطيـة لمشـاهدات متغيـر الاسـتجابة، بمـا أن $b = (X'X)^{-1}X'Y$ وحيـث إن المقـدار $(X'X)^{-1}X'$ عبـارة عـن ثابـت عليه تكون مقدرات المربعات الصغرى دوال خطية في قيم متغير الاستجابة Y .

عدم التحيز Unbiasedness

هـذه الخاصـية تعنـي أن$\boldsymbol{b}$غيـر متحيـزة Unbiasedبالنسـبة إلـى $\boldsymbol{\beta}$، أي أن توقـع قـيم المتجة $\boldsymbol{b}$ تساوي القيم المناظرة لها في المتجه $\boldsymbol{\beta}$ ويمكن إثبات ذلك كما يلي:

$$b = (X'X)^{-1}X'Y$$

$$\therefore \; E(b) = E[(X'X)^{-1}X'Y] \qquad 15.2$$

نعوض عن Y=Xβ + ε

$$\therefore \; E(b) = E[(X'X)^{-1}X'(X\beta + \varepsilon)]$$

$$= E[(X'X)^{-1}X'X\beta + (X'X)^{-1}X'\varepsilon]$$

بما أن E(ε)=0 ، $(X'X)^{-1}X'X = 1$ فإن

$$E(b) = \beta \qquad 16.2$$

الكفاءة أو أقل تباينEfficiency

تعتبـر مقـدرات المربعـات الصـغرى ذات أقـل تبـاين مقارنـة بجميـع المقـدرات الخطيـة غير المتحيزة ويمكن إثبات ذلك كما يلي:

إذا كان

$$b = (X'X)^{-1}X'Y$$

$$b = AY$$

حيث $A = (X'X)^{-1}X'$

فـاذا افترضـنا أن المتجـه b^* مقـدر خطـي آخـر للمتجـه β مـن غيـر مقـدرات المربعـات الصغرى فإن:

$$b^* = (A + C)Y$$

حيث C تعتبر مصفوفة ثوابت

وبما أن $Y = X\beta + \varepsilon$ فإن

$$b^* = (A + C)(X\beta + \varepsilon)$$

$$= (A + C)X\beta + (A + C)\varepsilon$$

$$= [(X'X)^{-1}X' + C]X\beta + [(X'X)^{-1}X' + C]\varepsilon$$

$$= \beta + CX\beta + (X'X)^{-1}X'\varepsilon + C\varepsilon \qquad 17.2$$

ولكـي يكـون المتجـه b^* غيـر متحيـز بالنسـبة إلـى المتجـه β يجـب أن تكـون المصـفوفة CX مصفوفة صفرية $CX = 0$ وبذلك فإن :

$$b^* = \beta + (X'X)^{-1}X'\varepsilon + C\varepsilon \quad 18.2$$

$$b^* - \beta = (X'X)^{-1}X'\varepsilon + C\varepsilon$$

وبذلك يعرف تغاير – تباين المتجه b^* بالمعادلة التالية:

$$Var - cov(b^*) = E(b^* - \beta)(b^* - \beta)' \quad 19.2$$

$$Var - cov(b^*) = E[(X'X)^{-1}X'\varepsilon + C\varepsilon][(X'X)^{-1}X'\varepsilon + C\varepsilon]'$$

$$= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1} + C\varepsilon\varepsilon'X(X'X)^{-1} + (X'X)^{-1}X$$

$$+(X'X)^{-1}X'\varepsilon\varepsilon'C' + C\varepsilon\varepsilon'C']$$

$$Var - cov(b^*) = \sigma^2(X'X)^{-1} + \sigma^2CX(X'X)^{-1} + \sigma^2(X'X)^{-1}X'C' + \sigma^2CC'$$

$$Var - cov(b^*) = \sigma^2(X'X)^{-1} + \sigma^2CC' \quad 20.2$$

حيث إن $CX=0$ و $X'C' = 0'$

عليه فإن

$$Var - cov(b^*) = Var - Cov(\beta) + \sigma^2CC' \quad 21.2$$

أي أن

$$Var - cov(b^*) \geq Var - Cov(\beta) \quad 22.2$$

وعندما $C = 0$ فإن

$$Var - cov(b^*) = Var - Cov(\beta) \quad 23.2$$

2 – 4: مصفوفة تباين – تغاير معاملات الانحدار

تنبع أهمية مصفوفة تباين – تغاير معاملات الانحدار في أنها تمكن من حساب تباينات تقديرات معاملات الانحدار β على القطر الرئيسي لمصفوفة تباين – تغاير معاملات الانحدار وبما أن[22][6][2]:

$$Var - cov(b) = E[(b - \beta)(b - \beta)'] \quad 24.2$$

$$Var - cov(b) = E\{[(X'X)^{-1}X'\varepsilon][(X'X)^{-1}X'\varepsilon]'\}$$

حيث $b - \beta = (X'X)^{-1}X'\varepsilon$

عليه فإن

$$Var - cov(b) = (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}$$

$$Var - cov(b) = \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$Var - cov(b) = \sigma^2(X'X)^{-1}$$

وبما أن التباين مجهول عادة يتم استخدام المقدر S^2 وبذلك يحسب مقدر تباين وتغاير مقدرات المربعات الصغرى بالمعادلة (25.2)

$$S^2(b) = S^2(X'X)^{-1} \qquad 25.2$$

2 – 5: خصائص البواقي Residuals Properties

البواقي أو الرواسب هي القيم المقدرة لحد الخطأ العشوائي ε_i وهي تعبر عن الفروق بين القيم الفعلية لمتغير الاستجابة والقيم المقدرة المقابلة لها،أنظر شكل (2–1) ويرمز لها بـ e_i وتساوي[27][6][2] :

$$e = Y - \hat{Y} = Y - Xb \qquad 26.2$$

وبالتعويض عن b في المعادلة (2.12) بـ $b = (X'X)^{-1}X'Y$

$$e = Y - X(X'X)^{-1}X'Y = Y - HY$$

حيث $H = X(X'X)^{-1}X'$ مصفوفة مربعة وتعرف بمصفوفة القبعة HatMatrix وهي تستخدم في تحويل قيم متغير الاستجابة (Y) إلى القيم المقدرة (التوفيقية) Fitted Values$(\hat{Y})$ حيث:

$$\hat{Y} = Xb = X(X'X)^{-1}X'Y = HY \qquad 27.2$$

وبما أن مصفوفة القبعة تعتبر مصفوفة جامدة Idempotent Matrix فإن التعبيرات التالية صحيحة.

$$HH' = H'H = H \qquad 28.2$$

وبذلك يمكن التعبير عن متجه البواقي بما يلي:

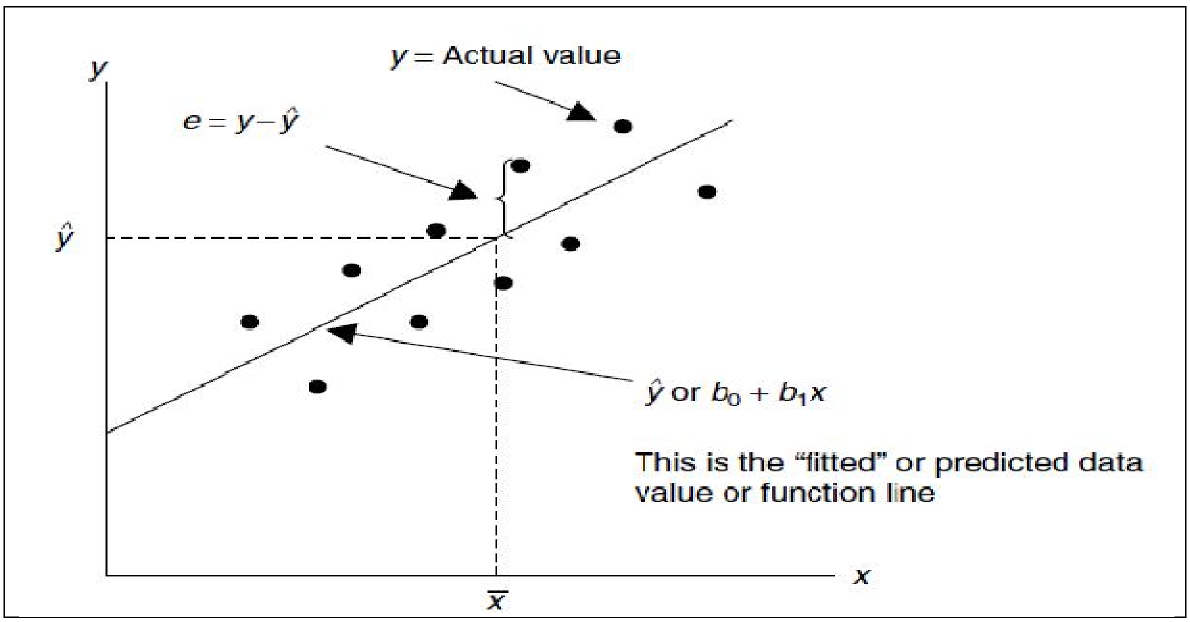
$$e = Y - HY \qquad 29.2$$

$$= (I - H)Y$$

$$= RY$$

حيث $R = (I - H)$ وهي كذلك مصفوفة جامدة .

شكل (2 – 1): نموذج الانحدار الخطي.



المصدر: مرجع [57]

القيمة المتوقعة للمتجه e تساوي صفراً:

$$E(e) = 0 \qquad 30.2$$

حيث

$$e = (I - H)Y$$

$$= (I - H)(X\beta + \varepsilon)$$

$$= X\beta - HX\beta + (I - H)\varepsilon$$

$$= X\beta - X(X'X)^{-1}X'X\beta + (I - H)\varepsilon$$

$$= X\beta - X\beta + R\varepsilon$$

$$= R\varepsilon$$

وأخذ التوقع لطرفي المعادلة يوضح أن توقع متجه البواقي يساوي صفراً كما يلي:

$$E(e) = E(R\varepsilon) = R\,E(\varepsilon) = 0 \qquad 31.2$$

البواقي مستقلة عن المتغيرات المفسرة:

$$X'e = 0 \qquad 32.2$$

$$X'e = X'(Y - Xb)$$

$$= X'Y - X'X(X'X)^{-1}X'Y$$

$$= X'Y - X'Y$$

$$= 0$$

البواقي مستقلة عن القيم المقدرة لمتغير الاستجابة:

$$\hat{Y}'e = 0 \qquad 33.2$$

بما أن $\hat{Y} = Xb$ فإن

$$\hat{Y}'e = b'X'e \qquad 34.2$$

$$= b'0$$

$$= 0$$

تباين متجه البواقي:

$$V(e) = \sigma^2 R \qquad 35.2$$

حيث إن

$$V(e) = E\{[e - E(e)][e - E(e)]'\} \qquad 36.2$$

وبما أن $E(e) = 0$ فإن

$$V(e) = E(ee') \qquad 37.2$$

وبما أن $e = R\varepsilon$ فإن

$$V(e) = E(R\varepsilon\varepsilon' R') \qquad 38.2$$

$$= R\sigma^2 I R'$$

$$= \sigma^2 R R'$$

$$= \sigma^2 R$$

حيث R تعتبر مصفوفة جامدة.

مقدر التباين:

$$\hat{\sigma}^2 = S^2 = \frac{e'e}{n-p-1} \qquad 39.2$$

يعتبر مقدر غير متحيز إلى σ^2 حيث:

$$e'e = [e_1 e_2 e_3 \quad . \quad . \quad . \quad e_n] \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ . \\ . \\ . \\ e_n \end{bmatrix} = \sum_{i=1}^{n} e_i^2 \qquad 40.2$$

ولكن $\sum_{i=1}^{n} e_i^2 = trace\,(ee')$

حيث:

$$ee' = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix} [e_1 e$$

أي أن

$$e'e = trace\ (ee') \qquad 42.2$$

$$E(e'e) = E(trace\ (ee'))$$

$$= trace\ E(ee')$$

$$= trace\ E(R\varepsilon\varepsilon' R')$$

$$E(e'e) = trace\ (\sigma^2 R)$$

$$= \sigma^2 trace\ (I - X(X'X)^{-1}X'$$

$$= \sigma^2 trace\ (I - X(X'X)^{-1}X'$$

$$= \sigma^2[trace\ I - trace\ (X'X)^{-1}X'X]$$

$$= \sigma^2\big(trace\ I_{nn} - trace\ I_{(p+1)(p+1)}\big) \qquad 43.2$$

بما أن عناصر القطر الرئيسي لـ I هي عبارة عن الواحد الصحيح فان:

$$E(e'e) = \sigma^2(n - p - 1) \qquad 44.2$$

عليه فإن مقدر التباين هو:

$$\frac{e'e}{n - p - 1} = \frac{\sum_{i=1}^{n} e_i^2}{n - p - 1} \qquad 45.2$$

وهو يعتبر مقدر غير متحيز إلى σ^2.

مجموع مربعات البواقي:

يحسب مجموع مربعات البواقي بالمعادلة التالية:

$$e'e = Y'Y - b'X'Y \qquad 46.2$$

ويمكن إثبات ذلك كما يلي:

$$e'e = (Y - Xb)'(Y - Xb)$$

$$= Y'Y - Y'Xb - b'X'Y + b'X'Xb$$

$$= Y'Y - Y'Xb - b'X'Y + b'X'X(X'X)^{-1}X'Y$$

$$= Y'Y - Y'Xb - b'X'Y + b'IX'Y$$

ولكن $Y'Xb = b'X'Y$ عليه فإن

$$e'e = Y'Y - b'X'Y \qquad 47.2$$

عليه يمكن التعبير عن مقدر التباين كما يلي:

$$S^2 = \frac{e'e}{n-p-1} = \frac{Y'Y - b'X'Y}{n-p-1} \qquad 48.2$$

2 – 6: الاستدلال الاحصائي Statistical Inference

المشاكل التي تصاحب عملية توفيق نموذج الانحدار الخطي المتعدد تقتضي الوصول إلى النموذج الأمثل، من خلال استخدام طرق الاستدلال الاحصائي المتمثلة في اختبار الفروض حول معلمات نموذج الانحدار وتقدير فترات الثقة لهذه المعلمات، وفيما يلي نستعرض هذه الطرق بدءًا بحساب معامل التحديد الذي يوضح مساهمة المتغيرات التفسيرية مجتمعة في تباين متغير الاستجابة ومن ثم نستعرض الطرق الاستدلالية من خلال بناء جدول تحليل التباين لنموذج الانحدار الخطي المتعدد.

2 – 6 – 1: معامل التحديد Coefficient of Determination

معامل التحديد المتعدد يكشف عن نسبة التباين في متغير الاستجابة التي تعود إلى المتغيرات التفسيرية، وبذلك فهو مقياس يعبر عن مدى القوة التفسيرية للنموذج، تتراوح قيمته بين الصفر

والواحد الصحيح، حيث كلما كانت قيمته قريبة من الواحد الصحيح دل ذلك على مقدرة النموذج في تفسير قدر كبير من التغير في متغير الاستجابة والاستقراءات حول المعلمات، لذلك يجب التأكد من خلو النموذج من مشاكل القياس المختلفة. يحسب معامل التحديد ويرمز له بـ R^2 بتقسيم مجموع المربعات الكلي Total Sum of Squares (SST) إلى جزئين أولهما يعرف بمجموع المربعات العائد إلى الانحدار Regression Sum of Squares (SSR) أو مجموع المربعات المفسرة وهو يشير إلى مجموع التباين في متغير الاستجابة الذي يعود إلى المتغيرات التفسيرية، وثانيهما يعرف بمجموع مربعات الخطأ Sum of Squares for Error (SSE) وهو يوضح الاختلاف غير المفسر الذي يعود إلى الصدفة والعشوائية، والمعادلة التالية تربط بين التباين المفسر والتباين غير المفسر[88] [92][86][6][2].

$$SST = SSR + SSE \qquad 49.2$$

معامل التحديد المتعدد R^2 عبارة عن النسبة بين مجموع مربعات الانحدار أو التباين المفسر والتغير الكلي، وتعبر عنه المعادلة (50.2).

$$R^2 = \frac{SSR}{SST} = \frac{b'X'Y - n\bar{Y}^2}{Y'Y - n\bar{Y}^2} \qquad 50.2$$

أيضا:

$$R^2 = 1 - \frac{SSE}{SST} \qquad 51.2$$

خصائص معامل التحديد:

1. تتراوح قيمة معامل التحديد المتعدد بين $0 \leq R^2 \leq 1$ إذا كانت جميع قيم $\hat{\beta}_j s$ تساوي صفرا، بإستثناء $\hat{\beta}_o$ إذا كانت قيمتها صفرا فإن $R^2 = 0$ ويستحيل حصول ذلك إذا كنا نتعامل مع بيانات مستمرة. إذا كانت جميع قيم Y تقع على سطح الاستجابة لنموذج الانحدار المتعدد أي أن $Y_i = \hat{Y}_i \,;\, i = 1,2,\ldots,n$ فإن $R^2 = 1$.

2. معامل الارتباط بين القيم التوفيقية والقيم المشاهدة لمتغير الاستجابة، يساوي معامل الارتباط المتعدد الذي يمثل الجذر التربيعي لمعامل التحديد المتعدد.
3. إضافة أي متغير مستقل لنموذج الانحدار المتعدد يؤدي فقط إلى زيادة قيمة معامل التحديد المتعدد R^2 أي يستحيل أن تنقص قيمته ويرجع ذلك إلى أن $R^2 = 1 - \frac{SSE}{SST}$، عليه لكي تتناقص قيمة معامل التحديد يجب أن يؤدي إضافة المتغير التفسيري إلى زيادة قيمة مجموع مربع الخطأ SSE بينما طريقة المربعات الصغرى تسعى إلى خفض قيمة SSE إلى أقصى درجة ممكنة.
4. إذا كان $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ فإن $E(R^2) = \frac{p}{n-1}$.
5. قيمة R^2 لا يمكن تقسيمها إلى p من المكونات الفريدة التي يعزى كل منها إلى X_j ، إلا إذا كانت قيم $X's$ متعامدة مع بعضها البعض.

يجب ملاحظة أن القيمة الكبيرة لمعامل التحديد المتعدد R^2 لا تعني بالضرورة أن النموذج جيد فقد يصاحب ذلك زيادة في قيمة متوسط تباين الخطأ MSE بما يجعل عملية الاستقراء غير مفيدة في الحالات التي تتطلب الدقة، هناك عدة حالات قد تظهر معها قيمة كبيرة لـ R^2، مثل بعض مشكلات القياس أو عمليات الاعتيان غير الجيدة أو عند إضافة متغيرات مستقلة جديدة للنموذج فقد تؤدي فقط إلى زيادة قيمة R^2 لأن p سوف تشكل نسبة كبيرة من n بالرجوع إلى الخاصية (4)، وكذلك قيمة التباين المفسر SSR لا تزداد أبدا مع إضافة متغيرات جديدة للنموذج، عليه يُقترح أحيانا استخدام معامل التحديد المعدل Adjusted Coefficient of Determination وهو عادة يعطي قيمة أقل وقد تكون سالبة مقارنة بقيم معامل التحديد العادي ويحسب وفقا للمعادلة (52.2)[102][6][2].

$$R^2_{adj} = \frac{\left(R^2 - \frac{p}{n-1}\right)(n-1)}{n-p-1} = \frac{(n-1)R^2 - p}{n-p-1} \qquad 52.2$$

أو بالمعادلة (53.2)

$$R^2_{adj} = 1 - \frac{\frac{SSE}{n-p-1}}{\frac{SST}{n-1}} = 1 - \frac{MSE}{MST} \qquad 53.2$$

2 – 6 – 2: معامل الارتباط المتعدد Multiple Correlation Coefficient

معامل ارتباط المتعدد R يقيس علاقة الارتباط بين قيم متغير الاستجابة المشاهدة Y والقيم المتوقعة $\hat{Y}$ ، وبعبارة أخرى يقيس مستوى تمثيل نموذج الانحدار المتعدد للتأثير المشترك لجميع المتغيرات التفسيرية على تباين متغير الاستجابة أو ملائمة النموذج للبيانات ومعامل الارتباط المتعدد عبارة عن الجذر التربيعي لمعامل التحديد المتعدد $R = \sqrt{R^2}$ وكذلك يحسب وفقاً لمعادلة بيرسون للارتباط معادلة (54.2)[6][84][2].

$$R_{Y\hat{Y}} = \frac{\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{S_{YY} * S_{\hat{Y}\hat{Y}}}} \qquad 54.2$$

وتتراوح قيمته بين $0 \leq R \leq 1$ أي أنه لا يأخذ قيما سالبة، وكلما إقتربت قيمته من الواحد الصحيح دل ذلك على أن القيم المقدرة قريبة جداً من القيم الفعلية لمتغير الاستجابة.

2 – 6 – 3: اختبار (ف) وجدول تحليل التباين F-Test and ANOVA Table

أهم نتيجة ينتهي إليها بناء جدول تحليل التباين ANOVA Table في نموذج الانحدار الخطي المتعدد هي قيمة احصاءةالاختبار (ف) F-Test Statistic ومستوى المعنوية المصاحب لها، وهي تجيب على التساؤل المتعلق بمعنوية تأثير المتغيرات التفسيرية مجتمعة على مستوى تباين قيم متغير الاستجابة، ويتم التوصل إلى ذلك من خلال اختبار صحة فرض العدم مقابل الفرض البديل حيث يصاغان كما يلي[6][2]:

$$H_o: \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

$$H_1: \beta_1 \neq \beta_2 \neq \cdots \neq \beta_p \neq 0$$

وتحسب احصاءة الاختبار $F_c = \frac{MSR}{MSE}$ كما على الجدول (2 – 1).

ويـتم ايجـاد قيمـة F_t الجدوليـة مـن جـدول توزيـع F بـدرجات حريـة P و (n-p-1)، وتكون قاعدة القرار تحت مستوى المعنوية α كما يلي:

إذا كان $F_c \geq F_{t(1-\alpha;n-p-1;p)}$ نرفض فرض العدم H_o ونقبل الفرض البديل أما،

إذا كان $F_c \leq F_{t(1-\alpha;n-p-1;p)}$ نقبل فرض العدم ونرفض الفرض البديل.

ولتكـوين جـدول تحليـل التبـاين وحسـاب الإحصـاءةF_c لابـد مـن إيجـاد مجـاميع مربعـات ومتوسطات التباين أو المكونات الأساسية لمعادلة الانحدار المتعدد وهي كما يلي:

مجموع المربعات الكلي (SST) وهو:

$$SST = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum Y_i{}^2 - n\bar{Y}^2 = Y'Y - n\bar{Y}^2 \qquad 55.2$$

مجموع المربعات العائد إلى الانحدار (SSR) وهو:

$$SSR = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2 = e'e = Y'Y - b'X'Y \qquad 56.2$$

مجموع المربعات العائد إلى الخطأ أو البواقي (SSE) وهو:

$$SSE = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2 = SST - SSR = b'X'Y - n\bar{Y}^2 \qquad 57.2$$

وعـادة تترافـق مـع مجمـوع المربعـات الكلـي SST(n – 1) درجـة حريـة، بينمـا تترافـق مـع مجمـوع المربعـات العائـد إلـى الانحـدار SSR(p) درجـة حريـة تمثـل عـدد المتغيـرات التفسـيرية فـي النمـوذج، فـي حـين ترافـق مجمـوع المربعـات العائـد الـي البـواقي SSE(n – p – 1) درجـة حريـة باعتبـار أننـا نحتـاج إلـى تقـدير (p – 1) مـن معلمـات دالـة الانحدار المتعدد. ويوضع جدول تحليل التباين للانحدار كما على الجدول (2 – 1)[6][2]:

جدول (2–1): جدول تحليل التباين للانحدار المتعدد.

Source of Variation	Sum of Squares	df	Mean Squares	F value
Regression	$SSR = b'X'Y - n\bar{Y}^2$	p	$MSR = \frac{b'X'Y - n\bar{Y}^2}{p}$	$\frac{MSR}{MSE}$
Residuals	$SSE = Y'Y - b'X'Y$	n-p-1	$MSE = \frac{Y'Y - b'X'Y}{n-p-1}$	
Total	$SST = Y'Y - n\bar{Y}^2$	n-1		

وتحسب قيمة F_c وفقا للمعادلة (58.2).

$$F_c = \frac{MSR}{MSE} = \frac{SSR/p}{SSE/(n-p-1)} = \frac{(b'X'Y - n\bar{Y}^2)/p}{(Y'Y - b'X'Y)/(n-p-1)} \qquad 58.2$$

وتجدر الاشارة إلى أنه قد يتم الحصول على قيمة كبيرة لـ F_c وتكون ذات دلالة إحصائية مما يعني وجود تأثير للمتغيرات التفسيرية على تباين متغير الاستجابة، بينما تكون المعلمات التي تم تقديرها جميعاً ليست ذات دلالة إحصائية، وهذه الحالة تعتبر من المؤشرات الدالة وجود مشكلة التعدد الخطي.

2 – 6 – 4: اختبارالفرضيات المتعلقة بمعاملات الانحدار الجزئية:

يهدف اختبار المعنوية لمعاملات الانحدار الجزئية إلى معرفة أهمية كل متغير من حيث تأثيره على متغير الاستجابة Y في ظل وجود المتغيرات التفسيرية الأخرى في النموذج، وفي هذه الحالة يتم استخداماختبار (ت) T-test لاختبار ما إذا كانت β_k تأخذ قيمة محددة حيث يتم اختبار فرض العدم التالي[6][2]:

$$H_o: \beta_k = 0$$

مقابل الفرض البديل.

$$H_1: \beta_k \neq 0$$

وتحسب احصاءة الاختبارT وفقا للمعادلة (59.2).

$$t = \frac{b_k - 0}{S_{b_k}} \sim t_{n-p-1} \qquad 59.2$$

ويكــون القــرار رفــض فــرض العــدم وقبــول الفــرض البــديل عنــدما تكــون قيمــة t المحســوبة المطلقــة أكبــر مــن قيمــة t الجدوليــة عنــد درجــات حريــة $(n-p-1)$ ومســتوى احتمــال$(1/2\,\alpha)$ ويــدل ذلــك علــى أن قيمــة β_k لا تســاوي الصــفر، أي أن المتغيــر التفســيري (X_k) يسهم بدلالة إحصائية في تفسير تباين متغير الاستجابة.

2 – 6 – 5: تقدير فترة الثقة لمعاملات الانحدار الجزئية:

بما أن متجه معاملات الانحدار يتبع التوزيع الطبيعي فإن تقدير فترة الثقة لمعاملات الانحدار الجزئية، وبما أن $\frac{b_k-\beta_k}{S_{b_k}}$ تتبع توزيع t فإن فترة الثقة $\{100(1-\alpha)\}$ لـ β_k تأخذ الشكل:

$$Pro\left(b_k - t\alpha_{/2(n-p-1)S_{b_k}} \le \beta_k \le b_k + t\alpha_{/2(n-p-1)S_{b_k}}\right) = 1-\alpha \qquad 60.2$$

ويمكن كتابة فترة الثقة لـ β_k كما في المعادلة (2.28).

$$b_k \pm t\alpha_{/2(n-p-1)S_{b_k}} \qquad 61.2$$

2 – 6 – 6: تقدير فترة الثقةلمتوسط الاستجابة:

للحصــول علــى تقــدير بنقطــة غيــر متحيــز لمتوســط الاســتجابة $\hat{Y}_0$ يــتم التعــويض فــي نمــوذج الانحــدار المقــدر $\bar{Y}_X = \hat{Y}_0 = X_0'b$ حيــث X_0' عبــارة عــن متجــه صــفي لقــيم المتغيرات التفسيرية المراد عندها تقدير قيمة متوسط متغير الاستجابة، أي أن [6][2]:

$$\bar{Y}_X = \hat{Y}_0 = \begin{pmatrix}1 & X_{10}X_{20} & \dots & X_{p0}\end{pmatrix}\begin{pmatrix}b_0\\ b_1\\ b_2\\ \vdots\\ b_p\end{pmatrix} \qquad 62.2$$

وبــافتراض أن حــد الخطــأ العشــوائي يتبــع التوزيــع الطبيعــي، فــإن $\hat{Y}_0$ تتبــع التوزيــع الطبيعــي بوســط حســابي قــدره $E(Y/X_0) = b_0X_0 + b_1X_1 + \cdots + b_pX_p$ وتبــاين

قــدره $\sigma^2 X_0'(X'X)^{-1}X_0$، وحيــث إن σ^2 مجهــول يــتم اســتخدام التبــاين المقــدر S^2 ، وبــذلك يمكن تقدير فترة الثقة لمتوسط الاستجابة وفقا للمعادلة التالية:

$$Pro\left(\hat{Y}_0 - t\alpha_{/2(n-p-1)} * S\sqrt{X_0'(X'X)^{-1}X_0} \le Y_0 \le \hat{Y}_0 + t\alpha_{/2(n-p-1)} * S\sqrt{X_0'(X'X)^{-1}X_0}\right) = 1 - \alpha \qquad 63.2$$

أو

$$\hat{Y}_0 \pm t\alpha_{/2(n-p-1)} * S\sqrt{X_0'(X'X)^{-1}X_0} \qquad 64.2$$

2 – 7: تشخيص النموذج Model Diagnostic

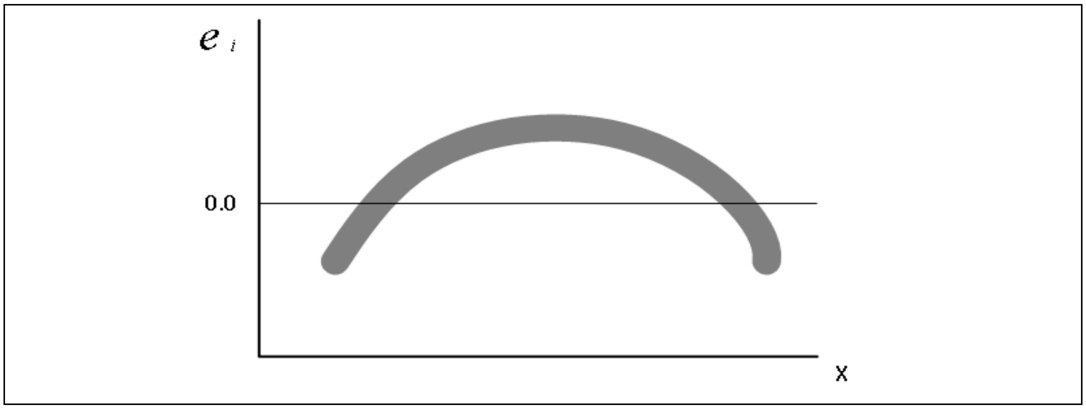
فحص وتشخيص نموذج الانحدار المتعدد يعد خطوة مهمة قبل الانتقال إلى خطوات تحليلية متقدمة تعتمد على النموذج، لأن الواقع التطبيقي يحفل بكثير من النماذج أو بعض خصائصها مثل خطية دالة الانحدار أو التوزع الطبيعي لحدود الخطأ التي قد تجعل النموذج غير مناسب للبيانات محل الدراسة. هناك نوعان من طرق فحص النموذج هما الطرق غير الرسمية وهي تعتمد على عمل مجموعة من الرسوم البيانية، والطرق الرسميةهي عبارة عن مجموعة من الاختبارات الاحصائية التي تحدد صلاحية النموذج[85][78][6] [42][2].

2 – 7 – 1: الطرق غير الرسمية لتشخيص النموذج

تبدأ مراحل تشخيص النموذج بالطرق غير الرسمية والتي تتمثل في دراسة الرسومات الخاصة بمتغير الاستجابة مقابل كل من المتغيرات التفسيرية أو استخدام الرسوم الخاصة بالبواقي للتأكد من عدم وجود مخالفات لفرضيات نموذج الانحدار الخطي المتعدد، ولإتمام مثل هذه التشخيصات عادة يتم استخدام عدة أنواع من الرسومات يمكن من خلالها اكتشاف بعض الاختلالات في النموذج مثل عدم خطية النموذج والالتواء أو عدم التوزع طبيعيا بالإضافةاكتشاف القيم الشاذة والمتطرفة، فمثلا في رسم الانتشار إذا كان انتشار النقاط ذو نمط شبيه بالمنحنى فيعتبر مؤشراً لعدم خطية دالة الانحدارشكل (2 – 2)، وفي حال إذا كان توزع النقاط على المستوى أشبه بالمروحة

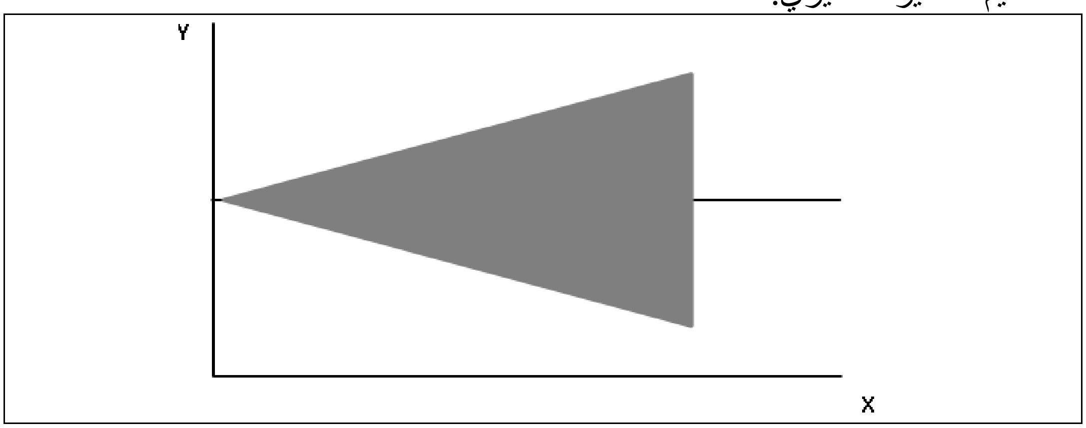
شكل (2 – 3) فإن ذلك مؤشر إلى عدم تساوي أو ثبات تباين حد الخطأ، وللكشف عن وجود قيم شاذة Outliers أو متطرفة Extremes عادة يتم استخدام الرسم الصندوقي Boxplot (2 – 4) ورسم الانتشار Scatterplot (2 – 5) ورسم الجذع والورقة Stem-and-leaf plots (2 – 6) [78][6] [51] [64][2].

شكل (2 – 2): انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة عدم خطية دالة الانحدار.



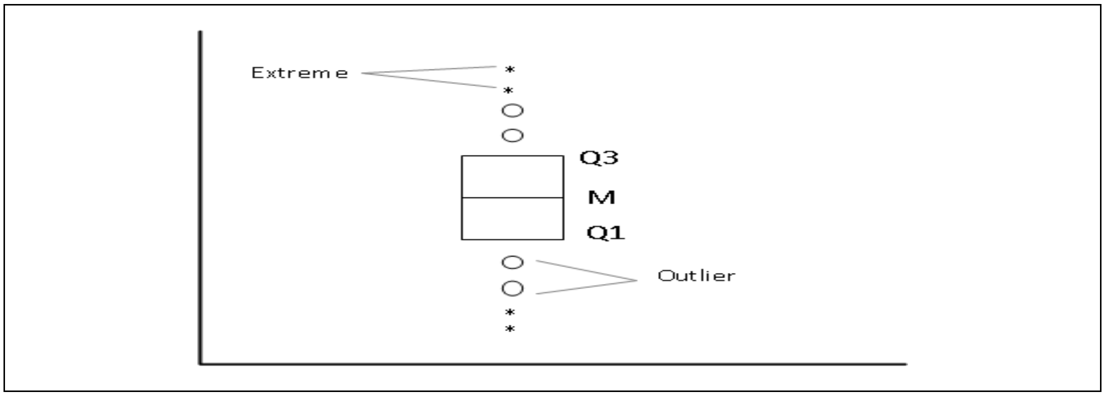
المصدر: تصميم الباحث.

شكل (2 – 3): انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تزايد التباين عند قيم المتغير التفسيري.



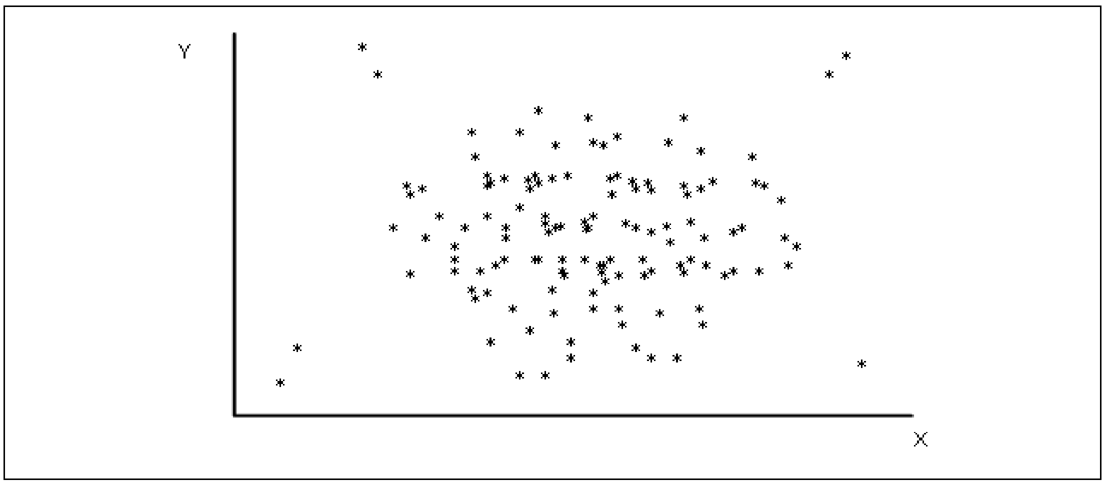
المصدر: تصميم الباحث.

شكل (2 - 4): الرسم الصندوقي للكشف عن القيم الشاذة والمتطرفة.



المصدر: تصميم الباحث.

شكل (2 – 5): رسم الانتشار لقيم متغير الاستجابة والمتغير التفسيري للكشف عن القيم الشاذة والمتطرفة.



المصدر: تصميم الباحث.

شكل (2 – 6): رسم الجذع والورقة للكشف عن القيم الشاذة والمتطرفة.

0	1225
1	1112222
1	55556667788899
2	1122222333333444444
2	5555666777788888888889999
3	00011122233 444444444
3	5555666778888999
4	0011223333
4	5556

المصدر: تصميم الباحث.

البواقي Residuals

تأتي أهمية فحص البواقي للتأكد من أهلية نموذج الانحدار الخطي من جهة أنها الأجدر في الكشف عن اختلالات نموذج الانحدار الخطي بصورة أكثر وضوحا، وذلك لأن قيم متغير الاستجابة دالة في قيم المتغير التفسيري لذا يتم تشخيص متغير الاستجابة في نموذج الانحدار الخطي بصورة غير مباشرة، تتمثل في فحص البواقي وهي عبارة عن الفرق بين قيم متغير الاستجابة الأصلية والقيم المقابلة المُتنبأ بها بواسطة نموذج الانحدار الخطي $e_i = Y_i - \widehat{Y}_i$ والباقي هنا يعبر عن الخطأ الملاحظ على خلاف الخطأ الحقيقي ε_iالذي يعتبر خطأ غير معروفا. ومن المسلم به افتراض أن ε_i متغير عشوائي يتوزع طبيعيا ومستقل بمتوسط يساوي صفر وتباين ثابت σ^2 وفي حال كان النموذج مناسبا للبيانات فإن البواقي e_i التي يتم الحصول عليها تعكس الخواص المتوقعة لـ ε_i وهي الفكرة الأساسية التي تقوم عليها عملية فحص البواقي للتأكد من صلاحية النموذج للبيانات.

خواص البواقي:

من خواص البواقي المهمة إن متوسط البواقي e_i يساوي صفراً حيث إن $\bar{e} = \sum e_i/n = 0$ لكل n من المشاهدات، وبما أن القيمة المتوقعة لـ e_i دائما تساوى صفر فهي في حقيقة الأمر لا توفر معلومات عن القيمة المتوقعة للأخطاء الحقيقة $E(\varepsilon_i) = 0$ ، ويعرف التباين لأي عدد n من البواقي على أنه $MSE = \frac{SSE}{n-p} = \frac{\sum e_i^2}{n-p} = \frac{\sum(e_i-\bar{e})}{n-p}$ ويكون MSE مقدِّرا غير منحاز لتباين حدود الخطأ σ^2 . تعتبر البواقي e_i متغيرات عشوائية غير مستقلة لأنها تحمل جزءًا من خواص القيم المقدرة $\widehat{Y}_i$ التي يعتمد تقديرها على معلمات النموذج $b_i \ , i = 0,1,\dots,n$ ، وتقل أهمية عدم استقلال البواقي مع كبر حجم العينة مقارنة بحجم المعلمات في نموذج الانحدار الخطي بما يسمح بتجاهله لمعظم الأغراض.

البواقي المعيارية Standardized Residuals

تعتبر البواقي المعيارية مفيدة جداً في الكشف عن القيم أو المشاهدات المتطرفة. تستخدم البواقي المعيارية في فحص البواقي باستخدام التباين لحدود الخطأ ε_i حيث يقدر بـ $V(e_i) = (1 - h_{ii})S^2$ عند إستخدام خصائص مصفوفة القبعة حيث h_{ii} عبارة عن العنصر القطري رقم iفي مصفوفة القبعة أو $MSE = \sigma^2$وباستخدام العنصر القطري يعرف المتبقي المعياري كما يلي[6] [78] [93][2]:

$$e_i' = \frac{e_i - \bar{e}}{\sqrt{V(e_i)}} = \frac{e_i}{\sqrt{(1 - h_{ii})S^2}} \qquad 65.2$$

Or

$$e_i' = \frac{e_i - \bar{e}}{\sqrt{MSE}} = \frac{e_i}{\sqrt{MSE}} \qquad 66.2$$

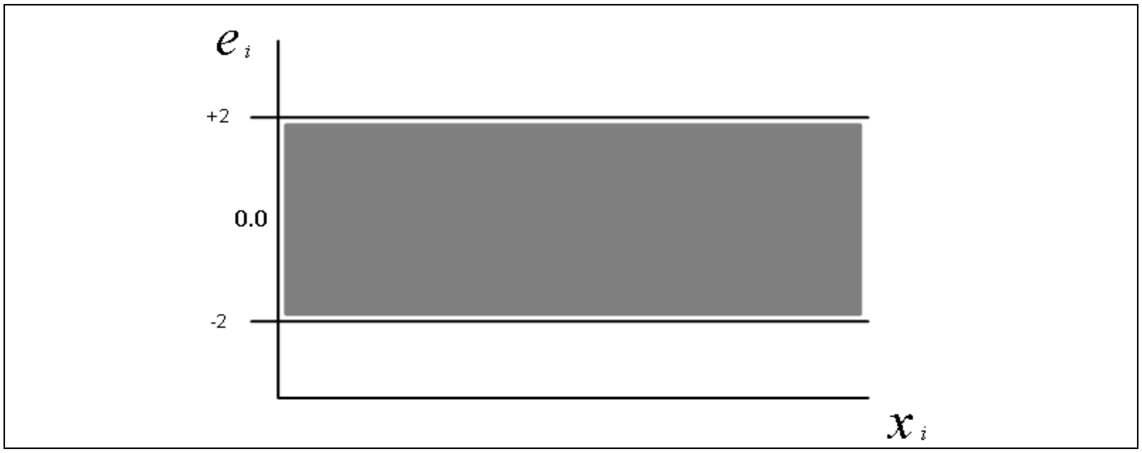
انحرافات النموذج

عملية فحص البواقي توفر طرقاً جيدة غير مباشرة للكشف عن استيفاء النموذج لفروض الانحدار الخطي والتأكد من ملائمته للبيانات محل الدراسة، وذلك من خلال اختبار ست أنواع من الانحرافات المحتملة للنموذج عن فروض الانحدار الخطي وهي؛ عدم خطية دالة الانحدار، عدم ثبات تباين حدود الخطأ، عدم استقلال حدود الخطأ، انحراف توزيع حدود الخطأ عن التوزيع الطبيعي، الكشف عن المشاهدات الشاذة أو المتطرفة، وفيما يلي استعراض لهذه الانحرافات بشيء من التفصيل من خلال استخدام بعض الرسومات التشخيصية للبواقي، ولكن نبدأ الاستعراض بالحديث عن الحالة المثالية أو الحالة التي يكون فيها النموذج ملائم للبيانات محل الدراسة.

حالة النموذج الملائم:

في حالة ملائمة نموذجالانحدار الخطي للبيانات محل الدراسة نجد أن شكل انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية أو مقابل القيم المقدرة بواسطة النموذج يعكس تبعثر عشوائي للنقاط بحيث تكون منحصرة بين خطين أفقيين بمتوسط يساوي صفراً وتباين ثابت، ويتوقع أن يكون 95% من النقاط محصورا بين حدين أفقيين (-2 و +2) والشكل (2 – 7) يوضح هذه الحالة.

شكل (2 – 7): انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، لحالة ملائمة النموذج.



المصدر: تصميم الباحث.

حالة عدم خطية دالة الانحدار:

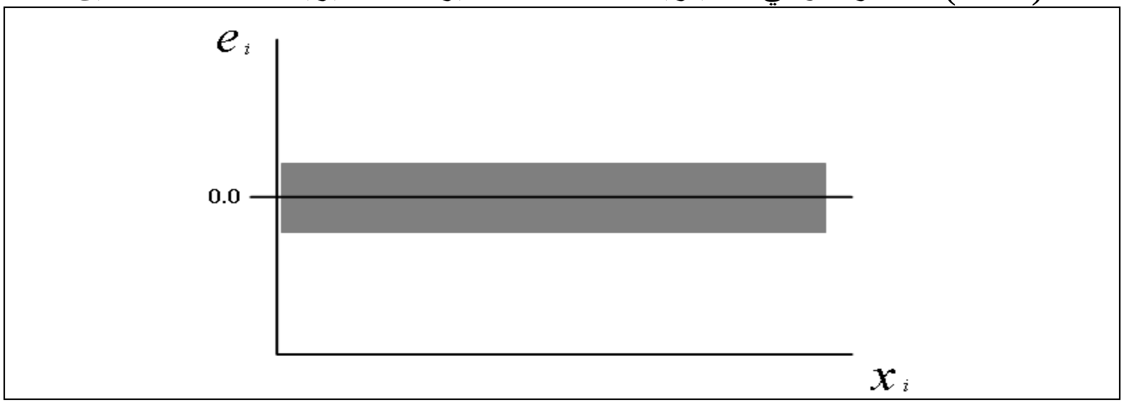
هنـاك عـدة رسـومات يمكـن مـن خلالهـا دراسـة مـا إذا كانـت دالـة الانحـدار لمجموعـة مـن المشـاهدات خطيـة مثـل رسـم البـواقي مقابـل المتغيـر التفسـيري أو رسـم البـواقي مقابـل القيم المُقـدَرة وجـدير بالـذكر أن هـذه الرسـومات يمكـن أن تعطـي صـورة أدق مـن رسـم الانتشـار وقـد تكـون مغـايرةً تمامـاً. عنـدما يكـون نمـوذج الانحـدار غيـر ملائـم للبيانـات محـل الدراسـة نجـد أن شـكل انتشـار البـواقي يأخـذ نمـط متناسـق حـول الصـفر حيـث نجـد أن البـواقي المقابلـة لقـيم المتغيـر التفسـيري الصـغرى تكـون سـالبة بينمـاتكـون موجبـةً مقابـل القيم المتوسـطة وتأخـذ قيمـا سـالبة مـرة أخـرى مقابـل قـيم المتغيـر التفسـيري الكبـرى شـكل (2 – 2) يوضـح هـذه الحالـة وهي توضح علاقة تربيعية.

حالة عدم ثبات تباين حدود الخطأ:

للكشـف عـن ثبـات تبـاين حـدود الخطـأ يـتم رسـم انتشـار البـواقي مقابـل المتغيـر التفسـيري أو قـيم التنبـؤ، ففـي حالـة ملائمـة النمـوذج للبيانـات فـإن حـد الخطـأ يكـون ثابتـا ويوضـح رسـم البـواقي مقابـل المتغيـر التفسـيري أو قـيم التنبـؤ شـكل أشـبه بالمسـتطيل حـول الصـفر والشـكل (2 – 8) يوضـح ذلـك، أمـا فـي حالـة عـدم ملائمـة النمـوذج للبيانـات فتختلـف قـيم حـد الخطـأ، ويعكـس رسـم البـواقي ثـلاث حـالات هـي: تنـاقص تبـاين حـدود الخطـأ مـع زيـادة قـيم المتغيـر التفسـيري شـكل (2 – 9)، زيـادة تبـاين حـدود الخطـأ مـع تزايـد قـيم المتغيـر التفسـيري شـكل (2 – 10)، وتنـاقص تبـاين حـدود الخطـأ ليصـل حـده الأدنـى مقابـل القـيم

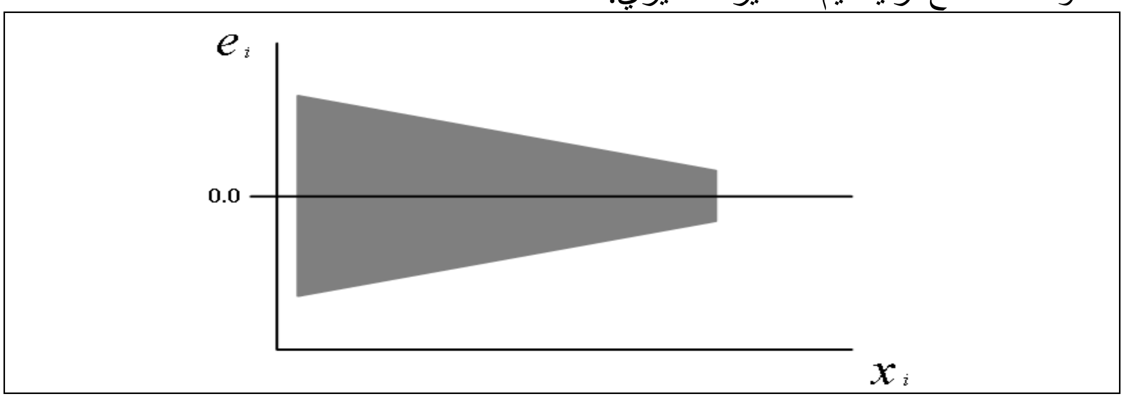
المتوسطة للمتغير التفسيري ثم يزداد تدريجيا ليصل حده الأعلى مقابل القيم الكبيرة للمتغير التفسيري شكل (2 – 11).

شكل (2 – 8): انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة ثبات التباين.



المصدر: تصميم الباحث.

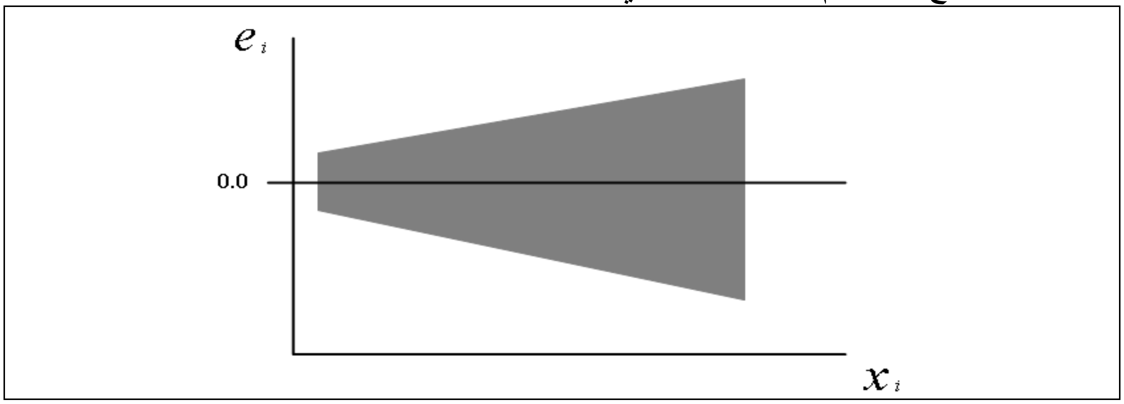
شكل (2 – 9): انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تناقص تباين حدود الخطأ مع تزايد قيم المتغير التفسيري.



المصدر: تصميم الباحث.

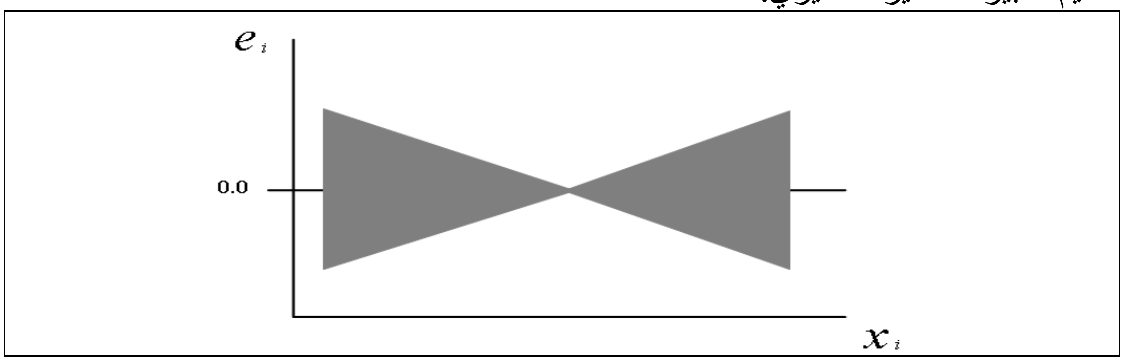
شكل (2 – 10): انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تزايد تباين

حدود الخطأ مع تزايد قيم المتغير التفسيري.



المصدر: تصميم الباحث.

شكل (2 – 11): انتشار البواقي المعيارية مقابل أحد المتغيرات التفسيرية، حالة تناقص التباين ليصل حده الأدنى مقابل القيم المتوسطة للمتغير التفسيري، ثم تزايدها ليصل حده الأعلى مقابل القيم الكبيرة للمتغير التفسيري.

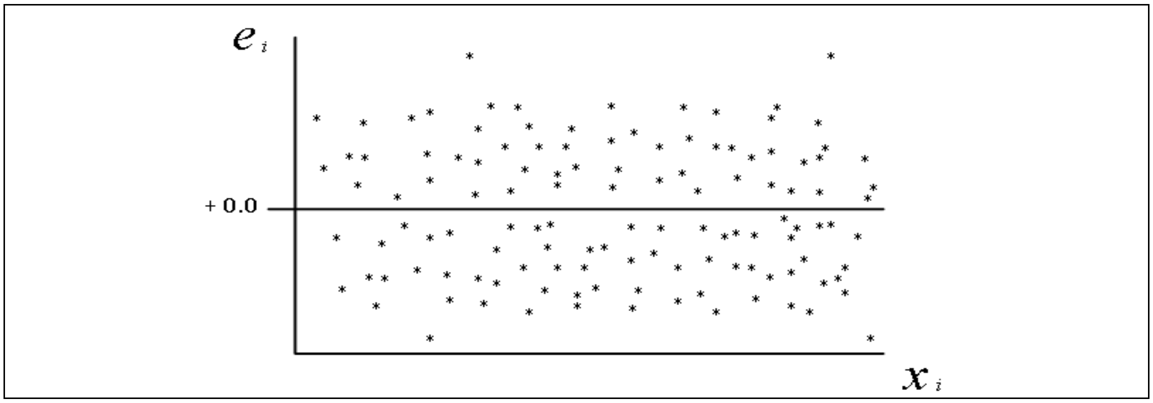


المصدر: تصميم الباحث.

الكشف عن القيم الشاذة أو المتطرفة:

القيم الشاذة أو المتطرفة يمكن أن تخلق إشكالاً كبيراً يؤدي إلى نقل النموذج لمعلومات مضللة عن الظاهرة محل الدراسة، لذا تأتي أهمية التأكد من عدم وجود قيم شاذة أو متطرفة، ومن الأفضل رسم الانتشار للبواقي المعيارية مقابل قيم المتغير التفسيري أو قيم التنبؤ أنظر شكل (2 – 12) ومن خلاله تظهر القيم الشاذة أو المتطرفة بعيدة عن مجموعة النقط، وكذلك يمكن استخدام عدة رسوم أخرى أهمها الرسم الصندوقي شكل (2-4- أ) حيث يمثل طوله الفرق بين قيمتي الربيع الأعلى والربيع الأدنى ويقع 50% من القيم داخل هذا المستطيل، ويرسم خط يتوسط المستطيل ليوضح قيمة الوسيط[6][2].

شكل (2 – 12): انتشار البواقي المعيارية مقابل المتغير التفسيري للكشف عن وجود قيم شاذة أو متطرفة.

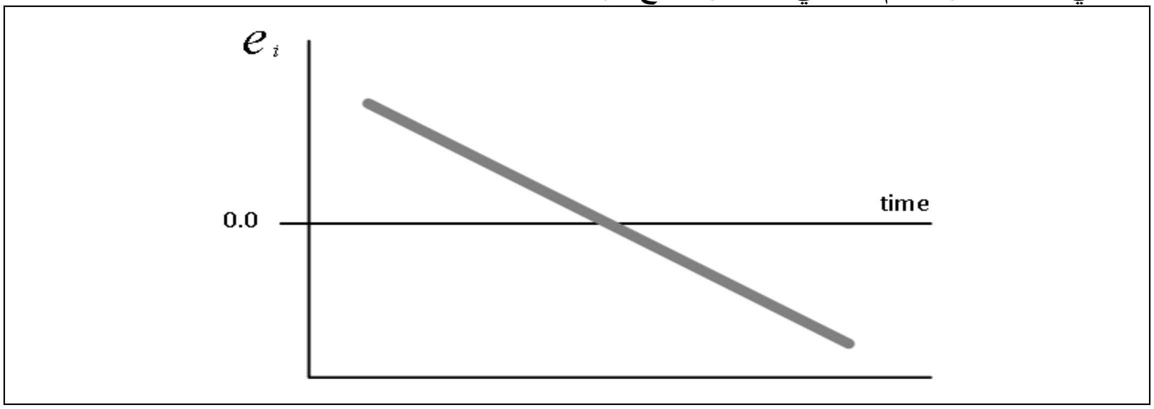


المصدر: تصميم الباحث.

حالة عدم استقلال حدود الخطأ:

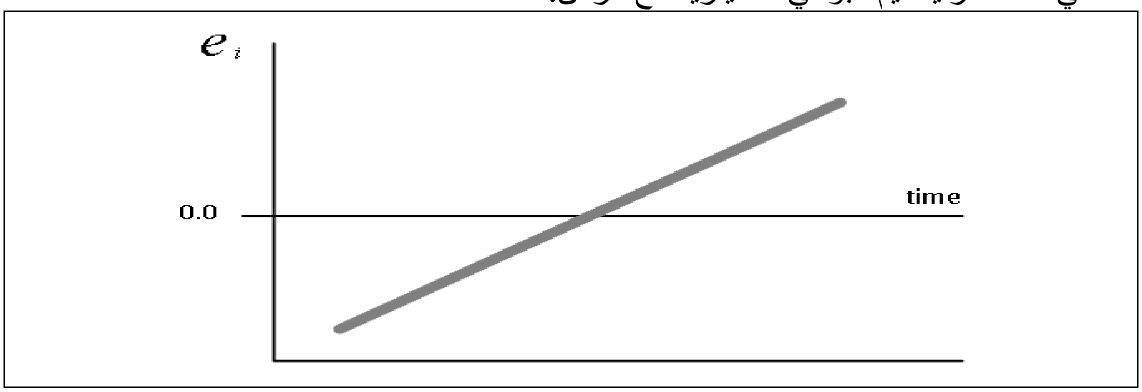
عند رصد البيانات في تتابع زمني (سلاسل زمنية)، أو عند حذف أحد المتغيرات التفسيرية المهمة، نجد أن حد الخطأ في فترة معينة e_i يكون مرتبطاً مع حد الخطأ في فترة زمنية قبله e_{i-1} ، فتنشأ ما تسمى بمشكلة عدم استقلالية حدود الخطأ أو الارتباط الذاتي، وتؤدي مشكلة عدم استقلالية حدود الخطأ إلى عدم تكافؤ مقدرات المربعات الصغرى، مما يتسبب في الوصول إلى اختبارات إحصائية وفترات ثقة غير صحيحة، وللكشف عن وجود هذه المشكلة : يتم رسم انتشار البواقي المعيارية مع الزمن، وعندما تكون حدود الخطأ مستقلة عن بعضها البعض يسفر رسم الانتشار عن توزع عشوائي للنقاط حول خط الأساس صفر، أما حالة عدم استقلالية حدود الخطأ أو الارتباط الذاتي تعبر عنها بيانيا عدة أنماط من رسوم الانتشار منها الرسم (2 – 13) حيث يعبر عن اتجاها تنازليا لقيم البواقي المعيارية مع الزمن والرسم (2 – 14)وهو يوضح اتجاهاً تصاعديا لقيم البواقي المعيارية عبر الزمن، في حين أن الرسم (2 – 15)يوضح علاقة ارتباطا ذاتيا للبواقي المعيارية مع الزمن ذات نمط متقلب دوريا تتغير معه إشارة الارتباط الذاتي، وكذلك من الطرق البيانية الأخرى للكشف الارتباط الذاتي يتم رسم انتشار البواقي المعيارية مقابل البواقي المعيارية في فترة سابقة، حيث إذا كانت قيمها موجبة تتبعها قيم موجبة شكل(2 – 16)أو قيم سالبة تتبعها قيم سالبة كما في الشكل (2 – 17)دل ذلك على وجود ارتباط ذاتي[6][2].

شكل (2 – 13): انتشار البواقي المعيارية مقابل الزمن، عدم استقلالية حدود الخطأ أو الارتباط الذاتي، حالة تنازل قيم البواقي المعيارية مع الزمن.



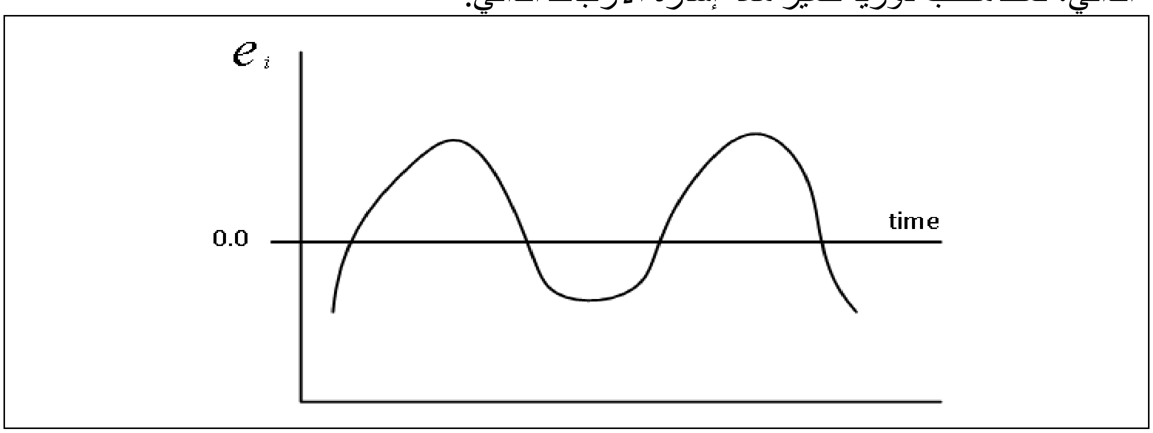
المصدر: تصميم الباحث.

شكل (2 – 14): انتشار البواقي المعيارية مقابل الزمن، عدم استقلالية حدود الخطأ أو الارتباط الذاتي، حالة تزايد قيم البواقي المعيارية مع الزمن.



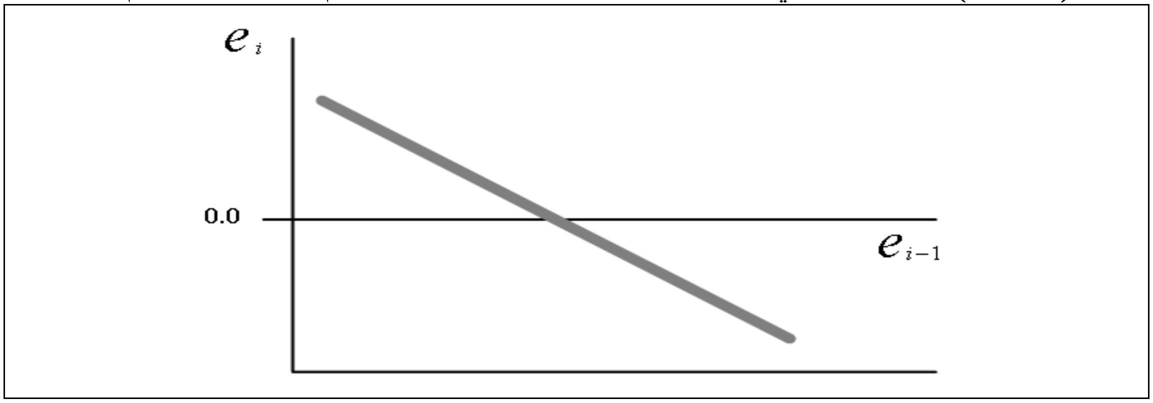
المصدر: تصميم الباحث.

شكل (2 – 15): انتشار البواقي المعيارية مقابل الزمن، عدم استقلالية حدود الخطأ أو الارتباط الذاتي، نمط متقلب دوريا تتغير معه إشارة الارتباط الذاتي.



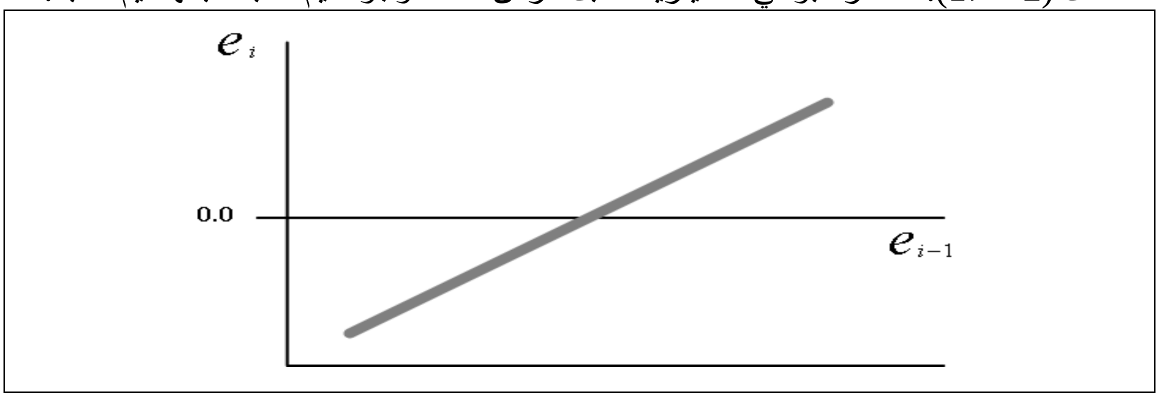
المصدر: تصميم الباحث.

شكل (2 – 16): انتشار البواقي المعيارية مقابل الزمن، حالة وجود قيم موجبة تتبعها قيم موجبة.



المصدر: تصميم الباحث.

شكل (2 – 17): انتشار البواقي المعيارية مقابل الزمن، حالة وجود قيم سالبة تتبعها قيم سالبة.



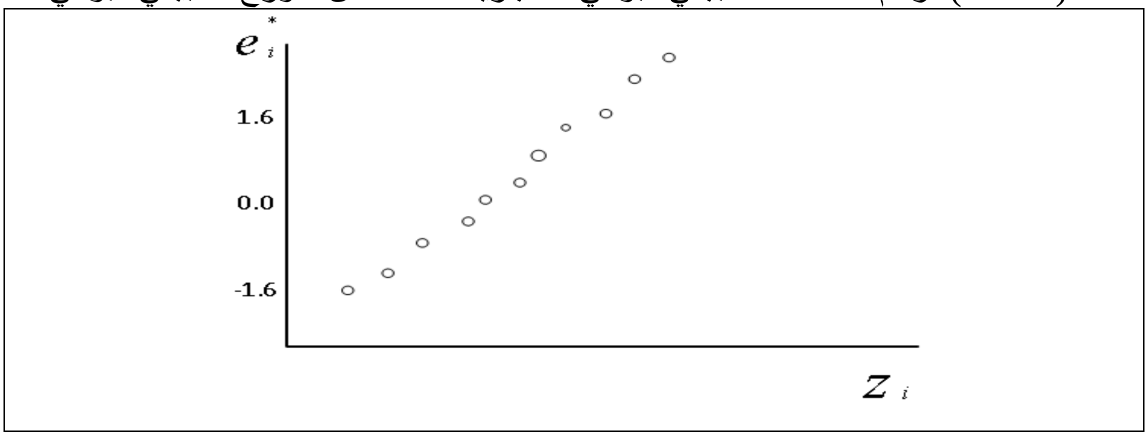
المصدر: تصميم الباحث.

حالة عدم التوزع الطبيعي للبواقي:

اتباع توزيع حدود الخطأ للتوزيع الطبيعي يعتبر أحد فروض نموذج الانحدار الخطي، وتوجد عدة أنواع من الرسومات التي تستخدم للكشف عن التوزيع الاعتدالي لحدود الخطأ، تتمثل في رسم الصندوق الذي يمكن من الكشف عن وجود مشاهدات متطرفة، ورسم المدرج التكراري للبواقي أو البواقي المعيارية الذي يحدد ما إذا كان شكل البواقي يشبه التوزيع الطبيعي، ومقارنة التكرارات الفعلية للبواقي المعيارية مع التكرارات المتوقعة المقابلة لها، ولكن رسم الاحتمال الطبيعي Normal Probability plot يعتبر الأكثر استخداما للكشف عن عدم اتباع توزيع البواقي للتوزيع الطبيعي شكل (2 – 18)، وفيه تتم مقارنة كل حد خطأ مع القيمة المتوقعة له، عندما يكون التوزيع طبيعياً، وإذا كان توزع النقاط قريبا من الخط المستقيم دل ذلك على أن توزيع حدود الخطأ يتبع التوزيع الطبيعي وإذا كان غير ذلك يكون توزيع حدود الخطأ غير طبيعي، وتحسب القيم المتوقعة تحت

الطبيعيــة لحــد الخطــأ وفقــا للصــيغة التقريبيــة $\sqrt{MSE}\left[Z\left(\frac{i-0.375}{n+0.25}\right)\right]$ ، ويفضــل عــادة تــأخير عمليــة الكشــف عــن إنحــراف توزيــع حــدود الخطــأ عــن التوزيــع الطبيعــي، نظــرا لتــأثر توزيعهــا بجملــة مــن الانحرافــات الأخــرى مثــل عــدم مناسـبة دالــة الانحــدار المســتخدمة أو لعــدم ثبــات تباين الخطأ [6][2].

شكل (2 – 18): رسم الاحتمال الطبيعي للبواقي المعيارية للكشف عن التوزع الطبيعي للبواقي.



المصدر: تصميم الباحث.

2 – 7 – 2: الطرق الرسمية لتشخيص النموذج

الطــرق الرســمية لتشــخيص نمــوذج الانحــدار الخطــي تضــم مجموعــة مــن الاختبــارات الاحصــائيةالتي تتطلــب اســتقلالية المشــاهدات مــع أن البــواقي غيــر مســتقلة، إلا أن معضــلــة عــدم الاسـتقلالية يضـعف أثرهـا مـع العينـات الكبيـرة، وفيمـا يلـي نتنـاول بعـض هـذه الاختبـارات وفقـا لتصــنيف يقــوم علــى الخصــائص التــي يــتم اختبارهــا، وهــي اختبــارات ثبــات التبــاين، التــوزع الطبيعي، وجودة التوفيق.

اختباراتثبات التباين:

هناك عدة اختبارات تستخدم للكشف عن مشكلة عدم ثبات التباين مثل اختبار بـارك Park Test واختبـار جليجسر Glejser Test واختبـار جولدفيلد – كواندت Goldfeld-Quandt Test واختبـارBreusch-Pagan-Godfrey Test واختبـار سـبيرمان للارتبـاط الرتبـي Spearman's Rank Correlation Test، وسـنتناول بالشـرح اختبـارPark Test الـذي قدمـه بـارك فـي العـام

(1966م) ضـمن بحـث بعنـوان "Estimation with Heteroscedastic Error Terms" المنشور في مجلة Econometrica وهو محاولـة لتحويل طريقـة الرسم إلـى طريقـة رسمية للكشـف عن مشكلة عدم ثبات التباين، حيث اقترح المعادلـة $\sigma_i^2 = \sigma^2 X_i^\beta e^{v_i}$ للعلاقة بين تبـاين حد الخطـأ σ_i^2 والمتغير المستقل X_i وبوضع طرفـي المعادلـة فـي صيغتهما اللوغاريثمية نحصـل علـى المعادلـة $ln\sigma_i^2 = ln\sigma^2 + \beta lnX_i + V_i$ ويعبر V_i عن حد الخطأ العشوائي، وحيث إن قيمـة σ_i^2 مجهولـة اقترح بارك استخدام e_i^2 كتقريب لهـا، وتنفيذ نمـوذج الانحـدار $lne_i^2 = \alpha + \beta lnX_i + V_i$ حيث $\alpha = ln\sigma^2$ واذا وجد أن βدالة احصائيا فان ذلك يعتبر دليلاً على وجود مشكلة عدم ثبـات التبـاين، والنتيجة المعاكسة تقم دليلاً على ثبات التباين[6] [78] [2].

اختبارات التوزع الطبيعي:

توجـد مجموعـة مـن الاختبـارات للتأكـد مـن الأخطـاء تتـوزع طبيعيـا منهـا اختبـار Jarque – Bera (JB) Test of Normality، واختبـارAnderson – Darling Normality Test، واختبـار مربـع كـاي Chi-Square Test، واختبـار كولموقروف – سميرنوف Kolmogorov – Smirnov Test، وهنـا نسـتعرض بشـيء مـن التفصـيل اختبـار Jarque – Bera Test of Normality فهـو يعتبـر مـن اختبـارات العينـات الكبيـرة وهـو يعتمـد علـى بـواقي طريقـة المربعـات الصـغرى الاعتياديـة OLS، لاسـتخدام هـذا الاختبـار أولا يـتم حسـاب الالتـواء Skewness والـتفلطح Kurtosis للبـواقي، ومـن ثـم يـتم تطبيـق اختبـار(JB) ومعادلتـه $JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$ حيـث K, S, n تمثل التفلطح والالتـواء وحجم العينـة. عنـدما تكـون الاخطـاء موزعـة وفقـاًللتوزيـع الطبيعـي فان $(S = 0, k = 3)$ ويتوقع أن تكون قيمة $(JB = 0)$ عليه إذا كانت قيمة JB قريبة من الصـفر لا نستطيع رفض فرضية توزيع الاخطاء طبيعياً[6] [78] [2].

اختباراتالخطية وجودة التوفيق:

من اختبارات جودة التوفيق الأكثر استخداما اختبار مربع كاي Chi-square Test for Goodness، واختبار كولموقروف – سميرنوف Kolomogorov-Simirnov، تستخدم لتحليل البواقي والتأكد من أن حدود الخطأ تتوزع طبيعيا.

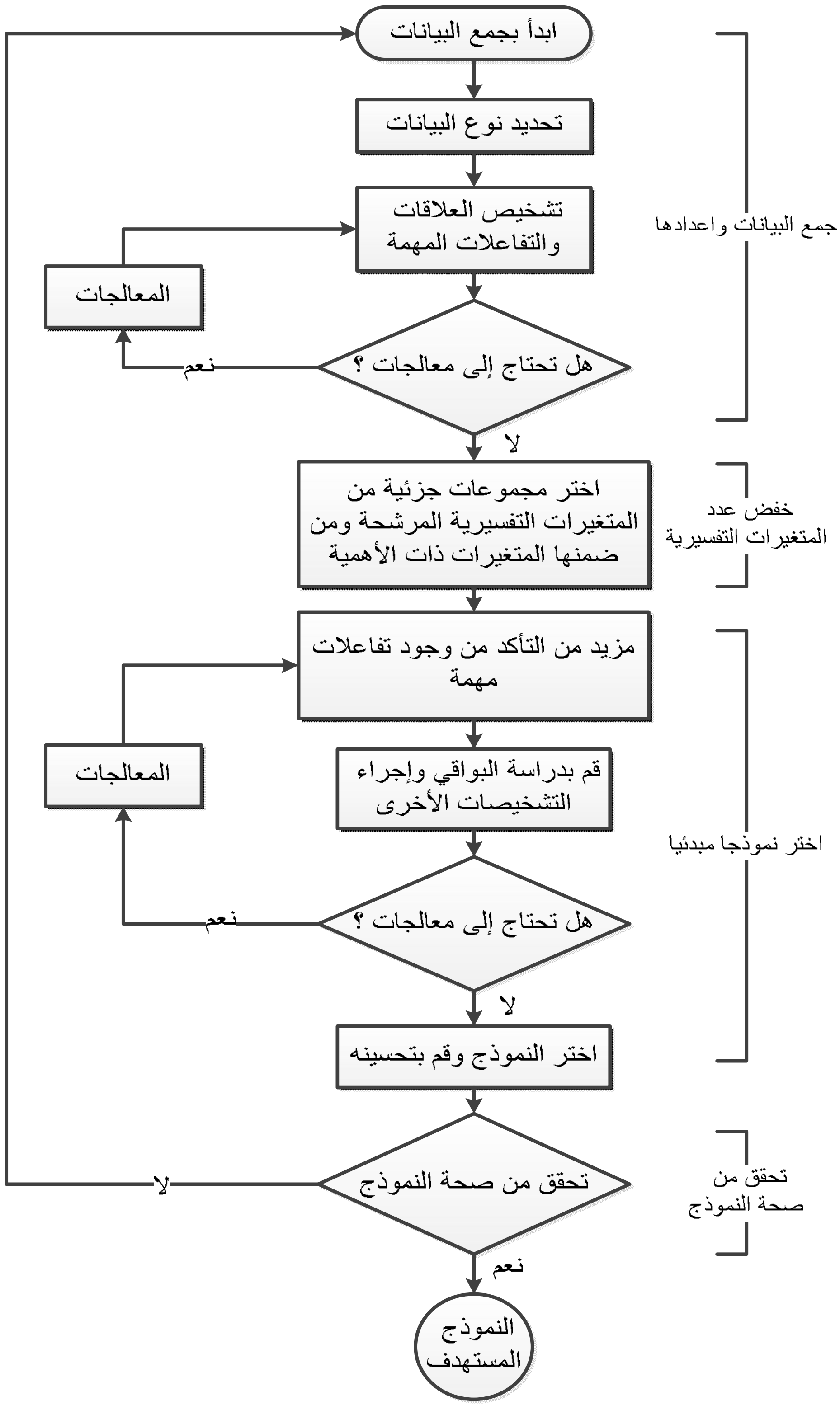
2 – 8: بناء النموذج واختيار المتغيرات

بناء النموذج مرحلة مهمة في تحليل الانحدار الخطي، حيث من خلالها يقوم الباحث بتوصيف النموذج، وتحديد المتغيرات التفسيرية المتوقع أن يكون لها إسهام جيد في تباين متغير الاستجابة بناء على المفهوم النظري والتطبيقي، وأيضا تشتمل عملية بناء النموذج على خوارزميات مختلفة لاختيار المجموعة الجزئية من المتغيرات التفسيرية الأكثر تأثيراً على متغير الاستجابة، وفقا لأحد المعايير الإحصائية المعروفة، والمقارنة بين مجموعة من النماذج للوصول للنموذج الأفضل تمثيلاً للبينات محل الدراسة[6] [30] [78] [2].

2 – 8 – 1: بناء نموذج الانحدار المتعدد

بناء نموذج الانحدار المتعدد يتم وفقا لعدة خطوات تبدأ بجمع واعداد البيانات، أو تسمية المتغيرات التفسيرية وفقا لعلاقتها بمتغير الاستجابة سواء أكان ذلك من الناحية النظرية أو التطبيقية، وهي مرحلة عادة ينتج عنها كم كبير من المتغيرات التفسيرية، لذلك تليها مرحلة خفض هذا العدد الكبير من المتغيرات التفسيرية وفقا لأهميتها في النموذج، وتأتي بعدها مرحلة اختيار المتغيرات التفسيرية الأكثر تأثيرا في تباين متغير الاستجابة، ثم تأتي أخيرا مرحلة التحقق من صحة النموذج، والشكل (2 – 19) يوضح مراحل وخطوات بناء نموذج الانحدار الخطي المتعدد.

شكل (2 – 19): مخطط بناء نموذج الانحدار الخطي المتعدد



المصدر: منقول بتصرف من المرجع [1]

2 – 8 – 2: معايير تقييم النماذج الفرعية للانحدار

كما يتضح من الشكل (2 – 19) الخاص بخطوات بناء نموذج الانحدار هناك عدة مواقف فيها يجب أن يتخذ الباحث قراراً بشأن دقة وصحة النموذج، إلى أن يصل إلى مرحلة اختيار النموذج النهائي وهنا لابد من وجود معيار للتمييز بين مجموعة من نماذج الانحدار المتعدد المرشحة لتمثيل الظاهرة، وهنا تجدر الاشارة إلى أنه عند وجود متغير استجابة واحد لكل نموذج تتم المقارنة بين مجموعة من النماذج الفرعية أو المخفضة Reduced Model، والنموذج الكامل Full Model. ومن المعايير الأكثر استخداماً للتمييز بين النماذج :معامل التحديد R^2، واختبار F الجزئي، ومتوسط مربعات الخطأ MSE، واحصاء ملاوس C_p، وفيما يلي نتناول باختصار هذه المعايير الأربعة [2] [6] [30] [78]:

معيار معامل التحديد R^2:

معامل التحديد يفسر نسبة أكبر تباين في قيم متغير الاستجابة يعود إلى تأثير المتغيرات التفسيرية في النموذج، لذا يمكن استخدامه للمقارنة بين نماذج الانحدار، ولكن يعاب عليه تأثره بعدد المتغيرات التفسيرية المضمنة في النموذج، حيث إضافة أي متغير تفسيري للنموذج تؤدي إلى زيادة قيمة معامل التحديد، بغض النظر عن مستوى تأثير المتغير التفسيري الذي تتم إضافته، وبما أن المقارنة عادة تتم بين النموذج المخفض، والنموذج الكامل، فإن الأفضلية في الغالب تكون للنموذج الكامل، عليه يفضل استخدام معامل التحديد المعدل $\bar{R}^2$ للتمييز بين النماذج بدلا من الاعتيادي R^2.

معيار F الجزئي:

يستخدم اختبار F الجزئي لاختبار معنوية نقص دقة التنبؤ نتيجة لحذف واحد أو مجموعة من المتغيرات التفسيرية من النموذج، يتم. يستخدم هذا المعيار للتأكد من أن الفرق بين مجموع مربعات البواقي للنموذج المخفض، وللنموذج الكامل يختلف اختلافاً معنوياً عن الصفر. ويتم حساب قيمة الاختبار وفقا للصيغة (2.52):

$$F_p = \frac{\left(RSS_p - RSS_k\right)/(k-p)}{RSS_k/(n-k-1)} \quad , F_{(k-p)(n-k-1)} \qquad 2.52$$

ويعبـر الرمـزان k و p علـى التـوالي عـن عـدد المتغيـرات التفسـيرية فـي النمـوذج الكامل، وفـي النمـوذج المخفـض، بينمـا تعبـر n عـن عـدد المشـاهدات، و RSS_p تمثـل مجمـوع مربعات بواقي النموذج المخفض، وتمثل RSS_k مجموع مربعات بواقي النموذج الكامل.

معيار $\mathbf{MSE}_p$:

يسـتخدم معيـار MSE_p للمفاضـلة بـين النمـوذج المخفـض والنمـوذج الكامـل بحيـث إذا كـان MSE_p للنمـوذج المخفـض يقـل عـن أو يسـاوي MSEللنمـوذج الكامـل، حينئـذ فـالنموذج المخفض تكـون لـه قـدرةٌ تفسـيرية مماثلـة، أو قريبـة مـن القـدرة التفسـيرية للنمـوذج الكامـل. عـادة مـن خـلال اسـتخدام هـذا المعيـار يـتم السـعي وراء عـدد بسـيط مـن المجموعـات الجزئيـة التـي يكـون MSE_p لهـا فـي حـده الأدنـى أو قريبـاً جداً مـن القيمـة الـدنيا، بحيث معـه ينعـدم معـه وجـود تأثير يذكر لأي متغير تفسيري آخر تتم إضافته للنموذج.

معيار احصاء ملاوس $Mallows'C_p$:

يهـتم هـذا المعيــار بمتوســط مربعــات الخطــأ الكلــي أو بالتبــاين لكــل مــن النمــوذج المخفــض والنمــوذج الكامــل، تســاوي عــادة قيمــة C_p (p + 1) إذا كــان تبــاين النمــوذج المخفض تسـاوي تبـاين النمـوذج الكامـل، والنمـوذج الجيـد هـو الـذي يسـاوي عـدد معلماتـه (عـدد المتغيـرات + 1). عليـه فـان معيـار مـلاوس يقـدم معلومـات مفيـدة فيمـا يتعلـق بعـدد المتغيـرات التفسيرية التي يجب تضمينها في النموذج. وصيغة الاختبار كما على المعادلة (2.53):

$$C_p = \frac{RSS_p}{MRSS_k} - [n - 2(p+1)] \qquad 2.53$$

حيــث تمثــل RSS_p مجمــوع مربعــات البــواقي للنمــوذج المخفــض، وتمثــل $MRSS_k$ متوسط مجموع مربعات البواقي للنموذج الكامل وهو يساوي التباين.

2 – 8 – 3: طرقاختيارالمجموعات الجزئية

تختلـف مراحـل التصـفية الأولـى للمتغيـرات التفسـيرية الداخلـة فـي نمـوذج الانحـدار الخطـي المتعـدد عـادة بـاختلاف عـدد كبيـر مـن هـذه المتغيـرات، وبـذلك يكـون الباحـث بحاجـة إلـى الوصـول إلـى فئـة محـدودة مـن المتغيـرات التفسـيرية تكـون ممثلـة تمثـيلا جيـداً للنمـوذج، وهنـا تـأتي أهميـة الخوارزميـات مثـل طريقـة جميـع الانحـدارات الممكنـة All-Possible Regression Procedure، و طريقـة الاختيـار إلـى الأمـام خطـوة فخطـوة Forward Selection Procedure، و طريقـة الحـذف إلـى الخلـف Backward Elimination Procedure، و طريقـة الانحـدار التـدريجي Stepwise Regression التـي تمكـن مـن الوصول إلى فئة تعكس طبيعة الظاهرة التي يمثلها النموذج بشكل جيد.

2 – 8 – 4: التحقق من صحة النموذج

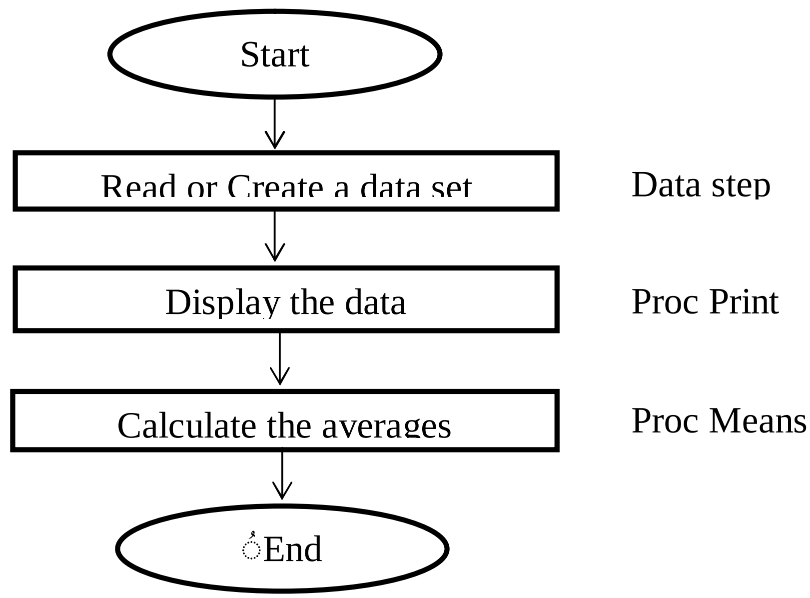
هنـاك ثـلاث طـرق اسـتقر اسـتخدامها بـين البـاحثين والمشـتغلين بالتحليـل الاحصـائيللتأكد مـن صـحة النمـوذج هـي: تتمثـل الطريقـة الأولـى فـي عمليـة فحـص النمـوذج للتأكـد مـن قدرتـه التنبؤيـة مـن خـلال بيانـات جديـدة، ومـن ثـم يـتم مقارنـة خـواص النمـوذج الجديـد المختلفـة بخـواص النمـوذج القـديم. وتقـوم الطريقـة الثانيـة علـى مقارنـة نتـائج النمـوذج بالمعطيـات النظريـة والتجريبيـة السـابقة أو بإسـتخدامنتائج المحكـاة. أمـا الطريقـة الثالثـة فتقـوم علـى الاحتفـاظ بجـذء مـن العينـة،لأغراض التحقـق مـن صـحة النمـوذج مـن خـلال تقيـيم قدرتـه التنبؤيـة. وأهـم مقـاييس قـدرة النمـوذج التنبؤيـة هـو متوسـط مربعـات الخطـأ MSE بحيـث كلمـا صغُرت قيمة MSE كلما دّل ذلك على قدرة النموذج التنبؤية.

2 – 9: نظام التحليل الاحصائي Statistical Analysis System (SAS)

نظام التحليل الاحصائي (ساس) عبارة عن مجموعة من البرامج تم وضـعها قبـل معهد نظـام التحليـل الاحصـائي SAS Institute للتحلـيلات المتقدمـة، وذكـاء الأعمـال، وإدارة البيانـات، والتحليلات التنبؤيـة. يعد معهد SAS ذو الحصـة السوقية الأكبر فـي مجال التحليلات المتقدمـة. تم تطوير برنامج SAS في جامعة ولاية كارولينا الشمالية في الفترة من عام 1966م إلـى عام 1976م، حيث شهدت هذه الفترة تأسيس معهد SAS. أيضا شهد البرنامج تطورا في الفترة ما بين عام 1980م

إلى العام 1990م شمل إضافة إجراءات إحصائية جديدة، ومكونات إضافية، ومقدمة إلى JMP. كما تم إضافة واجهة مستخدم رسومية مع الاصدارة التاسعة في العام 2004م [35][32][11][7] [100][81].

شكل (2 – 20): مخطط لتحليل مبسط للبيانات في برنامج SAS



المصدر: تصميم الباحث

نظام التحليل الاحصائي SAS عبارة عن مجموعة من البرمجيات التي تستكشف، وتعدل، وتدير، وتسترجع البيانات من مجموعات متنوعة من المصادر، ومن ثم إجراء التحليل الاحصائي عليها. يحوى البرنامج واجهة رسومية تمكن المستخدم العادي من إجراء بعض التحليلات الإحصائية، بينما تقدم لغة البرمجة في برنامج SAS خيارات متقدمة في مجال التحليل الاحصائي. عادة يبدأ برنامج SAS بجزء للبيانات DATA Step من خلاله يتم استرجاع ومعالجة البيانات قبل اجراء التحليل الاحصائي عليهاشكل (2 – 20)، فتتكون خطوة البيانات من جزئين لتجميعوتنفيذ الأوامر، فيمرحلةالتجميعتتممعالجةالجمل الاعلانية واستكشافالأخطاءفيبناءالجمل الاعلانية،ومن ثم يتمتنفيذجميع الأوامر بالتتابع. بينما يتكون الجزء الثاني PROC Step من مجموعة من إجراءات تحليل البيانات واعداد التقارير الخاصة بعمليات تحليل البيانات، والأشكال البيانية.

الفصل الثالث

التعدد الخطي

Multicollinearity

الفصل الثالث

التعدد الخطي

3 – 1: تمهيد

يتطـرق هـذا الفصـل إلـى تعريـف مشـكلة التعـدد الخطـي متنـاولا أسـباب حـدوثها وطـرق الكشـف عنهـا وتأثيراتهـا علـى نمـوذج الانحـدار الخطـي المتعـدد واسـتعراض طـرق وأساليب المعالجة، حيث سيتم تناول طريقة انحدار الحافة وطريقة المكونات الرئيسية.

3 – 2: التعدد الخطي Multicollinearity

يعود تاريخ مشكلة التعدد الخطي إلى الورقة التي نشرها Frisch في عام (1934م) حيث وضح فيها مفهوم التعدد الخطي بأن المتغيرات التي تتعامل معها قد تكون واقعة تحت تأثير علاقتين أو أكثر. تحدث مشكلة التعدد الخطي Multicollinearity Problemعندما تكون هناك علاقة خطية بين اثنين أو أكثر من المتغيرات التفسيرية، وبما أن أحد الشروط الواجب توفرها في نموذج الانحدار هو شرط الرتبة Rank Condition حيث يقال أن المصفوفة$\boldsymbol{X}$كاملة الرتبة إذا كان $rank(\boldsymbol{X}) = p$ حيث إن $\boldsymbol{X}$ مصفوفة من مرتبة $(n \times p)$ لمشاهدات المتغيرات التفسيرية، وعندما تكون المتغيرات التفسيرية مستقلة خطياًLinearly independent فانه يمكن إيجاد معكوس المصفوفة $X'X$ الذي يترتب عليه إمكانية إيجاد $(\hat{\beta})$ مقدرات طريقة المربعات الصغرى الاعتيادية (OLS). عليه فان التعدد الخطي يعني وجود علاقة ارتباط خطي بين متغيرين اثنين أو بين جميع المتغيرات التفسيرية ويمكن التعبير عنه رياضيا بما يلي[38][31][27][26]:

إذا كانت المتغيرات التفسيرية $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_p$ ووجد أن p من الثوابت $C_1, C_2, \ldots, C_p$ ليست جميعها أصفاراً، وتحقق المعادلة (1.3) التالية.

$$\sum_{j=1}^{p} C_j \boldsymbol{X}_j = \mathbf{0} \qquad 1.3$$

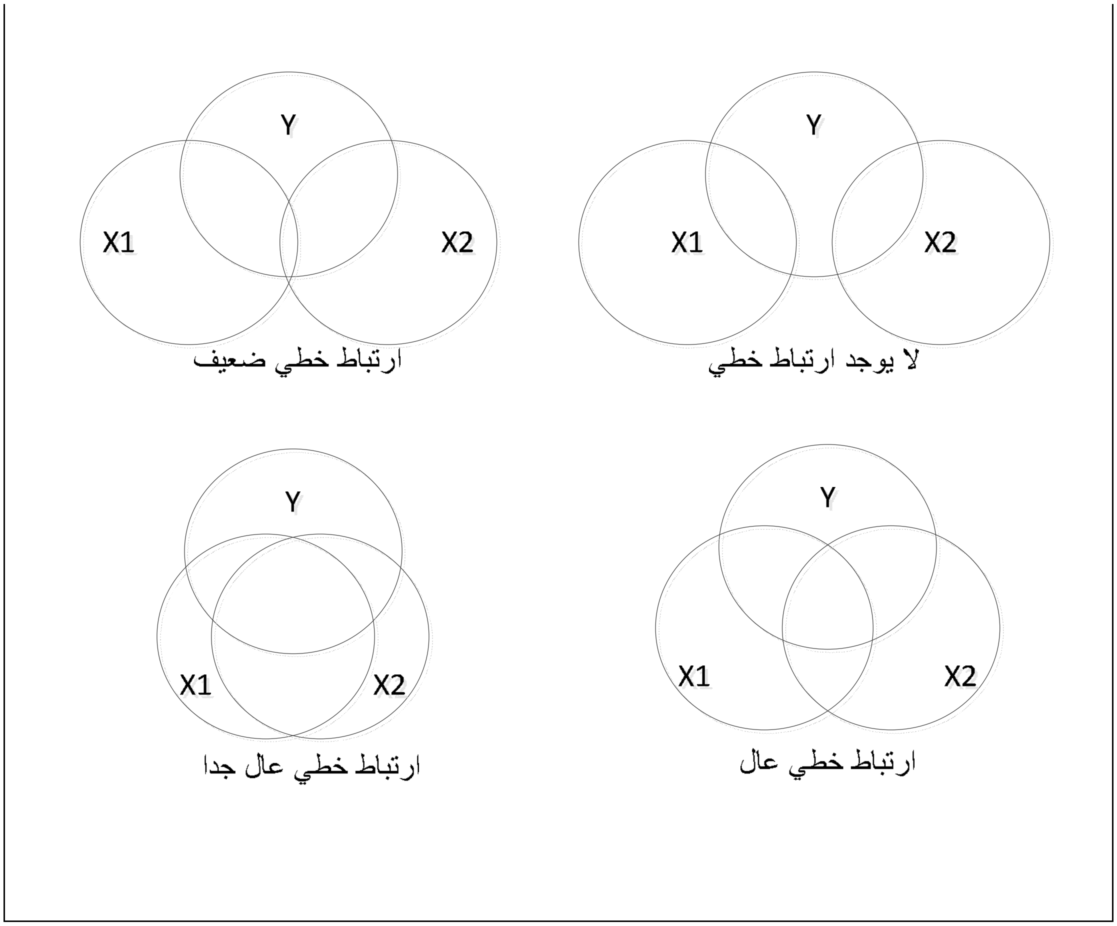
وتحقــق العلاقــة (3.1) يبــين أن المتجهــات غيــر مســتقلة خطيــاً، واذا كــان $C_1 = C_2 = \cdots = C_p$ أيضــا تحــق العلاقــة (1.3). بينمــا إذا كانــت $C_j \neq 0$ فــيمكن الحــل بالنســبة إلى :

$$X_j = \frac{1}{C_1}\left[C_1X_1 + C_2X_2 + \cdots + C_pX_p\right] \qquad 2.3$$

ومــن المعادلــة (2.3) يتضــح أنــه إذا كــان p مــن المتجهــات مرتبطــة خطيــا فهــذا يــدل علــى أن أحــدها عبــارة عــن تركيــب خطــي مــن المتجهــات الأخــرى، ويمكــن الحصــول علــى أي متجه بدلالة المتجهات الأخرى.

للتعدد الخطي أنواع هي تعدد خطـي تـام Exact or Perfect Multicollinearity ويحدث في حالة وجود ارتباط تـام بين متغيرين تفسيريين أو أكثر، والنوع الثـاني التعدد الخطـي غير التـام Non-exact or perfect Multicollinearity وفيه تظهر مشكلة التعدد الخطـي إذا كـان الارتبـاط عال أو قريبا من الارتباط التام سواء أكان بين متغيرين تفسيريين أو بين جميع المتغيرات التفسيرية أنظــر الشــكل (1.3)، وفيمــا يلــي استعراضــا تفصــيليا لأنــواع التعــدد الخطــي بــين المتغيــرات التفسيرية[90] [6] [43][2]:

شكل (3 – 1): مستويات الارتباط الخطي غير التام بين المتغيرات التفسيرية.



المصدر: تصميم الباحث.

3 – 2 – 1: التعدد الخطي التام Exact or Perfect Multicollinearity

تكون العلاقة الخطية بين عمودين أو أكثر من أعمدة مصفوفة المتغيرات التفسيرية $\boldsymbol{X}$ علاقة تامة إذا كانت ثوابت العلاقة (2.3) المتمثلة في $C_1 = C_2 = \cdots = C_p$ على الأقل أحدها لا يساوي صفراً، وعند هذه الحالة لا يمكن تقدير معلمات النموذج بطريقة المربعات الصغرى حيث إنها تتطلب ايجاد معكوس المصفوفة $X'X$، ولكن لسوء الحظ فإن محدد المصفوفة $X'X$ يساوي صفرا $|X'X| = 0$ عندما ينتهك شرط الرتبة بحيث تكون المصفوفة غير كاملة الرتبة أي $Rank(\boldsymbol{X}) < p$ أقل من عدد المتغيرات المستقلة [23] [26][18].

3 – 2 – 2: التعدد الخطي غير التام Non Exact or perfect Multicollinearity

تحدث حالة التعدد الخطي غير التام عندما تكون بعض المتغيرات التفسيرية دالة في تركيبة المتغيرات الأخرى أنظر شكل (3 – 1) مع وجود قيم عشوائية (δ_i) إذا تحققت العلاقة التي تمثلها المعادلة (3.3).

$$\alpha_1 x_1 + \alpha_2 x_2 + \cdots + \alpha_p x_p + \delta_i = 0 \qquad 3.3$$

$$i = 1,2, \ldots , p$$

حيث إن $\alpha_1, \alpha_2, \ldots, \alpha_p$ ثوابت على الأقل أحدها لا يساوي صفرا، كذلك ($\delta_i \neq 0$) وفي هذه الحالة لايساوي محدد المصفوفة $X'X$ صفرا ولكنه يكون قريبا من الصفر $|X'X| \cong 0$، مما يؤدي إلى عدم دقة تقدير المعلمات بطريقة المربعات الصغرى الاعتيادية نتيجة إلى كبر تباين المعلمات المقدرة[23] [26][18].

3 – 2 – 3: أهمية مشكلة التعدد الخطي

قد لا تشكل مشكلة التعدد الخطي حالة مقلقة إذا كان الهدف من بناء النموذج هو التنبؤ بقيم المتغير التابع Y_i بناء على قيم المتغيرات التفسيريةX_j لأن القيم التنبؤية تظل تتمتع بقدر عال من الدقة، وكذلك قيم معامل التحديد أو معامل التحديد المعدل تقيسان بشكل جيد إلى أي مدى يتنبأ النموذج بقيم Y_i المتغيرالتابع.

أما إذا كان الهدف من تصميم نموذج الانحدار الخطي المتعدد هو ايجاد تقديرات لمعلمات نموذج الانحدار الخطي المتعدد، أو معرفة الأهمية النسبية لمساهمة أي من المتغيرات التفسيرية في تباين المتغير التابع، فإن التعدد الخطي يُعد مشكلة خطيرة تواجه نموذج الانحدار الخطي. حيث يؤدي إلى عدم استقرار معلمات مقدرات نموذج الانحدار الخطي، كما يؤدي إلى اتساع خطأ المعاينة لمقدرات طريقة المربعات الصغرى الاعتيادية $cov(b) = \sigma^2 (X'X)^{-1}$[18][26] [23][78].

3 – 2 – 4: أسباب مشكلة التعدد الخطي

تنشأ مشكلة التعدد الخطي لعدة أسباب منها ما يتسبب فيه الباحث ومنها ما يعود إلى طبيعة البيانات المستخدمة، كما أن منها ما يتسبب في التعدد الخطي التام وبعضها الآخر تبرز معه مشكلة التعدد الخطي غير التام، ومن المهم فهم الفروق بين المصادر التي تؤدي إلى الخطية المتعددة لأن تحليل البيانات وتفسير النموذج الناتج عنها إلى حد ما يعتمد على سبب المشكلة، وفيما يلي استعراضا لهذه الأسباب:

- إدراج نفس المتغير المستقل في النموذج مرتين بمستويي مقياس مختلفين كإدراج درجات الحرارة المئوية مرة ومرة أخرى بالسيلزيوس.
- الادراج الخاطئ للمتغيرات الصورية بحيث يتم إدراج متغيرات صورية بعدد مستويات المتغير الاسمي k في النموذج بدلا من إدراج (k-1) متغير صوري.
- الأسلوب المتبع في جمع البيانات المستخدمة يمكن أن يؤدي إلى بروز مشاكل الخطية المتعددة، مثلا عندما يقوم الباحث أو المحلل باختيار عينات جزئية من فضاء المعاينة الخاص بالمتغيرات التفسيرية المعرفة على المعادلة (1.3).
- القيود المفروضة على النموذج أو على المجتمع الذي منه يتم أخذ العينات يمكن أن يتسبب في الخطية المتعددة، كميل بعض المتغيرات إلى التغير سويا مما يؤدي إلى أن يكون بينها ارتباطاً أقوى.
- اختيار النموذج يمكن أن يتسبب في بروز مشكلة الخطية المتعددة، فإضافة متغير كثيرات الحدود يؤدي إلى أن تكون المصفوفة $\boldsymbol{X'X}$ مصفوفة عليلة، أيضا إضافة المتغير X^2 في ظل وجود المتغير Xيؤدي مستوى حاد للتعدد الخطي.
- إضافة عدد كبير من المتغيرات التفسيرية إلى النموذج بحيث يفوق عددها، عدد المشاهدات التي تمثلها.
- ادراج متغيرات تفسيرية متباطئة Lagged Variables كما يحصل في النماذج التي يراد لها أن تعالج بيانات تتعلق بأسعار محصول ما في فترات زمنية محددة والأسعار في الفترات السابقة كمتغيرات مفسرة لكمية انتاج المحصول.

يلاحظ على الأسباب السابقة المتعلقة ببروز مشكلة التعدد الخطي في نموذج الانحدار الخطي المتعدد، أن السببين الأولين يؤديان عادة إلى بروز مشكلة التعدد الخطي التام، بينما الأسباب الأخرى يمكن أن تؤدي إلى ظهور التعدد الخطي غير التام الذي قد تتفاوت مستويات حدته ودرجة تأثيره على النموذج[79][36][23] [26][22][18][5].

3 – 2 – 5: تأثيرات التعدد الخطي

إذا كان نموذج الانحدار الخطي المتعدد واقعا تحت وطأة مشكلة التعدد الخطي التام فإن محدد المصفوفة $X'X$ يكون مساوياً للصفر فتكون المصفوفة وحيدة Singular ويتعذر ايجاد معكوس المصفوفة، ويترتب على ذلك استحالة ايجاد معلمات النموذج والخطأ المعياري لها، عليه يستحيل استخدام طريقة المربعات الصغرى الاعتيادية لتقدير معلمات النموذج، ولكن لحسن الحظ يندر بروز مشكلة التعدد الخطي التام في الواقع العملي.

أما إذا وقع النموذج تحت تأثير التعدد الخطي غير التام بحيث يكون محدد المصفوفة $\boldsymbol{X'X}$ قريبا من الصفر $|\boldsymbol{X'X}| \cong 0$ يترتب عليه عدة تأثيرات يمكن إجمالها فيما يلي [5] [66] [89][23] [26][22][18][13][6][2]:

- يتوقع أن يكون لمقدرات طريقة المربعات الصغرى الاعتياديةOLS أخطاء معيارية كبيرة بما يؤدي إلى استحالة الحصول على تقديرات دقيقة، وذلك لأن اقتراب محدد المصفوفة $X'X$ من الصفر يؤدي إلى أن تكون قيم العناصر القطرية لمعكوس المصفوفة $X'X$ كبيرة، ويترتب على ذلك أيضا كبر حجم الأخطاء المعيارية $\mathrm{var}(b) = S^2(\mathbf{X'X})^{-1}$.
- نتيجة للكبر النسبي لأحجام الأخطاء المعيارية كما ورد في النقطة السابقة تكون فترات الثقة واسعة بما يؤدي إلى سهولة قبول فرض العدم Null Hypothesis، علما أن حجم الخطأ المعياري إلى حد ما يتأثر بمستوى العلاقة الارتباطية بين المتغيرات التفسيرية المضمنة في النموذج.
- في ظل تأثر النموذج بمشكلة التعدد الخطي قد تظهر بعض أو جميع قيم احصاءةالاختبارT-test لمعلمات النموذج صغيرة نسبيا نسبة لكبر حجم الخطأ المعياري لها، وذلك يؤدي إلى عدم معنوية هذه المقدرات على الرغم من أن قيمة معامل التحديد R^2 للنموذج تكون كبيرة نسبياً.

- تكون مقدرات طريقة المربعات الصغرى الاعتيادية OLS وأخطائها المعيارية حساسة لأي تغيرات مهما كانت صغيرة في المشاهدات وبذلك لا تكون النتائج قوية Robust.
- ينتج عنها تغيرات غير متوقعة في حجم أو قيم المعلمات، وكذلك قد تظهر إشارات المعلمات غير متفقة مع الواقع التطبيقي أو النظري للبيانات، على الرغم من معنوية النموذج ككل.

3 – 3: تشخيص التعدد الخطي Multicollinearity Diagnosis

مشكلة التعدد الخطي تتعلق بمستوى شدة تأثيره على النموذج وليس بوجوده أوعدمه، ذلك أن التعدد الخطي يعود إلى وجود علاقة ارتباطية بين اثنين أو جميع المتغيرات التفسيرية وهو الأمر الذي لا يكاد يخلو منه نموذج، وإنما كلما كانت العلاقة الارتباطية عالية كلما برزت مشكلة التعدد الخطي واشتد تأثيرها على النموذج، أي كلما كانت درجة التعدد الخطي كبيرة كلما أشتد تأثيرها على النموذج. هنالك عدة تقنيات مقترحة لاكتشاف وتشخيص مشكلة التعدد الخطي، يمكن تقسيمها إلى طرق رسمية تم استعراضها فيما يلي في البند الأول والبند الثاني، وطرق غير رسمية تم استعراضها ضمن البند الثالث:

3 – 3 – 1: عامل تضخم التباين Variance Inflation Factor (VIF)

قدمه في عام (1967) كل من Glaube و Ferrar وأطلق عليه Minquardt مسمى عامل تضخم التباين VIF في عام (1970). يعتبر VIF من الطرق الأساسية الواسعة الاستخدام للكشف عن مشكلة التعدد الخطي، وهو يقيس مدى تضخم تباينات معلمات الانحدار المقدرة في ظل وجود ارتباط خطي بين المتغيرات التفسيرية، وتعتبر العناصر القطرية لمعكوس مصفوفة معلومات النظام $X'X$ مفيدة في الكشف عن التعدد الخطي، ويمكن كتابة العناصر القطرية j^{th} للمصفوفة $\boldsymbol{C}$ وفقا لـ $C_{jj} = \left(1 - R_j^2\right)^{-1}$.

ويلاحظ أنه إذا كان X_j قريباً من التعامد مع المتغيرات التفسيرية الأخرى $p-1$ ، فإن معامل التحديد R_j^2 يكون صغيراً وتكون المصفوفة C_{jj} قريبة من الوحدة، بينما إذا كان X_j معتمدا على بعض المتغيرات التفسيرية، فان R_j^2 يكون قريباً من الواحد وتكون C_{jj} تكون كبيرة. وحيث إن تباين المعلمة j هو $\boldsymbol{C_{jj}\sigma^2}$ يمكننا النظر إلى C_{jj} على أنها عامل يؤدي إلى زيادة تباين $\widehat{\boldsymbol{\beta}}_j$ نتيجة

لقربها من الاعتماد خطيا على بعض المتغيرات التفسيرية المضمنة في النموذج، والمعادلة (4.3) تقيس ما يسمى بعامل تضخم التباين[99][97][71][65].

$$VIF = C_{jj} = 1/\left(1 - R_j^2\right) \qquad 4.3$$

حيث:

VIF_j: عبارة عن عامل تضخم التباين للمتغير التفسيري j.

R_j^2: معامل تحديد نموذج انحدار المتغير التفسيري j على المتغيرات التفسيرية المتبقية (p-1).

تفسير قيمة عامل تضخم التباين VIF

عامل تضخم التباين VIF يقيس أثر التعدد الخطي بين المتغيرات التفسيرية في نموذج الانحدار، وهو دائما أكبر من أو يساوي واحد، ورسمياً ليس هناك قيمة حدية لـ VIF لتحديد شدة تأثير التعدد الخطي الذي يتعرض له النموذج . غالباً ما تستخدم أكبر قيمة لـ VIF كمؤشر للتعدد الخطي غير المرغوب فيه وغالباً إذا تجاوزت قيمة VIF الـ 10تعتبر إشارة إلى إمكانية تأثير غير مقبول للتعدد الخطي المرتفع على مقدرات المربعات الصغرى الاعتيادية.

إذا كان هناك ارتباط تام بين المتغير المستقل X_j والمتغيرات المستقلة الأخرى بحيث $R_j^2 = 1$ فان عامل تضخم التباين يؤول إلى ما لا نهاية، لأن $VIF_j = 1/(1-1) = \infty$، وإذا كان المتغير X_j متعامد مع المتغيرات المستقلة الأخرى بحيث $R_j^2 = 0$ فان قيمة عامل التضخم تساوي واحد صحيح.

ويقدم متوسط قيم VIF معلومات عن خطورة التعدد الخطي بناءًا على التباعد بين معلمات الانحدار المعيارية المقدرة b'_k والقيم الحقيقية لها β'_k والذي يحسب بالصيغة [101] [103] (5.3).

$$\overline{VIF} = \frac{\sum_{j=1}^{p-1}(VIF)_k}{p-1} \qquad 5.3$$

القيم الكبيرة لـ VIF تنتج في المتوسط فروقا كبيرة من معلمات الانحدار المعيارية المقدرة ومعلمات الانحدار المعيارية الحقيقية. ومتوسط قيم VIF إذا كان أكبر بكثير من الواحد فهو مؤشر لتأثير كبير للتعدد الخطي على النموذج.

مقياس التحمل Tolerance

كذلك يتم استخدام معكوس عامل تضخم التباين في بعض حزم البرامج الاحصائية الحاسوبية ويعرف بمقياس التحمل Tolerance للكشف عن وجود التعدد الخطي بين المتغير X_j والمتغيرات المستقلة الأخرى، وهو من تعريفة يحسب بالصيغة (6.3).

$$Tolerance = \frac{1}{VIF_j} = 1 - R_j^2 \qquad 6.3$$

وقيم التحمل الحدية الأدنى المستخدمة في هذه البرامج لدخول المتغير المستقل النموذج هي (0.01 ، 0.001 ، 0.0001)[101] [103].

أثر التعدد الخطي على دقة تقدير (b_j) بدلالة VIF

يحسب تباين مقدر المربعات الصغرى للمعلمة b_j وفقا للصيغة التالية (7.3) ، ويلاحظ أن قيمة تباين المعلمة b_j يزداد مع ازدياد قيمة معامل التحديد للمتغير j.

$$Var(b_j) = \frac{S^2 VIF_j}{(n-1)S_j^2} = \frac{1}{1-R_j^2} \times \frac{S^2}{(n-1)S_j^2} \qquad 7.3$$

حيث:

S^2 = مقدر تباين نموذج انحدار Y على جميع المتغيرات المستقلة.

S_j^2 = تباين المتغير المستقل X_j حيث : $S_j^2 = \frac{\sum_{j=1}^{n}(X_{ji}-\bar{X}_j)^2}{n-1}$

VIF_j= عامل تضخم تباين المتغير المستقل X_j.

وكذلك عامل تضخم التباين للمتغير رقم j هو عبارة عن العنصر القطري رقم j لمعكوس مصفوفة معاملات الارتباط البسيط بين المتغيرات المستقلة حيث[101] [103]:

$$VIF_j = diag(R_{XX}^{-1})jj \qquad 8.3$$

3 – 3 – 2: القيم المميزة/الكامنة Eigenvalues

تعتبر القيم المميزة لمصفوفة معاملات الارتباط بين معلمات نموذج الانحدار مؤشرات جيدة لقياس التعدد الخطي، ولإيجاد القيم المميزة يتم تنفيذ الخوارزمية التالية:

- ايجاد المصفوفة Z حيث:

$$Z = (\boldsymbol{X'X}) \qquad 9.3$$

- ايجاد المصفوفة القطرية S حيث:

$$S = diag(\boldsymbol{X'X})^{-\frac{1}{2}} \qquad 10.3$$

- ايجاد المصفوفة SZS التي تكون عناصرها القطرية عبارة عن واحد صحيح.

$$\boldsymbol{SZS = S(X'X)S} \qquad 11.3$$

- يتم حساب القيم المميزة للمصفوفة SZS، وإذا كانت القيمة المميزة قريبة من الصفر دل ذلك على وجود تعدد خطي عال، أما إذا كانت قيمة أحد القيم المميزة تساوي الصفر دل ذلك على وجود تعدد خطي تام Perfect Multicollinearity[2][6][103] [101].

مؤشر الحالة Condition Index (CI)

مؤشر الحالة Condition Index (CI) يستخدم للكشف عن وجود التعدد الخطي، ويتم حسابه بواسطة القيم المميزة وفقا للصيغة (12.3):

$$CI_j = \sqrt{\frac{\lambda_{max}}{\lambda_j}} \quad , \quad j = 1,2, \dots , p \qquad 12.3$$

حيث:

λ_{max} = أكبر قيمة مميزة.

λ_j = القيمة مميزة رقم j.

رقم الحالة (Condition Number (CN

رقم الحالة (CN) Condition Number أيضا يستخدم للكشف عن وجود التعدد الخطيويحسب باستخدام القيم المميزة وفقا للصيغة (13.3):

$$CN = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \qquad 13.3$$

اقترح Johnston أن قيمة CN إذا كانت ما بين 20 إلى 30 فإنها تكون مؤشراً لوجود تعدد خطي مرتفع، بينما اقترح Belsley وآخرون إذا كانت قيمة CN ما بين 30 إلى 100 يعتبر ذلك مؤشرا إلى تعدد خطي مرتفعا جدا[101] [103][6][2].

3 – 3 – 3: تشخيصات غير رسمية للتعدد الخطي

فيما يلي استعراضا للتشخيصات غير الرسمية التي تقدم مؤشرات لاحتمال تعرض النموذج لتأثيرات مشكلة التعدد الخطي.

- اختبار مصفوفة معاملات الارتباط Examination of Correlation Matrix من أبسط طرق قياس التعدد الخطي الكشف عن العناصر غير القطرية r_{ij} للمصفوفة $X'X$ ، فاذا كانت المتغيرات التفسيرية X_1 و X_2 لها ارتباط عال ببعضها البعض فان القيمة المطلقة لـ

$|r_{12}|$ تكون قريبة من الوحدة وبذلك يتوقع تعرض النموذج لدرجة عالية من التعدد الخطـي. لكن لسوء الحظ إن هذه الطريقة لا تمكن من الوصول إلـى رأي قاطع بوقوع أو عدم وقوع النموذج تحت تأثير مشكلة التعدد الخطـي، لأن ضعف الارتباط بين المتغيرات التفسيرية لا يعني عدم بروز مشكلة التعدد الخطي في كل الأحوال[101] [103][87][6][2].

- ظهور تغيرات كبيرة في حجم وإشارات معلمات النموذج، في حال حدوث تغييرات طفيفة في حجم العينة كإضافة أو حذف مشاهدة أو مشاهدات قليلة.
- عندما يكون النموذج ككل معنوي وقيمة معامل التحديد للنموذج R^2 كبيرة، وتظهر معلمات الانحدار غير معنوية و/أو ذات إشارات غير مطابقة للنظرية أو الواقع التطبيقي.
- عند بناء عدد p نموذج انحدار، كل منها لانحدار أحد المتغيرات المستقلة X_j على المتغيرات المستقلة الأخرى (p-1)، فاذا وجد أن أحد معاملات التحديد $\left(R_j^2\right)$ يقترب من الواحد فان ذلك مؤشر لتأثر النموذج بالتعدد الخطي.
- ظهور فترات ثقة عريضة لمعلمات انحدار تمثل متغيرات مستقلة لها أهمية كبيرة في التجربة أو الظاهرة التي يمثلها نموذج الانحدار.

3 – 4: طرق معالجة مشكلة التعدد الخطي

توجد عدة تدابير مقترحة لتخطي مشكلة التعدد الخطي يمكن تصنيفها إلى طرق رسمية وأخرى غير رسمية، وفيما يلي يتم استعراض الطرق غير الرسمية المستخدمة مع طريقة المربعات الصغرى الاعتيادية، والطرق الرسمية الأخرى المقترحة لمعالجة مشكلة التعدد الخطي.

3 – 4 – 1: الطرق غير الرسمية لمعالجة التعدد الخطي Informal Remedies Methods for Multicollinearity

الطرق غير الرسمية لتخطي مشكلة التعدد الخطي عادة تستخدم مع طريقة المربعات الصغرى الاعتيادية، فهي عبارة عن معالجات تنصب نحو معالجة بعض الاختلالات في البيانات، دون التعديل في بنية معادلة طريقة المربعات الصغرى ومنها: عملية إعادة توصيف النموذج Model Respecification، وحذف أحد المتغيرات التفسيرية المرتبطة، واستخدام بيانات إضافية أو جديدة [101] [103][6] [70][2]Use Additional or New Data.

3 – 4 – 2: بعض الطرق الرسمية لمعالجة التعدد الخطي Some formal Remedies Methods for Multicollinearity

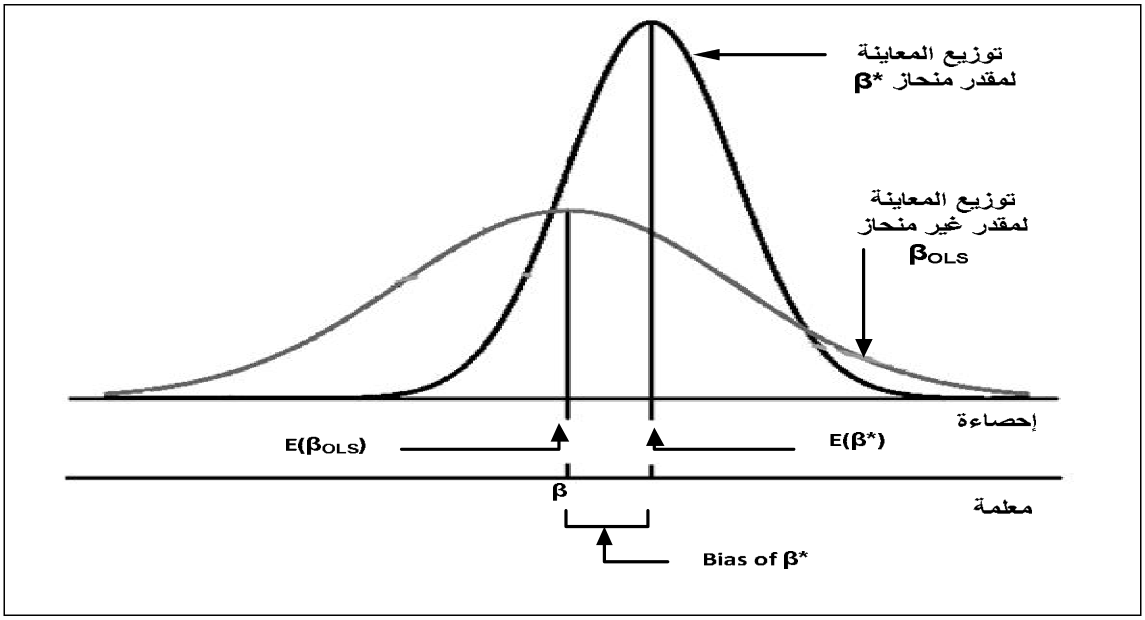
شهد النصف الأخير من القرن العشرين عدة مقترحات لمعالجة مشكلة التعدد الخطي تمثلت أهم مظاهرها في إدخال بعض التعديلات على طريقة المربعات الصغرى الاعتيادية OLS تراوحت بين التعديلات البسيطة والتعديلات الكبيرة، معظمها لها مقدرات متحيزة ولكنها تميزت بقدر كبير من الاستقرار جعلها محل الاهتمام، وعرفت هذه الطرق بطرق الانحدار المتحيزة، منها طريقة انحدار الجذور الصماء Method of Latent Roots Regression (LAT)، ومقدرات بيز Bayes Estimator (BYS) ، والمقدرات المنكمشة Shrunken Estimators (J&S)، والمقدرات بالتكرار Iteration Estimator (ITR)،وفيما يلي استعراض طريقتي انحدار الحافة وانحدار المكونات الرئيسية[40] [96] [104] [39][28][24][17].

3 – 5: انحدار الحافة Ridge Regression

تم تقديم انحدار الحافة أولا في عام (1962) من قبل Arthur E. Hoerl في مقال نشر له في مجلة الهندسة الكيميائية تحت عنوان: تطبيق تحليل الحافة على مشكلات الانحدار، حيث تطرق فيه إلى خبرته في تحليل الانحدار التي ركز فيها على أنه عندما يكون هناك ارتباط بين المتغيرات التفسيرية يكبح قدرة طريقة المربعات الصغرى الاعتيادية OLS على إعطاء مقدرات جيدة مما حدا به إلى اقتراح طريقة تقدم مقدرات أفضل عند وجود ارتباط بين المتغيرات التفسيرية. وفي عام (1970) قدم Arthur E. Hoerl و Robert W. Kennard مقالتين في مجلة Technometrics قدما فيهما انحدار الحافة Ridge Regression كحل متكامل لمشكلة التعدد الخطي [45] [78][58][56][55][54][34].

وعندما يتم الحصول على مقدر متحيزة بمقدار بسيط ويكون في نفس الوقت أكثر دقة بكثير من المقّدرات غير المنحازة فانه يعتبر الأفضل، لأنه حينئذ سيكون المقّدر الأكثر قربا للقيمة الحقيقية. الشكل (3.2) يوضح أن المقدر $\hat{\beta}$ لا يتميز بالقدر المطلوب من الدقة على الرغم من أنه غير متحيز، بينما يلاحظ أن المقدر $\hat{\beta}^*$ يمتاز بقدر عال من الدقة مع أنه متغير منحاز، عليه فان أحتمال وقوع المقدر $\hat{\beta}^*$ قرب القيمة الحقيقة β أكبر من من أحتمال وقوع المقدر $\hat{\beta}$ قرب القيمة الحقيقة.

شكل(3 – 2): مقارنة مقدر منحاز بتباين أصغر بآخر غير منحاز بتباين أكبر.



المصدر: بتصرف من الباحثمرجع [64].

لاستعراضاشتقاق معادلة مقدرات انحدار الحافة نستعرض فكرة مجموع مربعات الخطأ $MSE(b) = E\|b-\beta\|^2$لمقدرات المربعات الصغرى. حيث يعتبر مجموع مربعات الخطأ من المقاييس واسعة الانتشار التي تستخدم لتقييم جودة التقدير. يمكن تجزئة مجموع مربعات الخطأ إلى مربع التحيز زائدا التباين كما في المعادلة [78][64][56][55][54](15.3).

$$\begin{aligned} E\|b-\beta\|^2 &= \sum_j E\left(b_j-\beta_j\right)^2 \\ &= \sum_j \left\{E\left(b_j\right)-\beta_j\right\}^2 + \sum_j Var\left(b_j\right) \end{aligned} \qquad 15.3$$

ووفقا لنظرية Gauss-Markovفإن المربعات الصغرى لها أصغر تباين مقارنة بجميع المُقدرات الأخرى غير المتحيزة. ولكن هذا لا يعني بالضرورة أن يكون لها أقل مجموع مربعات خطأ MSE. وللتفريق بين مجموعة من المُقدرات سنرمز لمقدر المربعات الصغرى الاعتيادية بـ $\hat{\beta}_{LS}$ عليه فإن $E\left(\hat{\beta}_{LS}\right) = \beta$ ومصفوفة التباين $Var\left(\hat{\beta}_{LS}\right) = \sigma^2(X'X)^{-1}$. عليه يمكن كتابة مربعات الخطأ كما يلي:

$$MSE(\hat{\beta}_{LS}) = E\|\hat{\beta}_{LS}\|^2 - \|\beta\|^2$$

$$= tr\{\sigma^2(X'X)^{-1}\} = \sigma^2 tr\{(X'X)^{-1}\} \qquad 16.3$$

إذن

$$E\left(\|\hat{\beta}_{LS}\|^2\right) = \|\beta\|^2 + \sigma^2 tr\{(X'X)^{-1}\} \qquad 17.3$$

ويلاحظ هنا بوضوح أنه إذا كانت المصفوفة $X'X$ عليلة فإن مُقدر المربعات الصغرى الاعتيادية $\hat{\beta}_{LS}$ يكون ذو بُعد كبير $\|\hat{\beta}_{LS}\|^2$ يترافق معه إنفلات في التباين وعدم استقرار المقدرات. وهنا يأتي انحدار الحافة لتخطي مشكلة التعدد الخطي بإيجاد مُقدرات متحيزة ولكنها تمتاز بأن لها تباين أقل وأكثر استقرارا[98][78][56][55][54].

3 – 5 – 1: مُقدرات الحافة Ridge Estimators

للتحكم على انفلات التباين وعدم الاستقرار الذي يرافق مقدرات المربعات الصغرى في حال تعرض النموذج لمشكلة التعدد الخطي اقترحArthur E. Hoerl المقدر المنحاز "مُقدر الحافة" أو Ridge Estimator . لأي مُقدر b يمكن إعتبار معيار المربعات الدنيا هو أصغر ما يمكن الوصول إليه عند $\hat{\beta}_{LS}$ زائدا الصيغة التربيعية عند b :

$$Q(b) = \|Y - X\hat{\beta}_{LS} + X\hat{\beta}_{LS} - Xb\|^2$$

$$= (Y - X\hat{\beta}_{LS})'^{(Y-\hat{\beta}_{LS})} + (b - \hat{\beta}_{LS})'X'X(b - \hat{\beta}_{LS})$$

$$= Q_{min} + \varphi(b) \qquad 18.3$$

عليه فإن محيط المنحرف Contour لأي ثابت صيغة تربيعية لـ $\varphi(b)$ فهو مجسم ناقص Hyper-ellipsoids يتمركز حول أدنى مربع خطأ معياري اعتيادي LSE $\hat{\beta}_{LS}$ ، وبقراءة المعادلة (17.3) فإنه من المعقول قبول ذلك، عليه إذا تحركنا بعيداً عن Q_{min} فإننا سنكون متحركين في اتجاه يؤدي إلى تناقص بُعد b.

ولحل المشكلة باستخدام انحدار الحافة فيمكن تصغير أو خفض $\|\beta\|^2$ وفقا إلى $\left(\beta - \hat{\beta}_{LS}\right)' X' X \left(\beta - \hat{\beta}_{LS}\right) = \varphi_0$ لبعض الثوابت φ_0. وفي هذه الحالة فإن القيود المفروضة تضمن بواقي صغيرة نسبيا لمجموع مربعات $Q(\beta)$ عند مقارنتها إلى أدناها Q_{min}. الشكل (3 – 3) يصور محيط المنحرف لمجموع مربعات البواقي مع قيود الحافة L_2 في بُعدين. إذا نظرنا إليها كمسألة لاجرانج Lagrangianفهي موافقة لتصغير:

$$Q^*(\beta) = \|\beta\|^2 + (1/k)\left\{\left(\beta - \hat{\beta}_{LS}\right)' X' X \left(\beta - \hat{\beta}_{LS}\right) - \varphi_0\right\} \qquad 19.3$$

حيث $1/k$ تمثل المضروب يتم اختياره لاستيفاء القيود. عليه فإن المشتقة التالية:

$$\frac{\partial Q^*}{\partial \beta} = 2\beta + (1/k)\{2(X'X)\beta - 2(X'X)\hat{\beta}_{LS}\} = 0 \qquad 20.3$$

تعطي مُقدرات الحافة كما على المعادلة (21.3) التالية:

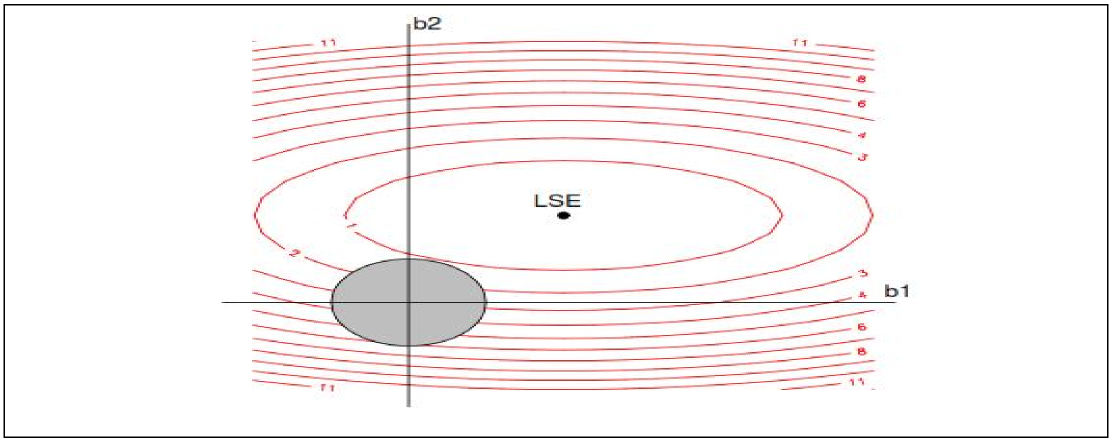
$$\hat{\beta}_{RR} = \{X'X + kI\}^{-1} X'Y \qquad 21.3$$

حيث:

I : مصفوفة الوحدة Identity Matrix ذات بُعد $(p \times p)$.

وعندما تكون $k = 0$ فإن تقديرات انحدار الحافة RR تساوي تقديرات طريقة المربعات الصغرى الاعتيادية، وعندما تكون $k > 0$ فإن مقدرات انحدار الحافة تميل إلى الاستقرار عند قيمة معينة نسبة للتغيرات في البيانات ولكنها تكون متحيزة، ويكون مجموع مربعات الخطأ لـ RR أقل من مجموع مربعات الخطأ لـ LS أي [59][57][56][55][54] $\text{MSE}(\beta_{RR}) < MSE(\beta_{LS})$.

شكل (3 – 3): التقدير بالمربعات الصغرى المقيدة في بُعدين، ومحيط المنحرف لمجموع مربعات البواقي والدالة انحدار الحافة المقيدة.



المصدر: مرجع [63].

بأخذ المعادلة (21.3) في الاعتبار يمكن صياغة العلاقة بين مقدرات انحدار الحافة ومقدرات طريقة المربعات الصغرى وفقا لما يلي[57][56][55][54]:

$$\hat{\beta}_{RR} = (X'X + kI)^{-1}(X'X)\hat{\beta}_{LS} \qquad 22.3$$

وبافتراض أن:

$$Z = \{I + k(X'X)^{-1}\}^{-1} \qquad 23.3$$

فإن:

$$\hat{\beta}_{RR} = Z_k\hat{\beta}_{LS} \qquad 24.3$$

عليه بما أن $E(\hat{\beta}_{RR}) = E(Z_k\hat{\beta}_{LS}) = Z_k\beta$، فان $\hat{\beta}_{RR}$ مقدر متحيز بالنسبة β. ويرمز إلى عامل التحيز بـ k، ومصفوفة التباين لمقدر انحدار الحافة $\hat{\beta}_{RR}$ هي:

$$V(\hat{\beta}_{RR}) = \sigma^2(X'X + kI)^{-1}X'X(X'X + kI)^{-1} \qquad 25.3$$

$$V(\hat{\beta}_{RR}) = \sigma^2 Z(X'X)^{-1}Z' \qquad 26.3$$

3 – 5 – 2: متوسط مربعات الخطأ Mean Square Error

متوسط مربعات الخطأ يعتبر من المقاييس المهمة الدالة على جودة التقدير لذلك غالبا ما يستخدم لمقارنة أداء نماذج الانحدار المختلفة، لأنه عبارة عن التباين مضافاً إلية مربع مقدار تحيز مقدر انحدار الحافة. بالرجوع إلى المعادلة (3.15) فيمكن تعريف مجموع مربعات الخطأ لمقدر انحدار الحافة كما يلي:

$$MSE(\hat{\beta}_{RR}) = V(\hat{\beta}_{RR}) + (bias\ in\ \hat{\beta}_{RR})^2 \qquad 27.3$$

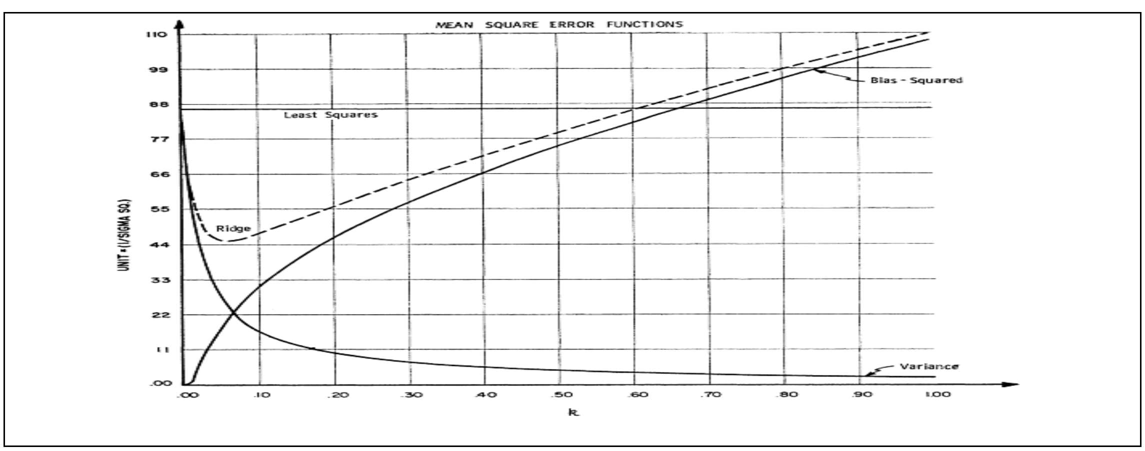
$$= \sigma^2 Tr[(X'X + kI)^{-1}X'X(X'X + kI)^{-1}] + k^2\beta'(X'X + kI)^{-2}\beta$$

$$= \sigma^2 \sum_{j=1}^{p} \frac{\lambda_j}{(\lambda_j + k)^2} + k^2\beta'(X'X + kI)^{-2}\beta \qquad 28.3$$

$$= \gamma_1(k) + \gamma_2(k) \qquad 29.3$$

إذن $\gamma_1(k)$ تعتبر دالة مطردة التناقص بالنسبة إلى k ، في حين $\gamma_2(k)$ تكون دالة مطردة التزايد بالنسبة إلى k . عليه إذا كان $k > 0$ فان مقدار التحيز في $\hat{\beta}_{RR}$ يزداد مع ازدياد قيمة k، بينما ينخفض التباين مع ازدياد k ، أي بالنسبة إلى عامل التحيز يكون التباين في تناقص بينما تزداد قيمة التحيز والشكل (3 – 4) يوضح ذلك[57][56][55][54].

شكل (3 – 4): دالة طريقة المربعات الصغرى ودالة انحدار الحافة بالنسبة إلى عامل التحيز k



المصدر: مرجع رقم [63].

3 – 5 – 3: مُقدرات الحافة المعيارية Standardized Ridge Estimators

من الأهمية بمكان التنبه إلى أن الحل المعطاء بواسطة انحدار الحافة غير ثابت بالنسبة إلى مستويات قياس المتغيرات الداخلة في النموذج، عليه أولاً لا بد من معايرة Standardization قيم المتغيرات التفسيرية ومتغير الاستجابة وفقا للصيغ التالية:

$$x'_{ij} = \frac{1}{\sqrt{n-1}} \frac{\left(X_{ij} - \bar{X}_j\right)}{S_{X_j}} \quad , \qquad i = 1,2,\dots,n \ , j = 1,2,\dots,p \qquad 30.3$$

$$y'_i = \frac{1}{\sqrt{n-1}} \frac{(Y_i - \bar{Y})}{S_Y} \quad , \qquad i = 1,2,\dots,n \qquad 31.3$$

وبناءًا على معايرة المتغيرات فان مصفوفة معلومات النظام $X'X$ والمصفوفة $X'Y$ تتحولان إلى R_{XX} و R_{XY} على التوالي، وتقدم المصفوفة R_{XX} معاملات ارتباط، لذلك تسمى بمصفوفة معاملات الارتباط بين المتغيرات التفسيرية $X_j's$، ويوضح المتجه R_{XY} معاملات ارتباط، ويعرف بمتجه معاملات الارتباط بين المتغير Y وجميع المتغيرات التفسيرية $X_j's$. وبذلك يمكن كتابة معادلة مُقدرات انحدار الحافة المعياري Standardized Ridge Regression وفقا للصيغة (32.3):

$$\hat{\beta}_{RR} = \{R_{XX} + kI\}^{-1} R_{YX} \qquad 32.3$$

في حالة تعامد المتغيرات التفسيرية حيث تكون $X'X = I$ فإنه من السهولة ملاحظة أن تقديرات الحافة تكون قيمة محجمة لمقدرات المربعات الصغرى أي أن $\hat{\beta}_{RR} = 1/(1+k) \cdot$ $\hat{\beta}_{LS}$ لبعض قيم ثابت التحيز الواقعة في المدى $0 \leq 1/(1+k) \leq 1$. بالإضافة إلى أن الحد الثابت Intercept ستؤول قيمته إلى الصفر $\beta_0 = 0$ عند إستخدام البينات المُعايرة Standardized Data.

بعد الحصول على مقدرات الحافة $\hat{\beta}_{RR}$ يتم تحويلها إلى قيم اعتيادية قبل بناء نموذج الانحدار الخطي الموفق بين القيم الأصلية لمتغير الاستجابة Y_i وقيم المتغيرات التفسيرية $X_j's$.

عليـــه لا بـــد مـــن توضـــيح عمليـــة معـــايرة Standardization البيانـــات وإعـــادة تحويلهـــا Transformation إلى قيمها الأصلية باستخدام جبر المصفوفات[57][56][55][54].

بـافتراض أن X_0 هـي مصـفوفة النظـام الأصـلية فـإن صـورتها المركزيـة Centered Version هي:

$$X_c = \left(I - \frac{j_n j_n'}{n}\right) X_0 \qquad 33.3$$

ونسختها المعيارية هي:

$$x = X_c L^{-\frac{1}{2}} \qquad 34.3$$

حيث j_n عبارة عن متجه في البُعد n عناصره واحدات، و L عبارة عن مصفوفة قطريـة عناصر قطرها من المصفوفة $X_c' X_c$ أي أن $L = diag(X_c' X_c)$، وعلى نفس المنوال فإن القيم الأصلية لمتجه الاستجابة Y_0 يمكن معايرتها وفقا لـ:

$$y = \frac{\left(I - \frac{j_n j_n'}{n}\right) Y_0}{S_Y} \qquad 35.3$$

حيث S_Y عبارة عن الانحراف المعياري لـ Y_0. معطى مقدر الحافة $\hat{\beta}_{RR}$ وفقا للمعادلة (21.3) وبافتراض أننا بصدد التنبؤ بمصفوفة بيانات جديدة X_{new} ذات أبعاد $m \times p$ للبينات الأصلية. لا بد من التنبه هنا إلى عدم إضافة j_m باعتبارة العمود الأول إلى المصفوفة X_{new} والذي يقابل الحد الثابت Intercept لأنه غير ظاهر في هذه الحالة. وفي هذه الحالة يعطى متجه متغير الاستجابة الجديد $\hat{Y}_{new}$ وفقا للمعادلة (36.3):

$$Y_{new} = S_Y \cdot \left\{\left(X_{new} - \frac{j_m j'_n}{n}\right) L^{-\frac{1}{2}} \hat{\beta}_{RR} + \frac{j_m j'_n Y}{n}\right\} \qquad 36.3$$

وفيما يلـي يمكـن حسـاب التوقـع والتبـاين للمقـدر $\hat{\beta}_{RR}$ ، ونكـون بحاجـة إلـى مقارنـة $\hat{\beta}_{RR}$ مـع $\hat{\beta}_{LS}$ لمعرفة إمكانية الحصول على متوسط مربعات الأخطاء MSE أصـغر لــ $\hat{\beta}_{RR}$ عنـد قيمـة معينـة لـ k[54][55][56][57].

3 – 5 – 4: انحدار الحافة المعمم Generalized Ridge Regression

أيضا اقترح كل من Hoerl وKennardامتدادا لانحـدار الحافـة الاعتيـادي، بحيث يمكـن مـن تخصـيص عامل تحيز خاص بكل متغير تفسيري يعرف بانحدار الحافة المعمم، ولاستعراضـه نضـع الصيغة العامـة للانحـدار الخطـي وفقـا للصـيغة القانونيـة/المتعـارف عليهاCanonical form بحيث تكـون المصـفوفة $X'X$ مصـفوفة قطريـة لهـا التحويـل المتعامـد P ويتبـع ذلـك أن $X'X = P'\Lambda P$ والمصـفوفة Λ و P عبـارة عـن مصـفوفتي القيم المميـزة والمتجهـات المميـزة للمصـفوفة $X'X$ وهـي تحقـق $P'X'XP = \Lambda = diagonal(\lambda_1, \lambda_2, \dots, \lambda_p)$ حيـث λ_j هـي القيمـة المميـزة رقـم j للمصـفوفة $X'X$ و $P'P = PP' = I_p$ وبـذلك يكـون النمـوذج المقابـل هو[78][57][56][55][54]:

$$Y = Z\alpha + \varepsilon \qquad 37.3$$

حيث:

$$\alpha = P'\beta \ , \qquad Z'Z = \Lambda \ , \qquad Z = XP \qquad 38.3$$

وتحسب الصيغة العامة لتقدير المربعات الصغرى لـ α وفقا لـ (3-28):

$$\hat{\alpha} = (Z'Z)^{-1}Z'Y = \Lambda^{-1}Z'Y \qquad 39.3$$

ويكون مقدر β هو $\hat{\beta} = P\hat{\alpha}$، أما الاجراء العام لتقدير انحدار الحافة المعمم فيمكن تعريفه وفقا للمعادلة (3.29):

$$\hat{\alpha}_{GR} = \left[I - kA_k{}^{-1}\right] \cdot \hat{\alpha} \qquad 40.3$$

حيث : $A_k = (\Lambda + kI_p)$ ، وتحسب معلمات طريقة انحدار الحافة المعمم وفقا للصيغة (41.3).

$$\hat{\beta}_{GR} = P\hat{\alpha}_{GR} \qquad 41.3$$

ويحسب متوسط مربع الخطأ لـ $\hat{\alpha}_{GR}$ وفقا للصيغة (3.31):

$$MSE(\hat{\alpha}_{GR}) = Var(\hat{\alpha}_{GR}) + [Bias(\hat{\alpha}_{GR})]^2$$

$$= \hat{\sigma}^2 \sum_{j=1}^{p} \lambda_j \Big/ (\lambda_j + k)^2 + k_j^2 \sum_{j=1}^{p} \alpha_j{}^2 / (\lambda_j + k)^2 \qquad 42.3$$

حيث: $\hat{\sigma}^2$ عبارة عن مقدر OLS لـ σ^2 ويحسب وفقا لـ $\hat{\sigma}^2 = \frac{Y'Y - \hat{\alpha}'Z'Y}{n-p-1}$ و $\alpha = P'\beta$.

ومعلمة الحافة التي تؤدي إلى الوصول إلى متوسط مربع أخطاء أصغر ما يمكن وفقا للصيغة (43.3) وهي تعبر عن الصيغة المقترحة لإيجاد معلمة الحافة k.

$$k = \frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}, j = 1,2, \dots, p \qquad 43.3$$

وبما أن القيمة المثلى لـ k_j تعتمد على معلمات غير معروفة σ^2 و α_jاقترح Hoerl وKennard طريقة تكرارية لتحديد قيم k_j إعتمادا على القيمة الابتدائية له التي يتم الحصول عليها وفقا لطريقة المربعات الصغرى الاعتيادية حيث تكون القيمة الابتدائية وفقا لـ $k_j^0 = \frac{\hat{\alpha}^2}{\hat{\alpha}_j^2}$ ويتم حساب القيمة الابتدائية لانحدار الحافة المعمم وفقا لـ $\hat{\alpha}_{GR}^0 = (\Lambda + K^0)^{-1}Z'Y$ حيث $K^0 = diag(K_1^0, K_2^0, \dots, K_p^0)$.

3 – 5 – 5: طرق اختيار معلمة الحافة (k)Methods of Choosing Ridge Parameter

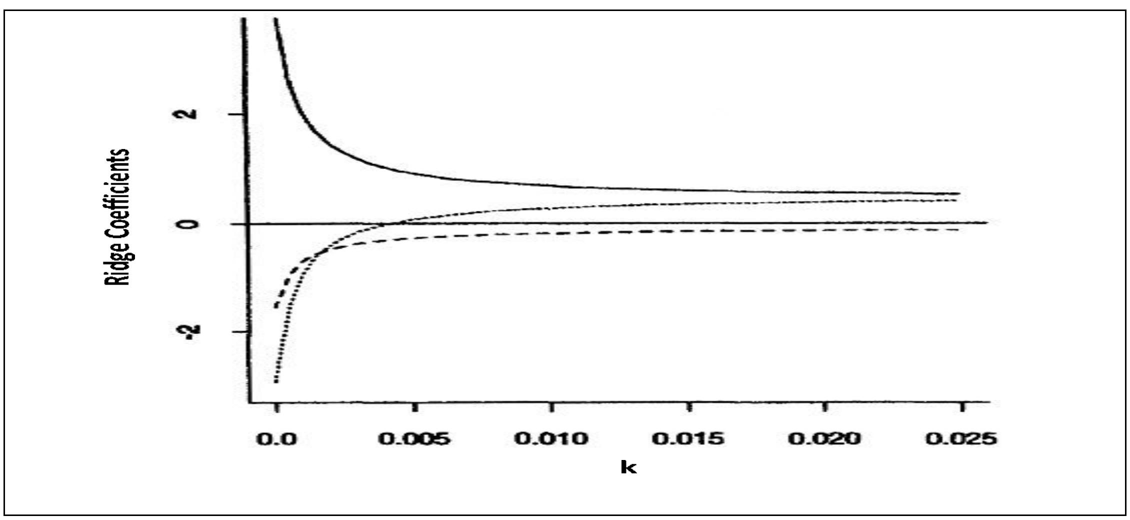
انحدار الحافة يعتبر دالة في معلمة الحافة أو عامل التحيز k ، حيث k عبارة عن قيمة موجبة تضاف إلى القطر الرئيس للمصفوفة $X'X$ كما شاهدنا على المعادلة (21.3)، وتقع قيمة

معلمة الحافة ضمن المجال $1 > k > 0$. يمكن التفريق بين طريقتين رئيستين لاختيار معلمة الحافة هما طريقة أثر الحافة Ridge Trace وطريقة تقدير قيمة k وفيما يلي إستعراضا للطريقتين:

طريقة أثر الحافة The Ridge Trace

طريقة أثر الحافة تعتمد على الرسم لتحديد القيمة المثالية لـ k ومن ثم تلعب خبرة الباحث أو الإحصائي دورا مهما في الاختيار لذلك تعتبر طريقة أثر الحافة طريقة غير موضوعية، وهي إذا افترضنا وجود p من العناصر لمقدر الحافة $\hat{\beta}_{RR}$ كدوال في k فإنه يتم رسمها جميعا مقابل قيم k في شكل واحد ضمن المدى $0 \leq k \leq 1$ ، ومع الزيادة التدريجية لقيمة عامل التحيز إبتداء من الصفر ضمن المدى المحدد يقوم الباحث باختيار قيمة $k = k_{trace}$ التي تشهد عندها الدالة استقرارا ملحوظا، والشكل (3 – 5) لثلاث دوال لانحدار الحافة حيث يلاحظ أنها أخذت تستقر عند $k = 0.010$ وبذلك يتم اختيار هذه القيمة لإيجاد قيم معلمات إنحدار الحفاة لهذا النموذج[69][58][53][52].

شكل (3 – 5): أثر الحافة لثلاث دوال لمعلمات انحدار الحافة.



المصدر : مرجع [58]

طرق تقدير Methods of Estimate k

طريقة تقدير معلمة الحافة تعتبر طريقة موضوعية Objective Method مقابل طريقة أثر الحافة حيث إن اختيار القيمة المثلى لمعلمة الحافة يعتمد فقط على قيم متغير الاستجابة Y. بصورة

عامة يهدف انحدار الحافة إلى إيجاد قيمة k التي من شأنها خفض التباين بشكل كبير مقابل الزيادة في قيمة مربع التحيز.

- **طرق Hoerl و Kennard و Baldwin:**

اقترح كل من Hoerl وKennard في ورقتهما التي نشرت في عام (1970) إيجاد قيمة k وفقاً للمعادلة (44.3) التي تؤدي إلى خفض MSE كأول معادلة لحساب قيمة معلمة الحافة.

$$k_{HK} = \frac{\hat{\sigma}^2}{\hat{\alpha}^2} \qquad 44.3$$

وفي عام (1975) قدم كل من Hoerl و Kennard و Baldwin صيغة جديدة لتحديد معلمة الحافة المثلى كتقدير لـ k وتمثلها المعادلة (45.3).

$$k_{HKB} = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} \qquad 45.3$$

حيث: $\hat{\alpha}$ و $\hat{\sigma}^2$ يتم الحصول عليهما من حل طريقة المربعات الصغرى الاعتيادية. وفي ورقتهما المنشورة عام (1976) وضع كل من Hoerl وKennard خوارزمية لتقدير قيمة معلمة الحافة تعتمد على المعادلة (45.3) وفقا للخطوات التالية:

$$\hat{\alpha}\ k_0 = \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$$

$$\hat{\alpha}_{RR}(k_0)k_1 = \frac{p\hat{\sigma}^2}{\hat{\alpha}'_{RR}(k_0)\hat{\alpha}_{RR}(k_0)}$$

$$\hat{\alpha}_{RR}(k_1)k_2 = \frac{p\hat{\sigma}^2}{\hat{\alpha}'_{RR}(k_1)\hat{\alpha}_{RR}(k_1)}$$

.

.

.

ويستخدم التغير النسبي في k_j لوقف التكرار إذا كانت $\frac{k_{j+1}-k_j}{k_j} > 20T^{-1.3}$، وهكذا تستمر الخوارزمية في التكرار أو في حالة الخروج يتم استخدام $\hat{\alpha}_{RR}(k_j)$ حيث $T = tr(X'X)^{-1}/p$[52][57].

- **طريقة Lawless و Wang:**

قدم كل من Lawless و Wang طريقتهما لإيجاد معلمة الحافة في ورقتهما التي نشرت في عام (1976) ويلاحظ أن طريقتهما تعتمد نوعاما على k_{HKB} حيث تمثلها المعادلة (3.34) [97] :

$$k_{LW} = \frac{p\hat{\sigma}^2}{\sum_{j=1}^{p} \lambda_j \hat{\alpha}_j^2} \qquad 46.3$$

- **طريقة Hocking و Sped و Lynn:**

اقترح كل من Hocking و Sped وLynn في عام (1976) طريقة جديدة لتحديد معلمة الحافة كما في المعادلة [14](47.3):

$$k_{HSL} = \hat{\sigma}^2$$

- **طريقةKibria:**

في عام (2003) قدم Kibria طريقتين جديدتين لإيجاد معلمة الحافة تعتمدان على فكرة الوسط الهندسي والوسيط وفيما يلي نتناول طريقة الوسيط كما على المعادلة[63](48.3).

$$k_{MED} = M$$

- **طريقةKhalaf و Shukur:**

اقترح كل من Khalaf و Shukur في عام (2005) طريقة جديدة كتعديل لـ k_{HK} لتحديد معلمة الحافة وفقا للمعادلة [19](49.3):

$$k_{KS} = \frac{}{(n}$$

- **طرق Alkhamisi وآخرون:**

أيضا اعتمادا على طريقة Kibria وطريقة Khalaf وShukur، قدم كل من Alkhamisi وآخرون في عام (2006) عدة طرق جديدة نتناول منها الطريقة التي تمثلها المعادلة [19](50.3):

$$k_{arith}^{KS} = \frac{1}{p}\sum_{j=1}^{p}\frac{\lambda_j\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_j\hat{\alpha}_j^2} \qquad 50.3$$

- **طرق .Alkhamisi, M و Shukur:**

في عام (2007) في ورقتهما Developing Ridge Parameters For SUR Model قدم كل من .Alkhamisi, M و Shukurعدة طرق لإيجاد معلمة الحافة اعتمادا على $K_j = \frac{1}{\hat{\alpha}_j^2}$ حيث في الطريقة $k_{j(Sqarith)}$تم استخدام فكرة الوسط الحسابي لـ $\sqrt{k_j}$ فكانت الطريقة وفقا للمعادلة [19](51.3):

$$k_{j(Sqarith)} = mean\left(\frac{1}{\sqrt{\hat{\alpha}_j^2}}\right) \qquad 51.3$$

وفي ورقتهما المنشورة في ذات العام (2007) تحت عنوان A Monte Carlo Study of Recent Ridge Parameters قاما باقتراح عدة طرق نتناول منها الطريقةk_{AS} وهي تعتمد k_{HK}صيغتها كما على المعادلة [19](52.3):

$$k_{AS} = \frac{\lambda_{max}\sigma^2 + \hat{\beta}_{max}^2}{\lambda_{max}\hat{\beta}_{max}^2} = \frac{\sigma^2}{\hat{\beta}_{max}^2} + \frac{1}{\lambda_{max}} = k_{HK} + \frac{1}{\lambda_{max}} \qquad 52.3$$

- **طريقة Al-Hassan:**

في عام (2010) قدم Al-Hassan طريقة جديدة لإيجادk معتمدا على التعديلات التي أدخلها كل من Alkhamisi و Shukur على الطريقة k_{HSL} التي قدمها Hocking وآخرون وفيما يلي نستعرضها[16].

$$k_{NHSL} = \frac{\hat{\sigma}^2 \lambda_{max} \sum_{j=1}^{p} \left(\lambda_j \hat{\alpha}_j\right)^2 + \left(\sum_{j}^{p} \lambda_j \hat{\alpha}_j\right)^2}{\lambda_{max} \left(\sum_{j=1}^{p} \lambda_j {\hat{\alpha}_j}^2\right)^2}$$

$$= \hat{\sigma}^2 \frac{\sum_{j=1}^{p} \left(\lambda_j \hat{\alpha}_j\right)^2}{\left(\sum_{j=1}^{p} \lambda_j {\hat{\alpha}_j}^2\right)^2} + \frac{1}{\lambda_{max}}$$

$$= k_{HSL} + \frac{1}{\lambda_{max}} \qquad 53.3$$

بما أن $k_{NHSL}\, is\ grater\ than\ k_{HSL}$ ،$1/\lambda_{max} > 0$

- **طريقة Dorugade و Kashid:**

في عام (2010) اقترح كل من Dorugade وKashid طريقة جديدة لإيجاد k تعتبر تعديلا لـ k_{HKB} حيث قاما بطرح عامل التضخم VIF بعد ضربه في حجم العينة والمعادلة هي[33]:

$$k_D = max\left(0, \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} - \frac{1}{n\left(VIF_j\right)_{max}}\right) \qquad 54.3$$

حيث إذا كان $\left(VIF_j\right)_{max}$ كبيرا جدا فإن k_D تقريبا تكون مساوية لـ k_{HKB} ، وإذا كان $\left(VIF_j\right)_{max}$ قريبا من الواحد فإن k_D يساوي الصفر أو $k_{HKB} - \frac{1}{n}$ ، عليه فإن $0 \le k_D \le$ k_{HKB} .

- **طرق Muniz وآخرون:**

في الورقة التي نشرت في عام (2012) قدمMuniz وآخرون عدة طرق لإيجاد معلمة الحافة، اعتمادا على Kabria (2003) و Khalaf and Skukur (2005) و Alkhamisi et al. (2006)، واعتمادا على فكرة تحويل الجذر التربيعي وفقا لـ *Alkhamisi and Shukur (2008)* قدم *Muniz* وآخرون عدة طرق نتناول منها الطريقتين (55.3) و (56.3)[79]:

$$k_{KM2} = max\left(\frac{1}{\sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}}\right) \qquad 55.3$$

$$k_{KM12} = median\left(\frac{1}{\sqrt{\frac{\lambda_{max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_{max}\hat{\alpha}_i^2}}}\right) \qquad 56.3$$

3 – 5 – 6: انحدار الحافة واختيار المتغيرات التفسيرية Ridge Regression and Variable Selection

على الرغم من أن عملية اختيار المتغيرات التفسيرية تلعب دوراً جيداً في ظل تعامد المتغيرات التفسيرية، نجد أن الخوارزميات الاعتيادية لاختيار المتغيرات المستقلة في نموذج الانحدار الخطي المتعدد لا تلعب ذات الدور في ظل بروز مشكلة التعدد الخطي. عليه فإن عملية تقريب المتغيرات التفسيرية من التعامد فيما بينها باستخدام عامل التحيز يخلق واقعاً جيداً لاختيار المتغيرات التفسيرية الأكثر تأثيرا في متغير الاستجابة. قدم كل من Hoerl وKennard في ورقتهما الثانية التي نشرت في ذات العام (1970) فكرة استخدام أثر الحافة كمعيار لاختيار المتغيرات التفسيرية وفقا لقاعدة حذف المتغيرات التفسيرية من النموذج الكامل والتي تلخصها الخطوات التالية[77][73][29][21]:

- حذف المتغير التفسيري ذو القدرة التنبؤية الصغرى من النموذج وإن كان مستقراً، ويستدل عليه بصغر معامله المعياري.
- حذف المتغير التفسيري الذي له معاملات غير مستقر وغير قادرة على عكس قدرته التنبؤية وتؤول إلى الصفر.
- حذف واحد أو مجموعة من المتغيرات التفسيرية التي لها معاملات غير مستقرة.

وتشكل المتغيرات المتبقية مجموعة جزئية عددها p يتم استخدامها في النموذج النهائي، ومن ثم يتم اختبار ما إذا كانت هذه المتغيرات التفسيرية تشكل مجموعة جزئية من المتغيرات التفسيرية شبه المتعامدة، حيث يمكن إجراء هذا الاختبار برسم $\hat{\beta}'_{RR}(k)\hat{\beta}_{RR}(k)$مربع طول متجه المعاملات

كدالة في k مقابل قيمة k. وعند تعامد المتغيرات التفسيرية فان مربع طول متجه مقدرات الحافة يكون $\hat{\beta}'\hat{\beta}/(1+k)^2$ حيث $\hat{\beta}$ عبارة عن مقدر المربعات الصغرى الاعتيادية لـ β . عليه إذا كان نموذج المجموعة الجزئية للمتغيرات التفسيرية قريب من التعامد فإن رسم الدوال $\hat{\beta}'_{RR}(k)\hat{\beta}_{RR}(k)$ و $\hat{\beta}'\hat{\beta}/(1+k)^2$ مقابل kيكون متشابه.

3 – 6: طريقة انحدار المكونات الرئيسية Principal Components Regression (PCR)

يعود تاريخ استخدام طريقة المكونات الرئيسية إلى أعمال كل من Beltrami في عام (1873) و Jordan في عام (1874) حيث قاما بشكل منفصل بوضع ما يسمى بـ Singular Value Decomposition (SVD) مما مهد إلى تعريف تحليل المكونات الرئيسية PCA من قبل كل من Pearson في عام (1901) و Hotelling في عام (1933)[78][61].

كما ذكرنا في بداية هذا الفصل تعتبر طريقة المكونات الرئيسية واحدة من النماذج الخطية المتحيزة الواسعة الاستخدام لتخطي مشكلة التعدد الخطي التي كثيرا ما يعاني منها نموذج الانحدار الخطي المتعدد في الواقع التطبيقي. تقوم طريقة المكونات الرئيسية على تحويل المتغيرات التفسيرية الأصلية المرتبطة دون حذف أي منها إلى متغيرات جديدة متعامدة تسمى بالمكونات الرئيسية، وكل مركب رئيسي عبارة عن تركيب خطي في المتغيرات المستقلة الأصلية.

تقدم المكونات الرئيسية قدر كبير من المعلومات عن مشاهدات المتغيرات الأصلية مثل أنماط تجمعاتها وعلاقتها بالمتغيرات الأصلية، وأيضا تقدم معلومات عن الارتباطات بين المتغيرات الجديدة والقديمة والمجموعات أو التصنيفات التي تحتويها البيانات أو المتغيرات. عادة يتم ترتيب المكونات الرئيسية وفقاً لمقدار التباين بحيث تكون المركبة الأولى هي المركبة ذات التباين الأكبر، ومن ثم يتم اعتماد عدد قليل من المكونات التي يتوقع أن تفسر أكبر قدر من التباين، ويتم إهمال المكونات ذات التأثير الأقل. وتعتبر عملية إيجاد المكونات الرئيسية خطوة مهمة لإزالة أثر التعدد الخطي تمهيداًلاستخدام طريقة المربعات الصغرى الاعتيادية لتقدير معالم نموذج الانحدار الخطي الأصلية للمتغيرات التفسيرية[50][25][20].

إذا كانت $(X_1, X_2, \dots, X_p)$ متغيرات تفسيرية، فيمكن تعريف توليفة متعامدة منها وفقاً للمعادلة (57.3).

$$Z = XA \qquad 57.3$$

حيث تمثل Z مصفوفة المكونات الرئيسية من الرتبة $(n \times p)$، بينما المصفوفة A عبارة عن مصفوفة متعامدة للمتجهات المميزة المعيارية المناظرة للجذور المميزة لمصفوفة معلومات النظام $(X'X)$ ورتبتها $(p \times p)$، عناصرها a_{ij} وأعمدتها A_j وهي تجعل المصفوفة $(X'X)$ مصفوفة قطرية، وباعتبار $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$ قيم مميزة للمصفوفة $(X'X)$ فإن المتغير Z_j يتوزع بمتوسط يساوي الصفر وتباين λ_j[8][61][68].

وللتعبير عن Y كدالة في المكونات الرئيسية بدلاً من المتغيرات المستقلة $(X_1, X_2, \dots, X_p)$ المرتبطة فيما بينها، وبما أن A مصفوفة متعامدة حيث $AA' = I$ فيمكن إعتبار $X = ZA'$ بالنسبة إلى معادلة نموذج الانحدار (2.2)فنحصل على نموذج الانحدار وفقاً للمعادلة (58.3) التالية:

$$Y = ZA'\beta + \varepsilon \qquad 58.3$$

وعلى افتراض أن $A'\beta = \gamma$ فتصبح المعادلة لنموذج الانحدار وفقاً للمعادلة (59.3) التالية:

$$Y = Z\gamma + \varepsilon \qquad 59.3$$

حيث γ تمثل متجه المعلمات $(\gamma_1, \gamma_2, \dots, \gamma_p)$ المناظرة للمركبات الرئيسية $(Z_1, Z_2, \dots, Z_p)$التي يمكن تقديرها باستخدام طريقة المربعات الصغرى الإعتيادية وفقاً للمعادلة (60.3) .

$$\gamma = (Z'Z)^{-1}Z'Y = \Lambda^{-1}ZY \qquad 60.3$$

والمصفوفة Λ تعتبر مصفوفة قطرية من الرتبة (p × p) عناصرها الجذور المميزة للمصفوفة $(X'X)$. والتوقع لهذه المعلمات هو $E(\hat{\gamma}) = \gamma$ وتباينها $var(\hat{\gamma}) = \Lambda^{-1}\sigma^2$، عليه يمكن القول إن متجه المعلمات $\hat{\gamma}$ له توزيع طبيعي بمتوسط γ وتباين $\Lambda^{-1}\sigma^2$. وبالتالي فإن تباين أي معلمة ضمن متجه المعلمات $\hat{\gamma}$ يحسب وفقا للصيغة $var(\gamma_j) = \frac{\sigma^2}{\lambda_j}$، مع الأخذ في الاعتبار أن $\hat{\gamma}_i = Z_i'Y/\lambda_i$ ، ويتم التنبؤ بقيمة المتغير التابع Y وفقا للمعادلة [68][61][8](61.3).

$$\hat{Y} = \sum_{j=1}^{p} Z_j\hat{\gamma}_j \qquad 61.3$$

ويكون التباين المخفض لتوفيق نموذج الانحدار باستخدام المركب الرئيسي Z_j يساوي المقدار $\lambda_j\gamma_j^2$ ، عليه فإن نسبة التباين المفسر في قيم متغير الاستجابة Y بواسطة المركب الرئيسي Z_j هي $100 \times \left(\lambda_j\gamma_j^2/Y'Y\right)$، وتساوي هذه النسبة مربع معامل ارتباط متغير الاستجابة والمركب الرئيسي Z_j مع ضرب الناتج في مائة. وبناءًا عليه يكون مربع الخطأ لتوفيق خط الانحدار هو:

$$MSE = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2 - \sum_{j=1}^{p}\lambda_j\hat{\gamma}_j^2 \qquad 62.3$$

وللحصول على معلمات المتغيرات التفسيرية الأصلية لنموذج الانحدار، يستفاد من العلاقة بين المعلمات الأصلية $\hat{\beta}$ ومعلمات نموذج الانحدار $\hat{\gamma}$ الخاصة بانحدار المتغير Y على المكونات الرئيسية Z وفقاً لما يلي:

إذا كان

$$A'\hat{\beta} = \hat{\gamma}$$

و

$A'A = I$

فان

$\hat{\beta} = A\hat{\gamma}$

وبما أن المعلمات $\hat{\beta}$ تتوزع طبيعياً بمتوسط $A\gamma$. فان القيمة المتوقعة للمعلمة β_i تحسب وفقاً لما يلي:

$$E(\hat{\beta}_i) = \sum_{j=1}^{p} a_{ij}\gamma_j \qquad 66.3$$

وبما أن المعلمات $\hat{\beta}$ لها التباين $\Lambda^{-1}A\sigma^2$. فان تباين المعلمة β_i يحسب وفقاً للمعادلة التالية:

$$var(\hat{\beta}_i) = \sigma^2 \sum_{j=1}^{p} {a_{ij}^2}\Big/{\lambda_j} \qquad 67.3$$

وعند استخدامالجذور المميزة لمصفوفة معاملات الارتباط بدلاً من مصفوفة التباين والتغاير كمدخلات في تحليل انحدار المكونات الرئيسية فانه يجب استخدام $n\lambda_j$ بدلا من λ_j .

وتوضح العلاقة (3.73) أن تباين معلمات نموذج الانحدار الأصلية المقدرة $\hat{\beta}$ أيضا تعتمد على الجذور المميزة للمصفوفة $(X'X)$وبناءًا على ذلك فهي تتأثر بوجود الجذور المميزة الصغيرة التي ينتج عنها تضخم التباينات. ووفقاً لتعريف المكونات الرئيسية فإن الجذور المميزة الصغيرة التي تسهم في تضخيم تباين معلمات نموذج الانحدار دائما تقابل المكونات الرئيسية الأخيرة للمصفوفة $(X'X)$، عليه يتطلب تخفيض التباين الكلي للمعلمات، استبعاد المكونات الرئيسية المقابلة لأصغر الجذور المميزة بالمصفوفة $(X'X)$.

اقتــرح بعــض البــاحثين أمثــال Jeffers وJolliffee و Price and Chatterjee أن يـتم اســتبعاد المكونـــات الرئيســية التـــي تقابـــل الجـــذور المميـــزة التـــي تقـــل عـــن 70%. كمـــا اقترحMorrisonاختيارالمكونـات الرئيسية التـي تفسر علـى الأقل 75% مـن التبـاين فـي قيم متغيـر الاسـتجابة. وهـذه النسـبة يمكـن الحصـول عليهـا بقسـمة مجمـوع الجـذور المميـزة المقابلـة لــ k مـن المكونــات الرئيسـية علــى مجمـوع الجـذور المميـزة عنـد اسـتخدام مصـفوفة التباينـات والتغـايرات للمتغيرات التفسيرية كمدخلات لتحليل المكونات الرئيسية وفقاً[91][68][61][8]:

$$\left(\sum_{j=1}^{k}\lambda_j \Big/ \sum_{j=1}^{p}\lambda_j\right) \times 100 \qquad 68.3$$

أما عند استخدام مصفوفة معاملات الارتباط كمدخلات لتحليل المكونات الرئيسية فعندئذ يتم استخدام عدد المتغيرات التفسيرية pبدلاً من مجموع الجذور المميزة.

يتم بناء نموذج الانحدار لمتغير الاستجابة Y علـى المكونـات الرئيسية المتبقيـة، بعد استبعاد بعضها الذي لا يحقق المعايير السابقة. بافتراض أن S من الجذور المميزة لها قيم كبيرة من بين p من الجذور المميزة للمصفوفة (X′X)، يكون هناك $(p-s)$ من الجذور المميزة ذات قيم صغيرة، أي يتم إستبعاد عدد $(p-s)$ من المكونـات الرئيسية Z ، ومن ثم يجري توفيق نموذج انحدار Y على المكونات الرئيسية المتبقية وبذلك تكون المعادلة التنبؤية كما يلي[68][61][8] [50]:

$$\hat{Y}_S = \sum_{j=1}^{S} Z_j \hat{\gamma}_j \qquad 69.3$$

ويحسـب مجمـوع مربعـات الخطـأ الخـاص بنمـوذج الانحـدار الموفـق بعـدد S مـن المكونـات الرئيسية وفقا لما يلي:

$$\hat{\sigma}_S^2 = MSE = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2 - \sum_{j=1}^{S}\lambda_j\,\hat{\gamma}_j^2 \qquad 70.3$$

ولحسن الحظ فإن خاصية التعامد لمقدرات المربعات الصغرى لـ γ سوف لن تختلف في حال استخدام جميع المكونات الرئيسية أو مجموعة جزئية منها، وتأسيساً على ذلك يتم تقدير معلمات نموذج الانحدار المخفض γ_S وفقا لما يلي:

$$\hat{\gamma}_S = \Lambda_S^{-1} Z_S' Y \qquad 71.3$$

حيث $\Lambda_S = diag(\lambda_1, \lambda_1, \dots, \lambda_S)$

ويتم الحصول على متجه المعلمات $\hat{\gamma}_S$ بتجزئة المصفوفة وفقا لما يلي:

$$A = \left[A_S \colon A_{p-s}\right] \qquad 72.3$$

حيث $A_S = [A_1, A_2, \dots, A_S]$ ، و $A_1 = [a_{11}, a_{12}, \dots, a_{1S}]$

ومصفوفة المكونات الرئيسية هي: $Z = \left[Z_S \colon Z_{p-S}\right]$

حيث $Z_S = [Z_1, Z_2, \dots, Z_S]$

ومتجه المعلمات هو $\gamma_S = \left[\gamma_S \colon \gamma_{p-s-1}\right]$

وعند الحصول على تقدير متجه المعلمات $\hat{\gamma}_S$ يمكن إستخدامه في تقدير متجه معلمات نموذج الانحدار الأصلي $\hat{\hat{\beta}}$ وتحسب وفقا لما يلي:

$$\hat{\hat{\beta}} = A_S \hat{\gamma}_S \qquad 73.3$$

ويمكن الحصول على معلمات الانحدار للمتغيرات التفسيرية باستخدام معلمات الانحدار للمركبات الرئيسية وفقا للمعادلة التالية:

$$\hat{\hat{\beta}}_i = \sum_{j=1}^{S} a_{ij} \hat{\gamma}_j \quad , i = 1,2, \dots, p \qquad 74.3$$

وتباين $\hat{\hat{\beta}}_i$ يقدر وفقا لما يلي:

$$var(\hat{\hat{\beta}}_i) = \sigma_S^2 \sum_{j=1}^{S} \gamma_{ij}^2/\lambda_j \qquad 75.3$$

ويعتبر المقدر $\hat{\hat{\beta}}$ مقدر متحيز، ويحسب مقدار تحيزه وفقا لما يلي:

$$Bias = E\left(\hat{\hat{\beta}}_i\right) - \beta_i \qquad 76.3$$

$$= -\sum_{j=S+1}^{p} a_{ij}\gamma_j \qquad 77.3$$

وبذلك فان متوسط مربع الخطأ للمقدر $\hat{\hat{\beta}}$ يحسب وفقا لما يلي:

$$MSE\left(\hat{\hat{\beta}}_i\right) = Var\left(\hat{\hat{\beta}}_i\right) + Bias\ of\ \left(\hat{\hat{\beta}}_i\right)^2 \qquad 78.3$$

$$= \sigma_S^2 \sum_{j=1}^{S} a_{ij}^2/\lambda_j + \left(-\sum_{j=S+1}^{p} a_{ij}\gamma_j\right)^2 \qquad 79.3$$

الفصل الرابع

الجانب التطبيقي

الفصل الرابع

الجانب التطبيقي

4 – 1: تمهيد

يتناول هذا الفصل الجانب التطبيقي من البحث حيث فيه سوف يتم تناول تصميم التجارب ووصف خطوات تصميم وتنفيذ محاكاة مونت كارلو، وتناول العوامل التي سوف يتم إدخالها إلى نموذج المحاكاة بمستويات مختلفة تتفق وأغراض البحث، كما سيتم عرض نتائج دراسة المحاكاة.

4 – 2: تصميم الدراسة Study Design

يهدف هذا البحث إلى المقارنة بين أداء طريقة المربعات الصغرى الاعتيادية وطريقة المكونات الرئيسية وانحدار الحافة عندما يكون النموذج واقعا تحت تأثير مشكلة التعدد الخطي التام أوغير التام عند مستويات مختلفة، كذلك يهدف البحث إلى المقارنة بين أداء مقدرات الحافة وفقاً لعدة طرق مقترحة لتحديد معلمة الحافة (عامل التحيز). لتحقيق أهداف هذا البحث تم تنفيذ الدراسة التطبيقية وفقا للتصميم التالي:

في هذه التجربة واتباعاً لدراسة كل من Al-Hassan and Al-kassab، ودراسة Al-khamisi and Shukur، ودراسة Mansson وآخرون، ودراسة Dorugade و Kashid، سوف يتم مقارنة أداء ثلاثة طرق لتقدير معلمات الانحدار لبيانات واقعة تحت تأثير مشكلة التعدد الخطي وهذه الطرق هي: طريقة المربعات الصغرى الاعتيادية OLS، وطريقة المكونات الرئيسية PC، وانحدار الحافة RR حيث سيتم استخدام عدة طرق لتحديد معلمة الحافة والمتمثلة في الطرق المقترحة من قبلHoerl, Kennard والمقترحة من قبلHoerl, Kennard, and Baldwin الواردة بالجدول (4 – 1)، والطريقة المقترحة من قبل كل من Lawless and Wang ، كما سوف يتم استخدام الطريقة المقدمة من Hocking وآخرون، والطرق المقترحة بواسطة Kibria، وطريقة Khalaf and Shukur، وطرق Alkhamisis وآخرون، وطرق Alkhamisis and Shukur ، وطريقة Al-Hassan، والطريقة التي اقترحها Dorugade and Kashid، والطرق التي قدمها Muniz وآخرون[63][19][16][15][14].

جدول (4 – 1): بعض الطرق المقترحة لتحديد معلمة الحافة k.

طريقة تحديد k	رمز	طريقة تحديد k	رمز
$= \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}}$	k_{HKB}	$= \frac{\hat{\sigma}^2}{\hat{\alpha}^2}$	k_{HK}
$= \hat{\sigma}^2 \frac{\sum_{j=1}^{p}\left(\lambda_j \hat{\alpha}_j\right)^2}{\left(\sum_{j=1}^{p} \lambda_j {\hat{\alpha}_j}^2\right)^2}$	k_{HSL}	$= \frac{p\hat{\sigma}^2}{\sum_{j=1}^{p} \lambda_j \hat{\alpha}_j^2}$	k_{LW}
$= \frac{\lambda_{max}\hat{\sigma}^2}{(n-p-1)\hat{\sigma}^2 + \lambda_{max} \cdot \hat{\alpha}_{max}^2}$	k_{KS}	$= Median\left\{\frac{\hat{\sigma}^2}{\hat{\alpha}_j^2}\right\}$	k_{MED}
$= mean\left(\frac{1}{\sqrt{\hat{\alpha}_j^2}}\right)$	$k_{j(Sqarith)}$	$= \frac{1}{p}\sum_{j=1}^{p}\frac{\lambda_j\hat{\sigma}^2}{(n-p)\hat{\sigma}^2 + \lambda_j\hat{\alpha}_j^2}$	k_{arith}^{KS}
$= k_{HSL} + \frac{1}{\lambda_{max}}$	k_{NHSL}	$= k_{HK} + \frac{1}{\lambda_{max}}$	k_{AS}
$= max\left(\frac{1}{\sqrt{\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}}}\right)$	k_{KM2}	$= max\left(0, \frac{p\hat{\sigma}^2}{\hat{\alpha}'\hat{\alpha}} - \frac{1}{n\left(VIF_j\right)_{max}}\right)$	k_D
$= median\left(\frac{1}{\sqrt{\frac{\lambda_{max}\hat{\sigma}^2}{(n-p)\hat{\sigma}^2+\lambda_{max}\hat{\alpha}_i^2}}}\right)$			k_{KM12}

4 – 2 – 1: معيار تقييم الأداء:

وللمقارنة بين الطرق الثلاثة OLS وPC وRRبطرقه المختلفة لتحديد معلمة الحافة الواردة بالجدول (4 – 1)، وبما أنه من المفترض أن يكون لانحدار الحافة MSE أصغر، عليه اتباعاً لمعظم البحوث والدراسات السابقة والتي أهمها دراسة Al-Hassan and Al-kassab، ودراسة Muniz وآخرون، ودراسة Manssonوآخرون، سوف يتم استخدامMSE كمعيار للمقارنة بين أداء الطرق المختلفة ومدى قدرتها التنبؤية، حيث لطريقة OLS يتم استخدام الصيغة $MSE = \frac{\sum_{i=1}^{R}(\hat{\beta}-\beta)'(\hat{\beta}-\beta)}{R}$، حيث $\hat{\beta}$ تمثل متجهة المعلمات المقدرة بواسطة OLS و PCR وRR بطرقها المختارة في هذا البحث، وR تمثل عدد مرات تكرار التجربة، وبما أن MSE دالة في المعلمة β في التباين σ^2 ومعلمة الحافة أو ثابت التحيز k، فاذا تم تثبيت المتغيرات التفسيرية يتبع ذلك تخفيض قيمة MSEعند اختيار متجهة المعلمات، وحينئذ يمكن المقارنة بين أداء الطرق المختلفة بحيث تكون الطريقة الأفضل هي التي لها أصغرMSE. أما بالنسبة RRفسوف يتم استخدام الصيغة (42.3)،وبالنسبة لطريقة PCRسوف يتم استخدام الصيغة (79.3) لحساب قيم [48]MSE.

4 – 3: دراسة محاكاة مونت كارلو Monte Carlo Simulation Study

تم تصميم دراسة المحاكاة التي تهدف إلى مقارنة أداء مقدرات OLS، ومقدرات PCR، ومقدرات RR وفقا لمجموعة من طرق تحديد معلمة الحافة k الواردة بالجدول (4–1)، سوف تتم المقارنة عند مستويات مختلفة للتعدد الخطي. وفيما يلي نستعرض مجموعة من العوامل المؤثرة في نموذج المحاكاة والتي سوف تؤخذ في الاعتبار عند تصميم نموذج المحاكاة[95][94].

4 – 3 – 1: المتغيرات التفسيريةExplanatory Variables

في معظم دراسات محاكاة مقدرات انحدار الحافة السابقة مثل بحث Al-Hassan and Al-kassab، بحث Al-khamisi and Shukur، وبحث Muniz and Kibria، وبحث Muniz et al.كان عدد المتغيرات التفسيرية الأكثر استخداما هو (ما بين متغيرين إلى أربعة متغيرات تفسيرية)، وبصورة عامة يراوح عدد المتغيرات التفسيرية المستخدمة

بـــين (2) متغيـــرين تفســـيريين إلـــى (20) متغيـــرا تفســـيريا. فـــي هـــذا البحـــث اعتمـــد الباحـــث استخدام العدد ما بين (2 ، 4 ، 8) متغيرا مستقلا [19][16][15][14].

4 – 3 – 2: حجم العينة Sample Size

زيــادة حجــم العينـــة يتوقـــع يـــؤثر إيجابـــا علـــى متوســـط مربعـــات الخطـــأ MSE، عليـــه لدراســـة أثـــر حجـــم العينـــة علـــى جميـــع المقـــدرات تحـــت الدراســـة روعـــي اســـتخدام عينـــات بأحجـــام تتـــراوح بـــين عينـــات صـــغيرة الحجـــم وعينـــات متوســـطة الحجـــم وعينـــات كبيـــرة الحجـــم، عليـــه اتباعـــا لدراســـة Al-khamisi and Shukur، ودراســـة Al-Hassan and Al-kassab ودراسات أخرى، تم استخدام ثلاث عينات بالأحجام [19][16](n = 10, 50, 100).

4 – 3 – 3: مستويات التعدد الخطي بين المتغيرات التفسيرية

لتوليد المتغيرات التفسيرية المرتبطة خطيا استخدم مجموعة من الباحثين منهم McDonald and Galarneau، وGibbons، وKibria، Al-khamisi and Shukur، Al-Hassan and Al-kassab، GiselaMuniz، وباحثون آخرون، استخدموا مجموعة من المستويات للارتباط بين المتغيرات التفسيرية شملت عدة مستويات مختلفة للارتباط بين المتغيرات التفسيرية، عليه اتباعا لتطبيقات هؤلاء الباحثين في هذا البحث تم استخدام عدة مستويات هي (ρ = .70 , .90 , and .99)[79][63][19][16][15][14].

4 – 3 – 4: توليد المتغيرات Generating Variables

لتوليد قيم المتغيرات التفسيرية عند مستويات مختلفة من الارتباط بينها، وبأحجام عينات مختلفة، ووفقا لدراسات McDonald and Galarneau، وGibbons، وKibria، و Al-khamisi and Shukur، Al-Hassan and Al-kassab، Mansson et al.، ومجموعة من الباحثين تم استخدام المعادلة (4.1)التي تتيح توليد المتغيرات التفسيرية بمستويات مختلفة للارتباط بينها[79][72] [75] [76][63][44][37][19][16][10][3].

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}}$$

حيث تمثل ρ قيمة الارتباط بين المتغيرات التفسيرية، ويتم توليد قيم z_{ij} المستقلة عن بعضها البعض باستخدام التوزيع الطبيعي المعياري، وفقاً للخوارزمية التالية:

الخطوة الأولى: تم توليد أرقام عشوائية U_i تتبع التوزيع المنتظم Uniform Distribution حيث $U\sim U(0,1)$.

الخطوة الثانية: تم توليد متغيرات عشوائية z_{ij} تتبع التوزيع الطبيعي المعياري Standard Normal Distribution حيث $Z_i\sim N(0,1)$.

الخطوة الثالثة: تم استخدام المعادلة (1.4) لتوليد قيم المتغيرات المستقلة وفقاً لأحجام العينات المختلفة، ووفقا لمستويات الارتباط المختلفة.

الخطوة الرابعة: تم تحويل المتغيرات المستقلة إلى متغيرات موزعة طبيعيا بمتوسطات مختلفة وانحراف معياري ثابت قدره (5) وحدات.

الخطوة الخامسة: تم توليد قيم المتغير التابع لكل مجموعة من المتغيرات التفسيرية المختلفة، باستخدام المعادلة (2.4) مع اعتبار قيمة المعامل $\beta_o = 0$، وقيم حد الخطأ e_i موزعة طبيعيا بـ $e_i\sim N(0,\sigma^2)$.

$y_i = \beta_o + \beta$

جدول (4 – 2): المتغيرات التفسيرية وفقاً لعدد المشاهدات ومستوى الارتباط.

عدد المتغيرات	عدد المشاهدات	مستوى الارتباط	رمز العينة
متغيرين	10	0.70	N10V2R70
		0.90	N10V2R90
		0.99	N10V2R99
	50	0.70	N50V2R70
		0.90	N50V2R90
		0.99	N50V2R99
	100	0.70	N100V2R70
		0.90	N100V2R90
		0.99	N100V2R99
اربعة متغيرات	10	0.70	N10V4R70
		0.90	N10V4R90
		0.99	N10V4R99
	50	0.70	N50V4R70
		0.90	N50V4R90
		0.99	N50V4R99
	100	0.70	N100V4R70
		0.90	N100V4R90
		0.99	N100V4R99
ثمانية متغيرات	10	0.70	N10V8R70
		0.90	N10V8R90
		0.99	N10V8R99
	50	0.70	N50V8R70
		0.90	N50V8R90
		0.99	N50V8R99
	100	0.70	N100V8R70
		0.90	N100V8R90
		0.99	N100V8R99
اثني عشر متغيرا	10	0.70	N10V12R70
		0.90	N10V12R90
		0.99	N10V12R99
	50	0.70	N50V12R70
		0.90	N50V12R90
		0.99	N50V12R99
	100	0.70	N100V12R70
		0.90	N100V12R90
		0.99	N100V12R99

4 – 4: نتائج دراسة المحاكاة:

تم تنفيذ دراسة المحاكاة باستخدام برنامج SAS V. 9.2 و SAS Enterprise V. 4.2لتوليد قيم المتغيرات المختلفة وفقا المعايير المذكورة في المبحث (4 – 3) وحساب معلمات الطرق المختلفة محل المقارنة، كما تم إستخدام برنامج Microsoft Excel V. 13 في حساب قيم MSE للطرق المختلفة وإعداد الرسوم البيانية الخاصة بمقارنة أداء الطرق المختلفة وفقا للمعايير المختلفة[100][74][60][7].

فيما يلي استعراض نتائج محاكاة مونت كارلو الواردة بالجداول (4– 3 إلى 4 – 14) قيم مربعات الأخطاء MSE لنموذج انحدار المكونات الرئيسيةPCR، ونموذج المربعات الصغرى الاعتيادية OLS، ونماذج انحدار الحافة RRالمختلفة. حيث تم عرض النتائج وفقا للمعايير المختلفة، الانحراف المعياري للقيم σ حيث تم استخدام مستويات مختلفة للانحراف المعياري , $\sigma = .5$ 1. , 3.وتم استخدام انحراف معياري كبير نسبيا بحجم $\sigma = 5$، وعينات بأحجام تتراوح بين عينة صغير $n = 10$ وحجم كبير $n = 50$وحجم كبير نسبيا$n = 100$، ومستويات مختلفة للارتباط بين المتغيرات التفسيرية تراوحت بين ارتباط عال $(\gamma = .7)$ وعال جدا$(\gamma = .9)$ وارتباط قريب من الواحد $(\gamma = .99)$، وفيما يلي استعراض النتائج وفقا لمستوى التباين:

4 – 4 – 1: قيم متوسط مربعات الخطأ عند $\sigma = .5$

جدول (4 – 3): قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = .5).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	1.641	0.121	0.053	0.310	0.141	1.971	1.099	2.831	4.320	**0.034**	1.452	0.666
OLS	1.942	0.132	0.054	0.313	0.207	2.042	4.390	3.392	5.290	0.444	5.370	8.550
K_{HK}	1.827	0.131	0.053	0.206	0.037	0.892	1.577	0.166	0.949	0.049	0.798	0.541
K_{HKB}	1.827	0.131	0.053	0.206	0.039	0.892	1.577	0.166	0.914	0.049	0.798	0.541
K_{LW}	1.757	0.127	0.052	0.154	0.033	1.278	1.543	0.137	0.770	0.049	0.734	0.541
K_{HSL}	1.837	0.125	0.052	0.128	0.033	2.024	1.332	**0.124**	0.622	0.053	0.378	0.448
K_{MED}	1.816	0.125	0.051	0.104	0.033	0.893	1.207	0.127	0.499	0.061	0.726	0.465
K_{KS}	1.801	0.122	0.051	0.174	0.046	0.997	1.577	0.203	0.537	0.072	0.465	**0.443**
K_{arith}^{KS}	1.741	0.120	0.051	0.174	0.098	1.576	1.331	0.130	0.609	0.187	0.856	0.670
K_j	1.734	0.118	0.047	0.100	0.711	1.799	1.095	0.150	0.568	0.187	0.790	0.670
K_{AS}	1.740	0.118	0.041	0.116	0.035	**0.743**	2.000	0.152	0.386	0.088	**0.296**	0.579
K_{NHSL}	1.685	0.121	0.050	0.102	**0.031**	1.490	1.952	0.186	0.548	0.088	0.790	0.457
K_D	1.747	0.116	0.042	0.143	0.038	0.964	**0.733**	0.208	0.431	0.187	0.636	0.579
K_{KM2}	1.668	0.113	**0.040**	**0.096**	0.048	0.934	0.808	0.145	**0.320**	0.088	0.636	0.650
K_{KM12}	**1.603**	0.112	0.041	0.116	0.628	1.624	1.060	0.167	0.386	0.346	0.636	0.650

المصدر: الباحث بواسطة البرنامج.

النتائج الواردة بالجدول (4 – 3) عبارة عن قيم MSEلنماذج PCR ونماذج OLS ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = .25$)بين مشاهدات كل من المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية (γ = .7 ,.9 ,.99)،وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 10$.

جدول (4 – 4): قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (σ = .5).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**0.051**	0.067	0.053	**0.034**	0.073	**0.159**	1.553	0.893	**0.073**	0.131	0.208	0.291
OLS	0.103	0.072	0.055	0.365	0.066	0.176	1.663	1.429	1.381	1.899	3.840	4.260
K_{HK}	0.094	0.071	0.055	0.071	0.049	0.166	0.399	0.961	0.934	0.612	0.640	0.604
K_{HKB}	0.094	0.071	**0.044**	0.071	0.049	0.168	0.399	0.961	0.924	0.612	0.640	0.604
K_{LW}	0.093	0.070	**0.044**	0.138	0.049	0.168	0.350	0.961	0.917	0.148	0.160	0.325
K_{HSL}	0.091	0.069	0.052	0.146	0.048	0.169	0.334	0.996	0.852	0.814	0.195	0.620
K_{MED}	0.091	0.068	0.052	0.073	0.047	0.170	0.604	0.979	0.848	0.148	0.266	0.139
K_{KS}	0.090	0.067	0.051	0.324	0.044	0.172	0.959	0.922	0.909	0.199	0.185	**0.138**
K_{arith}^{KS}	0.090	0.068	0.050	0.213	0.040	0.169	0.159	**0.905**	0.922	**0.121**	0.175	0.724
K_j	0.088	0.067	0.049	0.104	0.040	0.174	0.438	0.929	0.914	0.148	0.814	0.856
K_{AS}	0.087	0.066	0.048	0.052	0.039	0.161	0.414	0.982	0.901	0.202	0.153	**0.138**
K_{NHSL}	0.090	0.065	0.048	0.174	0.048	0.171	0.217	0.952	0.822	0.199	0.266	0.403
K_D	0.086	**0.064**	0.046	**0.034**	0.039	0.172	0.414	0.970	0.830	0.168	0.164	0.724
K_{KM2}	0.084	**0.064**	0.048	0.186	0.042	0.178	0.529	0.979	0.817	0.371	**0.137**	0.197
K_{KM12}	0.075	**0.064**	0.047	0.138	**0.038**	0.174	**0.144**	0.941	0.810	0.168	0.164	0.197

المصدر: الباحث بواسطة البرنامج.

توضح النتائج الواردة بالجدول (4 – 4) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، عند مستوى تباين قدره ($\sigma^2 = .25$) بين مشاهدات كل من المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية ($\gamma = .7, .9, .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 50$.

جدول (4 – 5): قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = .5).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**0.109**	0.106	0.105	**0.057**	**0.070**	0.043	0.046	**0.868**	**0.282**	**0.775**	0.663	0.616
OLS	0.136	0.124	0.107	0.547	0.428	0.054	0.071	1.806	0.795	1.514	9.335	0.795
K_{HK}	0.124	0.120	0.104	0.547	0.415	0.054	0.060	0.927	0.546	0.990	0.635	0.501
K_{HKB}	0.124	0.120	0.105	0.547	0.415	0.053	0.060	0.927	0.518	0.990	0.635	0.462
K_{LW}	0.121	0.118	0.102	0.534	0.412	0.052	0.055	0.908	0.498	0.964	0.206	0.498
K_{HSL}	0.119	0.115	0.096	0.538	0.408	0.053	0.054	0.913	0.529	0.948	0.365	0.513
K_{MED}	0.120	**0.100**	0.096	0.511	0.373	0.049	0.053	0.926	0.553	0.922	0.511	0.502
K_{KS}	0.119	0.110	0.099	0.514	0.360	0.053	0.052	0.921	0.530	0.925	0.593	0.467
K_{arith}^{KS}	0.121	0.109	0.098	0.514	0.346	0.048	0.052	0.918	0.532	0.941	0.511	0.502
K_j	0.117	0.121	**0.093**	0.499	0.350	0.051	0.050	0.919	0.538	0.906	0.365	0.471
K_{AS}	0.117	0.107	0.102	0.499	0.410	0.052	0.049	0.923	0.565	0.922	0.593	0.539
K_{NHSL}	0.117	0.111	0.097	0.528	0.394	0.052	0.047	0.918	0.499	0.928	**0.192**	0.501
K_D	0.111	0.104	**0.093**	0.491	0.366	0.043	0.049	0.925	0.588	0.960	**0.192**	0.510
K_{KM2}	0.110	0.103	0.103	0.515	0.379	**0.042**	0.049	0.922	0.537	0.959	0.511	**0.462**
K_{KM12}	0.091	0.105	0.098	0.511	0.343	0.052	**0.045**	0.928	0.532	0.941	**0.192**	0.506

المصدر: الباحث بواسطة البرنامج.

النتائج الواردة بالجداول (4 – 5)توضح قيم MSEلنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، عند مستوى تباين ($\sigma^2 = .25$)بين مشاهدات كل من المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية ($\gamma = .7, .9, .99$)،وأعداد مختلفة من المتغيرات التفسيرية،وحجم عينة $n = 100$.

4 – 4 – 2: قيم متوسط مربعات الخطأ عند $\sigma = .3$

جدول (4 – 6): قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = .3).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**0.047**	0.044	0.019	**0.025**	0.057	0.712	**0.132**	10.77	14.64	**0.019**	0.654	0.243
OLS	0.696	0.045	0.021	0.117	0.194	0.739	1.623	10.87	15.37	0.183	7.747	6.972
K_{HK}	0.658	0.045	0.021	0.078	0.019	0.305	0.634	0.071	0.334	0.071	0.165	**0.211**
K_{HKB}	0.658	0.045	0.021	0.078	0.011	0.305	0.634	0.071	0.316	0.071	0.165	**0.211**
K_{LW}	0.639	0.044	0.020	0.058	0.014	0.436	0.612	0.054	0.276	0.071	0.130	**0.211**
K_{HSL}	0.676	0.043	0.020	0.045	0.014	0.731	0.481	**0.046**	0.218	0.119	0.112	0.244
K_{MED}	0.665	0.043	0.020	0.037	0.014	0.306	0.408	0.047	0.185	0.038	0.140	0.280
K_{KS}	0.662	0.042	0.020	0.063	0.013	0.341	0.634	0.118	0.198	0.183	0.097	0.271
K_{arith}^{KS}	0.632	0.042	0.020	0.063	0.086	0.535	0.480	0.048	0.219	0.032	0.185	0.306
K_j	0.631	0.041	0.018	0.036	0.551	0.632	0.346	0.067	0.208	0.032	0.162	0.575
K_{AS}	0.638	0.041	**0.015**	0.043	0.009	**0.254**	0.910	0.068	0.135	0.021	0.179	0.423
K_{NHSL}	0.611	0.042	0.019	0.035	**0.010**	0.506	0.879	0.104	0.200	0.021	0.162	0.230
K_D	0.638	0.041	0.016	0.052	0.011	0.330	0.169	0.114	0.156	0.032	**0.115**	0.423
K_{KM2}	0.612	**0.040**	**0.015**	0.035	0.015	0.320	0.209	0.064	**0.114**	0.021	**0.115**	0.289
K_{KM12}	0.589	**0.040**	0.016	0.041	0.479	0.554	0.338	0.092	0.147	0.103	**0.115**	0.289

المصدر: الباحث بواسطة البرنامج.

النتائج الواردة بالجدول (4 – 6) عبارة عن قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = .09$) بين مشاهدات كل من المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية ($\gamma = .7 , .9 , .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 10$.

جدول (4 – 7): قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (σ = .3).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**0.022**	0.050	0.018	**0.013**	0.024	**0.056**	**0.043**	**0.318**	0.517	**0.014**	**0.031**	**0.019**
OLS	0.700	0.026	0.018	120.4	0.024	0.058	0.710	0.535	0.545	3.523	7.862	11.16
K_{HK}	0.650	0.026	0.018	0.022	0.019	0.059	0.069	0.344	0.366	0.395	0.408	0.317
K_{HKB}	0.650	0.026	0.018	0.022	0.019	0.060	0.069	0.344	0.362	0.395	0.408	0.317
K_{LW}	0.628	0.026	0.018	0.049	0.019	0.060	0.061	0.344	0.359	0.046	0.048	0.120
K_{HSL}	0.665	0.026	0.017	0.067	0.019	0.060	0.058	0.371	0.333	0.568	0.067	0.329
K_{MED}	0.655	0.025	0.017	0.022	0.018	0.060	0.141	0.362	0.331	0.046	0.115	0.082
K_{KS}	0.651	0.025	0.017	0.179	0.017	0.061	0.323	0.330	0.354	0.202	0.160	0.072
K_{arith}^{KS}	0.621	0.025	0.017	0.091	0.015	0.060	0.127	0.322	0.361	0.081	0.056	0.410
K_j	0.621	0.025	0.016	0.031	0.015	0.062	0.080	0.336	0.358	0.046	0.907	0.516
K_{AS}	0.627	0.025	0.016	0.016	**0.014**	0.058	0.073	0.372	0.351	0.076	0.046	0.072
K_{NHSL}	0.600	**0.024**	0.016	0.088	0.019	0.061	0.107	0.349	0.319	0.202	0.115	0.171
K_D	0.627	**0.024**	0.015	**0.013**	**0.014**	0.062	0.073	0.354	0.323	0.157	0.130	0.410
K_{KM2}	0.601	**0.024**	**0.014**	0.096	0.016	0.064	0.110	0.362	0.318	0.200	0.084	0.055
K_{KM12}	0.578	**0.024**	0.015	0.049	**0.014**	0.062	0.243	0.341	0.316	0.157	0.130	0.055

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 7) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره (σ^2 = .09) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية (γ = .7, .9, .99)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة n = 50 .

جدول (4 – 8): قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = .3).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**0.039**	0.037	0.032	**0.020**	0.021	0.016	**0.016**	**0.295**	**0.101**	**0.264**	**0.050**	**0.148**
OLS	0.051	0.040	0.037	0.454	0.003	0.019	0.025	0.696	0.284	0.553	3.235	0.319
K_{HK}	0.046	0.039	0.036	0.454	0.003	0.019	0.021	0.354	0.208	0.370	0.248	0.211
K_{HKB}	0.046	0.039	0.037	0.454	0.003	0.018	0.021	0.354	0.192	0.370	0.248	0.184
K_{LW}	0.045	0.039	0.036	0.442	0.003	0.018	0.019	0.346	0.180	0.361	0.060	0.207
K_{HSL}	0.044	0.037	0.034	0.446	0.003	0.018	0.019	0.350	0.199	0.357	0.097	0.216
K_{MED}	0.044	0.036	0.033	0.421	**0.002**	0.017	0.019	0.355	0.214	0.347	0.745	0.211
K_{KS}	0.044	0.036	0.035	0.424	**0.002**	0.018	0.019	0.351	0.197	0.347	0.225	0.188
K_{arith}^{KS}	0.044	0.035	0.034	0.424	**0.002**	0.017	0.019	0.350	0.197	0.353	0.745	0.211
K_j	0.043	0.040	0.032	0.411	**0.002**	0.018	0.018	0.351	0.201	0.342	0.097	0.190
K_{AS}	0.043	0.035	0.035	0.411	0.003	0.018	0.017	0.354	0.218	0.347	0.225	0.238
K_{NHSL}	0.043	0.036	0.034	0.437	0.003	0.018	0.017	0.350	0.177	0.349	0.064	0.211
K_D	0.041	0.034	0.033	0.404	**0.002**	**0.015**	0.017	0.355	0.232	0.360	0.064	0.217
K_{KM2}	0.041	**0.032**	0.036	0.425	**0.002**	**0.015**	0.017	0.353	0.200	0.361	0.745	0.184
K_{KM12}	0.043	0.033	**0.031**	0.421	**0.002**	0.018	0.018	0.356	0.197	0.354	0.064	0.213

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 7) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = .09$) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية ($\gamma = .7, .9, .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 100$.

4 – 4 – 3: قيم متوسط مربعات الخطأ عند $\sigma = .1$

جدول (4 – 9): قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = .1).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**0.005**	0.005	0.002	**0.001**	0.005	0.074	0.064	1.140	1.666	**0.013**	0.152	**0.024**
OLS	0.077	0.005	0.002	0.012	0.234	0.080	0.210	1.214	1.915	0.046	0.887	0.777
K_{HK}	0.068	0.005	0.002	0.009	0.023	0.034	0.102	0.020	0.034	0.232	0.051	0.053
K_{HKB}	0.068	0.005	0.002	0.009	0.001	0.034	0.102	0.020	0.030	0.232	0.051	0.053
K_{LW}	0.067	0.005	0.002	0.007	0.014	0.048	0.094	0.010	0.030	0.232	0.161	0.053
K_{HSL}	0.076	0.005	0.002	0.004	0.014	0.079	0.046	0.008	0.021	0.325	0.372	0.159
K_{MED}	0.073	0.005	0.002	0.004	0.014	0.034	0.027	**0.006**	0.023	0.155	0.078	0.206
K_{KS}	0.074	0.005	0.002	0.006	0.002	0.038	0.102	0.066	0.024	0.432	0.258	0.209
K_{arith}^{KS}	0.065	0.005	0.002	0.006	0.109	0.059	0.046	**0.006**	0.024	0.019	**0.041**	0.126
K_j	0.067	0.005	0.002	0.004	0.431	0.069	0.015	0.021	0.025	0.019	0.054	0.559
K_{AS}	0.069	0.005	0.002	0.006	0.002	**0.028**	0.229	0.021	0.013	0.095	0.590	0.367
K_{NHSL}	0.063	0.005	0.002	0.003	0.002	0.056	0.215	0.054	0.023	0.095	0.054	0.021
K_D	0.069	0.005	0.002	0.006	**0.001**	0.036	0.029	0.061	0.018	0.019	0.123	0.367
K_{KM2}	0.066	0.005	0.002	0.004	**0.001**	0.035	**0.013**	0.019	**0.012**	0.095	0.123	0.114
K_{KM12}	0.065	0.005	0.002	0.005	0.367	0.061	0.016	0.045	0.021	0.014	0.123	0.114

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 9) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = .01$) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية ($\gamma = .7, .9, .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 10$.

جدول (4 – 10): قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (1. = σ).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**0.015**	0.054	0.002	**0.001**	0.003	**0.006**	0.061	**0.032**	0.054	**0.014**	0.015	**0.016**
OLS	0.077	0.003	0.002	12.49	0.003	**0.006**	0.167	0.057	0.058	0.372	0.889	1.202
K_{HK}	0.068	0.003	0.002	0.016	0.003	**0.006**	0.068	0.035	0.040	0.238	0.235	0.128
K_{HKB}	0.068	0.003	0.002	0.016	0.003	0.007	0.068	0.035	0.040	0.238	0.235	0.128
K_{LW}	0.067	0.003	0.002	0.006	0.003	0.007	0.098	0.035	0.040	0.005	0.007	0.015
K_{HSL}	0.076	0.003	0.002	0.019	0.002	0.007	0.108	0.047	0.037	0.382	0.006	0.136
K_{MED}	0.074	0.003	0.002	0.002	0.003	**0.006**	**0.012**	0.043	0.037	0.025	0.028	0.125
K_{KS}	0.074	0.003	0.002	0.085	0.002	0.007	0.029	0.034	0.038	0.268	0.216	0.106
K_{arith}^{KS}	0.066	0.003	0.002	0.024	0.002	0.007	0.409	**0.032**	0.040	0.102	**0.005**	0.194
K_j	0.067	0.003	0.002	0.003	0.002	0.007	0.051	0.036	0.040	0.025	1.096	0.273
K_{AS}	0.070	0.003	0.002	0.005	**0.001**	0.007	0.061	0.050	0.038	0.016	0.008	0.106
K_{NHSL}	0.063	0.003	0.002	0.032	0.003	0.007	0.319	0.041	0.034	0.268	0.028	0.039
K_D	0.069	0.003	0.002	0.005	**0.001**	0.007	0.061	0.040	**0.035**	0.210	0.175	0.194
K_{KM2}	0.067	0.003	0.002	0.037	0.002	0.007	0.023	0.043	**0.035**	0.089	0.107	0.012
K_{KM12}	0.065	0.003	0.002	0.006	**0.001**	0.007	0.651	0.037	**0.035**	0.210	0.175	0.012

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 10) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = .01$) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية المختلفة ($\gamma = .7 , .9 , .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 50$.

جدول (4 – 11): قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = .1).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	0.005	0.004	**0.003**	**0.002**	0.002	0.002	0.002	**0.033**	**0.010**	**0.034**	**0.009**	**0.011**
OLS	0.006	0.004	0.004	0.376	0.001	0.002	0.003	0.073	0.031	0.059	0.350	0.033
K_{HK}	0.005	0.004	0.004	0.376	0.001	0.002	0.002	0.038	0.023	0.039	0.037	0.021
K_{HKB}	0.005	0.004	0.004	0.376	0.001	0.002	0.002	0.038	0.019	0.039	0.037	0.015
K_{LW}	0.005	0.004	0.004	0.365	0.001	0.002	0.002	0.037	0.016	0.037	0.106	0.020
K_{HSL}	0.005	0.004	0.004	0.369	0.001	0.002	0.002	0.038	0.021	0.037	0.010	0.021
K_{MED}	0.005	0.004	0.004	0.346	**0.000**	0.002	0.002	0.038	0.025	0.036	0.304	0.021
K_{KS}	0.005	0.004	0.004	0.349	**0.000**	0.002	0.002	0.037	0.019	0.035	0.030	0.016
K_{arith}^{KS}	0.005	0.004	0.004	0.349	**0.000**	0.002	0.002	0.037	0.020	0.036	0.304	0.021
K_j	0.005	0.004	**0.003**	0.338	**0.000**	0.002	0.002	0.038	0.021	0.036	0.010	0.016
K_{AS}	0.005	0.004	0.004	0.339	0.001	0.002	0.002	0.039	0.026	0.036	0.030	0.030
K_{NHSL}	0.005	0.004	0.004	0.360	0.001	0.002	0.002	0.037	0.014	0.036	0.123	0.021
K_D	0.005	0.004	0.004	0.332	**0.000**	0.002	0.002	0.039	0.031	0.037	0.123	0.023
K_{KM2}	0.005	0.004	0.004	0.350	**0.000**	0.002	0.002	0.038	0.020	0.038	0.304	0.015
K_{KM12}	0.005	0.004	**0.003**	0.346	**0.000**	0.002	0.002	0.039	0.020	0.037	0.123	0.022

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 11) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = .01$) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية المختلفة ($\gamma = .7 , .9 , .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 100$.

4 – 4 – 4: قيم متوسط مربعات الخطأ عند $\sigma = 5$

جدول (4 – 12): قيم MSE عند استخدام عينة بحجم (n = 10) وانحراف معياري (σ = 5).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**12.66**	12.45	4.951	**8.230**	13.31	17.92	**11.94**	13.84	**25.11**	**31.65**	**11.60**	**29.85**
OLS	19.46	13.36	4.987	31.45	12.00	19.19	43.74	34.61	29.97	40.62	21.83	38.57
K_{HK}	18.72	13.29	4.956	20.14	5.297	18.50	14.61	**11.91**	28.64	34.34	15.39	43.62
K_{HKB}	18.72	13.29	4.916	20.14	4.340	18.40	14.61	**11.91**	27.27	34.34	15.39	42.62
K_{LW}	17.93	12.86	4.849	14.47	4.917	12.55	14.57	13.22	25.22	34.34	18.42	42.62
K_{HSL}	18.40	12.71	4.794	13.27	4.936	19.02	14.46	13.44	25.48	33.37	14.60	43.63
K_{MED}	18.29	12.62	4.715	9.830	4.936	18.60	14.34	13.54	25.44	35.32	15.42	43.72
K_{KS}	18.07	12.40	4.741	17.73	5.418	18.60	14.61	13.63	26.57	32.43	14.98	42.72
K_{arith}^{KS}	17.82	12.16	4.776	17.73	8.224	**12.11**	14.49	13.88	27.48	38.31	15.73	45.81
K_j	17.68	11.99	4.414	9.560	9.746	16.87	14.14	13.68	26.96	38.31	15.43	41.33
K_{AS}	17.59	11.94	3.800	10.44	4.356	13.90	15.15	14.03	28.29	36.30	14.24	44.70
K_{NHSL}	17.31	12.26	4.633	10.33	**4.337**	14.46	15.01	13.06	25.22	36.30	15.43	44.32
K_D	17.72	11.79	3.902	14.08	4.329	15.40	13.58	15.36	29.01	38.31	15.36	44.70
K_{KM2}	16.87	11.44	**3.754**	9.230	4.687	15.50	13.01	13.37	27.48	36.30	153.6	46.41
K_{KM12}	16.18	**11.34**	3.832	11.16	9.368	15.41	13.38	13.75	26.77	40.37	15.36	46.41

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 12) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = 25$) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية المختلفة ($\gamma = .7 , .9 , .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 10$.

جدول (4 – 13): قيم MSE عند استخدام عينة بحجم (n = 50) وانحراف معياري (σ = 5).

Method	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**12.23**	**4.380**	4.980	**3.372**	6.548	16.80	**13.40**	**12.79**	**14.41**	**14.11**	**14.44**	**27.21**
OLS	56.23	21.45	18.25	34.36	26.75	37.59	28.50	24.95	59.98	43.12	62.35	75.36
K_{HK}	17.84	7.487	5.289	11.91	4.659	16.69	15.35	13.27	18.67	23.21	24.38	23.94
K_{HKB}	17.84	7.487	5.261	11.91	4.659	16.74	15.35	13.27	17.57	23.21	24.38	23.94
K_{LW}	17.08	7.425	5.152	13.03	4.629	16.76	14.46	13.63	15.43	19.64	20.45	31.58
K_{HSL}	17.54	7.339	5.084	10.02	4.572	16.91	14.14	14.03	19.77	24.13	20.52	24.05
K_{MED}	17.43	7.246	5.019	9.170	4.367	17.13	16.65	13.68	19.51	19.64	21.08	27.85
K_{KS}	17.22	7.158	4.964	14.94	4.151	17.36	14.28	14.14	16.92	16.83	17.83	27.85
K_{arith}^{KS}	16.98	7.182	4.904	15.39	4.047	17.01	18.37	13.28	17.30	17.86	20.06	34.49
K_j	16.85	7.134	4.811	12.81	**3.935**	17.59	16.06	12.81	16.49	19.64	15.25	35.15
K_{AS}	16.68	7.030	4.654	7.460	3.956	**16.20**	15.57	13.48	15.88	20.31	19.76	27.85
K_{NHSL}	16.49	6.975	4.704	10.16	4.466	17.09	13.46	14.19	17.39	16.83	21.08	32.31
K_D	16.88	6.859	4.507	4.190	3.962	17.27	15.57	14.12	18.29	17.13	17.86	34.49
K_{KM2}	16.07	6.890	**4.158**	10.40	4.036	17.78	15.36	13.68	16.54	21.36	18.30	29.84
K_{KM12}	15.41	6.829	4.558	13.03	3.945	17.46	18.30	13.51	15.69	17.13	17.86	29.84

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 13) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = 25$) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية المختلفة ($\gamma = .7 , .9 , .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 50$.

جدول (4 – 14): قيم MSE عند استخدام عينة بحجم (n = 100) وانحراف معياري (σ = 5).

Method	**متغيرين**			**4 متغيرات**			**8 متغيرات**			**12 متغيرا**		
	.99	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**	**.99**	**.90**	**.70**
PCR	**11.67**	09.98	08.95	**5.379**	5.869	04.98	**04.02**	**11.08**	05.95	**08.13**	**05.17**	6.404
OLS	13.78	11.63	10.01	7.900	9.310	08.10	06.14	16.40	18.67	13.49	14.87	13.48
K_{HK}	12.61	11.30	09.64	6.900	0.215	05.10	05.16	11.73	05.09	08.44	06.98	04.76
K_{HKB}	12.61	11.30	09.77	6.900	0.215	04.97	05.16	12.73	04.21	08.44	06.98	04.74
K_{LW}	12.42	11.15	09.50	6.853	0.193	04.88	04.88	14.82	**03.69**	09.40	05.23	04.39
K_{HSL}	12.07	10.81	09.08	6.863	0.172	04.97	04.78	11.63	04.03	09.82	07.16	04.35
K_{MED}	12.27	10.55	08.91	6.745	0.014	04.67	04.63	11.28	04.31	08.69	05.94	**04.04**
K_{KS}	12.17	10.36	09.21	6.747	0.000	04.99	04.58	11.40	04.50	10.01	05.22	04.19
K_{arith}^{KS}	12.29	10.29	09.08	6.765	0.018	04.54	04.54	12.11	04.53	10.40	05.94	04.94
K_j	11.84	11.37	08.62	6.672	0.016	04.83	04.47	12.78	04.57	08.85	05.19	04.16
K_{AS}	11.88	10.12	09.43	6.656	0.183	04.94	04.36	11.56	04.91	09.69	05.22	04.99
K_{NHSL}	11.92	10.49	**08.30**	6.824	0.092	04.92	04.17	11.11	04.22	09.20	05.25	04.76
K_D	11.86	09.84	08.91	6.614	**0.013**	04.12	04.33	11.93	05.08	09.98	05.25	04.47
K_{KM2}	11.78	09.76	09.60	6.774	0.028	**04.07**	04.29	11.98	04.62	09.44	05.94	04.41
K_{KM12}	11.69	**09.53**	08.40	6.745	0.023	04.90	04.06	11.39	04.57	09.17	05.25	04.40

المصدر: الباحث بواسطة البرنامج.

توضح النتائج على الجدول (4 – 14) قيم MSE لنماذج PCR، ونماذج OLS، ونماذج RR المختلفة وفقا لطرق تقدير معلمة الحافة K ، حيث توضح أداء هذه المقدرات عند مستوى تباين قدره ($\sigma^2 = 25$) بين مشاهدات المتغيرات التفسيرية ومشاهدات متغير الاستجابة، وعند مستويات مختلفة للارتباط بين المتغيرات التفسيرية المختلفة ($\gamma = .7 , .9 , .99$)، وأعداد مختلفة من المتغيرات التفسيرية، وحجم عينة $n = 100$.

جدول (4 – 15): طرق مقدرات الحافة الأفضل أداء عند مستويات الارتباط والتباين وأحجام العينات وعدد المتغيرات المحددة

σ	n	متغيرين			4 متغيرات			8 متغيرات			12 متغيرا		
		0.99	0.90	0.70	0.99	0.90	0.70	0.99	0.90	0.70	0.99	0.90	0.70
0.5	10	K_{KM12}	K_{KM12}	K_{KM2}	K_{KM2}	K_{NHSL}	K_{AS}	K_D	K_{HSL}	K_{KM2}	K_{HK} K_{HKB} K_{LW}	K_{AS}	K_{AS}
	50	K_{KM12}	K_{KM12}	K_{KHB} K_{LW}	K_D	K_{KM12}	K_{AS}	K_{KM12}	K_{arith}^{KS}	K_{KM12}	K_{arith}^{KS}	K_{KM2}	K_{KS}
	100	K_{KM12}	K_{MED}	K_j	K_{AS}	K_{KM12}	K_{KM2}	K_{KM12}	K_{LW}	K_{LW}	K_j	K_{NHSL}	K_{KM2}
0.3	10، 50، 100	(K_{KM12} وK_{KM2} ،K_D وK_{NHSL} و K_{AS} و K_j و K_{KS} وK_{arith}^{KS})						(K_{HK} و K_{HKB} و K_{LW} و K_{HSL} و K_{MED})					
0.1	10، 50، 100	(K_{KM12} وK_{KM2} ،K_D وK_{NHSL} و K_{AS} و K_j و K_{KS} وK_{arith}^{KS})						(K_{HK} و K_{HKB} و K_{LW} و K_{HSL} و K_{MED})					
5	10، 50، 100	(K_{KM12} وK_{KM2} ،K_D وK_{NHSL} و K_{AS} و K_j و K_{arith}^{KS} وK_{KS})						(K_{HK} و K_{HKB} و K_{LW} و K_{HSL} و K_{MED})					

الجدول (4 – 15) يلخص مقدرات RR الأفضل أداءًا من خلال قيم MSE، وذلك وفقا لمستوى شدة التعدد الخطي، ومستوى التباين داخل المتغيرات التفسيرية، ومتغير الاستجابة، وعدد المشاهدات أو حجم العينة، وعدد المتغيرات التفسيرية المضمنة في النموذج.

4 – 5: مناقشة النتائج

تمت مناقشة النتائج تحت كل مستوى من مستويات التباين، مع مستويات مختلفة من شدة التعدد الخطي،أعداد مختلفة منللمتغيرات التفسيرية المضمنة في النموذج، ومن خلال المناقشة تم إجراء مقارنات مختلفة بين أداء طريقة OLSوPCRوRR، ومن جهة أخرى تمت المقارنة بين طرق بطرقه المختلفة RRوفقا لطريقة حساب معلمة الحافة K، وتم ربط النتائج بفروض الدراسة للتأكد من صحتها ، وكذلك تم ربط النتائج بما توصلت إليه الدراسات السابقة.

4 – 5 – 1: أداء المُقدرات كدالة في $5 . = \sigma$:

بدراسة النتائج على الجداول (4 – 3 ، 4 – 4 ، 4 – 5) نجد أن قيم MSEعند مستويات الارتباط المختلفة $(\gamma = .7, .9, .99)$تحت الأعمدة (متغيرات) تبين أن مقدرات PCR، ومعظم مقدرات RR قدمت أداءًا أفضل من أداء مقدرات OLS.تتفق هذه النتيجة مع ما توصلت إليه دراسة كل من Richard F. Gunstو Robert L. Mason. وتؤكد هذه النتيجة صحة فرضية البحث الأولى في أن مقدراتPCR وRR تقدم أداءًا أفضل من مقدرات OLS عند وقوع النموذج تحت تأثير مشكلة التعدد الخطي غير التام.

عند مقارنة أداء PCR بأداء مقدرات RRالمختلفة، الواردة تحت العمود (متغيرين) على الجداول الثلاثة عند مستوى ارتباط$(\gamma = .99)$،نجد أن PCR أفضل أداءًا من جميع مقدرات RRباستثناء مقدر K_{KM12}حيث كان أفضل أداءًا كما على الجدول (4 – 3). تحقق هذه النتيجة الفرضية الثانية في أن PCR تقدم أداءًا أفضل من أداء OLS وRR في حال وجود ارتباط تام، أو شبه تام بين المتغيرات التفسيرية.وتتفق هذه النتيجة مع دراسةهيثم يعقوب يوسف وآخرون حيث أوصت هذه الدراسة باستخدام PCR عندما يكون مستوى التعدد الخطي فوق 90%، بينما عند مستوى الارتباط $(\gamma = .9)$نجد أن قيمMSEلـ PCRتبين أنها ليست أفضل أداءًا منجميع مقدرات RR، وتنطبق هذه النتيجة أيضا على أداء PCRعند مستوى الارتباط$(\gamma = .7)$،وهذه النتيجة إلى حد ما تتفق مع ما توصلت إليه دراسة Yazid M. Al-Hassan و Mowafaq M. Al-Kassab.

وعند المقارنة بين طرق انحدار الحافة على الجدول (4 – 3) حيث حجم العينة (n = 10) نجد أن K_{KM12}الأفضل أداءًا يليهاK_{KM2}وK_{NHSL}عند مستوى الارتباط (γ = .99)، وعند مستوى الارتباط (γ = .9)أيضاK_{KM12} هي الأفضلأداءًا ويليها K_{KM2}وK_{D}ثمK_{AS}و K_j، وعند مستوى الارتباط(γ = .7)كانت أفضل طرقتقدير معلمة الحافة أداءًا هيK_{KM2}وتليها طريقة K_{KM12}و K_{AS}. وتوضح النتائج على الجدول (4 – 4) عند حجم العينة (n = 50) ومستوى الارتباط (γ = .99)أفضل طرق تقدير معلمة الحافة أداءًا هي K_{KM12} ويليها K_{KM2}ثمK_{D} ، وعند مستوى الارتباط (γ = .9)أيضا الأفضل أداءًا هي الطرق K_{KM12} و K_{KM2}وK_{D} وتليها K_{NHSL}ثم K_{NHSL}.تتفق هذه النتائج مع ماتوصلت إليه دراسة Gisela Muniz وآخرون ودراسة Yazid M. Al-Hassan، وعند مستوى الارتباط(γ = .7)كانت الطرق الأفضل أداءًا هي K_{KHB}وK_{LW}وتليها K_{D}ثم K_{KM12}أيضا تتفق هذه النتيجة مع ما توصلت إليه دراسة Ghadban Khalaf. بينما النتائج الواردة بالجدول (4 – 5) حيث تم استخدام عينات بحجم (n = 100)، فعند مستوى الارتباط (γ = .99)كانت أفضل طرق تقدير معلمة الحافة أداءًا هي K_{KM12}وتليهاK_{KM2}ثم K_{D} ، وعند مستوى الارتباط (γ = .9)كانت الطريقة الأفضل أداءًاK_{MED}وتليها K_{KM2}ثم K_{D}، وعند مستوى الارتباط (γ = .7)أظهرت الطريقة K_j أداءًا أفضل وتلتها K_{HSL}وK_{MED}ثم تلتها K_{NHSL}.

النتائج الواردة تحت العمود (4 متغيرات) بالجداول الثلاثة توضح قيم MSEعند مستويات الارتباط المختلفة (γ = .7 ,.9 ,.99)، وعند مقارنة أداء PCR بأداء مقدرات RRالمختلفة عند مستوى ارتباط(γ = .99)نجد أن جميع مقدرات PCR كانت أفضل أداءًا من مقدرات RR عند حجم عينة (n = 50, 100)، بينما أظهرت مقدرات RR أداءًا أفضل عند حجم عينة (n = 10). تتفق هذه النتائج مع ما توصلت إليه دراسة هيثم يعقوب يوسف وآخرون. وعند مستوى الارتباط (γ = .9)جميع طرق مقدرات RRقدمت أداءًا أفضل من أداء مقدرPCRعند (n = 10, 50) باستثناء K_j . أيضا تتفق هذه النتائج مع ما توصلت إليه دراسة Gisela Muniz وآخرون، أما عند

($n = 100$) فكانت طرق PCR هي الأفضل أداءًا.وعند مستوى الارتباط($\gamma = .7$)كانت طرق RRالأفضل عند ($n = 10, 100$) بينما كانت الأفضلية لطرق PCR عند ($n = 50$).

وعند المقارنة بين طرق انحدار الحافة على الجدول (4 – 3) حيث ($n = 10$)نجد أن K_{KM2}الأفضل أداءًاتليهاK_jثمK_{NHSL}عند مستوى الارتباط ($\gamma = .99$) ، وعند مستوى الارتباط ($\gamma = .9$) أفضلها أداءًا K_{NHSL}تليها K_{LW}وK_{HSL}وK_{MED}ثم K_{AS}، وعند مستوى الارتباط= (γ .7)كانت أفضل طرقتقدير معلمة الحافة هيK_{AS}وتليها طريقة K_{HK} و K_{HKB}ثم K_{MED}. وعلى الجدول (4 – 4) وعند ($\gamma = .99$)الأفضل أداء ًا هي طريقةK_Dتليها K_{AS}ثم K_{HK}وK_{HKB}، وعند ($\gamma = .9$)قدمت الطريقةK_{KM12}أداءًاأ فضلتليها الطريقة K_D ثم K^{KS}_{arith} و K_j. وعند ($\gamma =$ .7)أفضلها أداءًاK_{AS}تيليها K_{HK}ثم K_{HKB}و K_{LW}. بينما على الجدول (4 – 5) وعند مستوى ارتباط ($\gamma = .99$)كانت الطريقةK_Dالأفضل أداءًا تليها الطريقتان K_j و K_{AS}ثم الطريقتان K_{MED}و K_{KM12}، وعند مستوى ارتباط ($\gamma = .9$)برزتK_{KM12}كأفضل الطريقة وتلتها الطريقةK^{KS}_{arith}ثم الطريقة K_j، وعند مستوى ارتباط($\gamma = .7$)أفضل الطرق هي K_{KM2}وتليهاK_D ثم K^{KS}_{arith}، تتفق هذه النتائج ما توصلت إليه دراسة كل من دراسة Gisela Muniz وآخرون، ودراسة Yazid M. Al-Hassan.

وبدراسة النتائج الواردة تحت العمود (8 متغيرات)على الجداول الثلاثة لقيم MSEتحت جميع مستويات الارتباط ($\gamma = .7, .9, .99$)، توضح مقارنة أداء PCR بأداء مقدرات RRالمختلفة عند مستوى ارتباط($\gamma = .99$)، نجد أن جميع مقدرات RR كانت أفضل أداءًا من مقدرات PCR عند حجم عينة ($n = 10$)وهذه النتيجة تتفق مع ما توصلت إليه دراسة Yazid M. Al-Hassan و Mowafaq M. Al-Kassab ، أما عند حجم عينة (100و 50 = n) أظهرت مقدرات PCR أداءًا أفضل عند مستوى ارتباط ($\gamma = .7, 90$)وكذلك تتفق هذه النتائج مع ما توصلت إليه دراسة مع ما توصلت إليه دراسة هيثم يعقوب يوسف وآخرون.

وللمقارنة بين طرق انحدار الحافة على الجدول (4 – 3) وحجم عينة $(n = 10)$أظهرت طريقة K_Dأداءًاأفضل وتليهاK_{KM2}ثمK_{KM12}عند مستوى الارتباط $(\gamma = .99)$،وهذه النتيجة تتفق مع دراسة A. V.Dorugade و D. N. Kashidفي أنK_Dلها أفضل أداء، وعند مستوى الارتباط $(\gamma = .9)$كانت الطريقةK_{HSL}تليها K_{MED}ثم K_{arith}^{KS} تتفق هذه النتيجة مع ما توصلت إليه دراسة Yazid M. Al-Hassanفي أن K_{HSL} كانت لها أفضل أداءًا، وعند مستوى الارتباط$= \gamma)$ $.7)$كانت أفضل الطرق أداءًاهيK_{KM2}تليها طريقة K_{AS}وK_{KM12}ثم K_D. وعلى الجدول (4 – 4) وحجم عينة $(n = 50)$عند مستوى ارتباط $(\gamma = .99)$كانتالطريقة الأفضل أداءًا هي طريقةK_{KM12}تليها K_{arith}^{KS}ثم K_{NHSL}، وعند $(\gamma = .9)$قدمت الطريقةK_{arith}^{KS}أفضل أداءًا تليها الطريقةK_{KS} ثم K_j. وعند $(\gamma = .7)$أفضلها أداءK_{KM12}تيليها K_{KM2}ثم K_{NHSL}. في حين على الجدول (4 – 5) وحجم عينة $(n = 100)$وعند مستوى الارتباط$(\gamma = .99)$كانت الطريقة الأفضل هي K_{KM12}تليها الطريقتان K_{NHSL} ثم الطرق K_{AS}وK_Dو K_{KM2}، وعند مستوى الارتباط $(\gamma = .9)$أظهرت الطريقةK_{LW}أداءًا أفضل وتلتها الطريقة K_{HSL}ثمK_{arith}^{KS}وK_{NHSL}، وعند مستوى الارتباط $(\gamma = .7)$ أفضل الطرق أداء K_{LW}وتلتها K_{NHSL}ثم K_{HKB}كذلك تتفق هذه النتائج مع توصلت إليه دراسةGisela Muniz وآخرون.

النتائج الواردة تحت العمود (12 متغيرا)على الجداول الثلاثة توضح قيم MSEعند مستويات الارتباط $(\gamma = .99, .9, .7)$، وعند مقارنة أداء PCR بأداء مقدرات RRالمختلفة عند مستوى ارتباط$(\gamma = .99)$ نجد أن مقدراتPCR كانت أفضل أداءًا من مقدرات RR عند حجم عينة $(n = 10, 100)$وتتفق هذه النتائج مع ما توصلت إليه دراسة هيثم يعقوب يوسف وآخرون، أما عند حجم عينة $(n = 50)$ فأظهرت مقدرات RR أداءًا أفضل، وعند مستوى ارتباط $(\gamma = .7, 90)$أظهرت مقدرات RR أداءًا أفضل وهذه النتيجة إلى حد ما تتفق مع ما توصلت إليه دراسة Yazid M. Al-HassanوMowafaq M. Al-Kassab ودراسة Norliza Adnan وآخرون.

عند مقارنة طرق RRعلى الجدول (4 – 3) وحجم عينة (n = 10) نجد أن الطرق K_{HK} و K_{HKB}وK_{LW} هي الأفضل وتليهاK_{HSL}ثمK_{KS}عند مستوى الارتباط (γ = .99) وهنا تتفق هذه النتيجة مع دراسة Yazid M. Al-Hassan حيث أظهر K_{HKB} أداءًا أفضل من المقدرات الأخرى، وعند مستوى الارتباط (γ = .9)كانت الطريقةK_{AS}تليها K_{HSL}ثمK_{KS}، وعند مستوى الارتباط(γ = .7)كانت أفضل الطرق أداءًاهيK_{KS}تليها K_{HSL}ثم K_{MED} هذه النتائج كذلك تتفق مع توصلت إلية دراسة Mahdi A. AlkhamisiوGhazi Shukur. وتوضح النتائج الواردة بالجدول (4 – 4)وحجم عينة (n = 50)عند (γ = .99)أن الطريقةالأفضل أداءًا هيK_{arith}^{KS}تليها K_{LW}وK_{MED} و K_j ثم K_DوK_{KM12}، وعند (γ = .9)قدمت الطريقةK_{KM2}أداءًاأفضل وتليهاK_{KS} ثم K_DوK_{KM12}. وعند (γ = .7)كانتأفضل الطرق أداءًاK_{KS}وK_{AS} تليها K_{MED}ثم K_{KM2}وK_{KM12}. أما على الجدول (4 – 5) وحجم عينة (n = 100) وعند مستوى الارتباط (γ = .99)كانت الطريقة الأفضل أداءًا هي K_jتليها الطريقتانK_{MED} و K_{AS} ثم الطريقة K_{KS}، وعند مستوى الارتباط (γ = .9)أظهرتK_{NHSL}وK_D وK_{KM12}الأداء الأفضل وتلتها K_{LW}ثمK_{HSL}وK_j، وعند مستوى الارتباط (γ = .7)كانتأفضل الطرق أداءًا هي K_{KM2}وتلتها K_{HKB}ثم K_{KS}. وكذلك تتفق معظم هذه النتائج مع ما توصلت إليه دراسة Yazid M. Al-Hassan و Mowafaq M. Al-Kassab ، ودراسة Yazid M. Al-Hassan ، ودراسة Gisela Muniz وآخرون.

4 – 5 – 2: أداء المُقدرات كدالة في 3. = σ :

عند دراسة النتائج على الجداول (4 – 6 ، 4 – 7 ، 4 – 8) تحت جميع الأعمدة (متغيرات)، يلاحظ أن قيم MSEعند مستويات الارتباط المختلفة (γ = .7 ,.9 ,.99) تبين أن مقدرات PCR، وجميع مقدرات RRحازت على الأفضل مقارنة بأداء OLS، وتتفق هذه النتيجة كذلك مع ما توصلت إليه دراسة G. R. Pasha وMuhammad A. Ali Shah. وأيضا هذه النتيجة تثبت صحة الفرضية الأولى في أن أداء PCRوأداء طرق RR أفضل من أداء OLSعند بروز مشكلة التعدد الخطي.

أوضحت المقارنة بين مقدرات PCR، ومقدرات انحدار الحافة RRالمختلفة تحت الأعمدة (متغيرات) ضمن الجداول الثلاثة عند مستوى ارتباط$(\gamma = .99)$،أوضحت أن أداء مقدرات PCR أفضل من أداء جميع مقدرات RR، في حين عند مستوى الارتباط $(\gamma = .9)$ أظهرت معظم مقدرات RR أداءًا أفضل من أداءمقدرات PCR، وعند مستوى الارتباط $(\gamma = .7)$ أيضا أظهرت معظم مقدرات RRأداءًا أفضل مقارنةببعض مقدرات PCR.

كذلك أوضحت دراسة المقارنة بين مقدرات انحدار الحافة المختلفة وفقا لطرق تقدير عامل التحيز أو معلمة الحافة K أن معظم المقدرات كانت لها مستويات متقاربة من الأداء وإجمالاً نجد أن الطرق مثل $(K_{KM12}$و K_{KM2}، K_Dو K_{NHSL} و K_{AS} و K_j و K_{arith}^{KS}و$K_{KS})$ قدمت أداءًا أفضل مع مختلف أحجام العينات، وعدد المتغيرات التفسيرية، وعند كل مستويات الارتباط أو شدة التعدد الخطي. في حين أبدت الطرق $(K_{HK}$ و K_{HKB} و K_{LW} و K_{HSL} و $K_{MED})$ أداءًا أفضل عندما يكونعدد المتغيرات التفسيرية كبيراً (ثمانية متغيرات واثني عشر متغير)، مع اختلاف أحجام العينات، ومستويات الارتباط أو شدة التعدد الخطي.

4 – 5 – 3: أداء المُقدرات كدالة في $\sigma = .1$:

النتائج الواردة بالجداول (4 – 9 ، 4 – 10 ، 4 – 11) تحت جميع الأعمدة (متغيرات) توضح قيم MSEعند مستويات الارتباط المختلفة $(\gamma = .7 , .9 , .99)$،حيث يلاحظ أن مقدرات PCR، وجميع مقدرات RR كانت الأفضل عند مقارنة أدائها بأداء مقدرات OLS. كذلك توضح هذه النتيجة صحة الفرضية الأولى، في أن طرق RRلها أداءًا أفضل من أداء OLSعندبروز مشكلة التعدد الخطي.

أوضحت المقارنة بين مقدرات PCR، ومقدرات RRتحت الأعمدة (متغيرات)، ضمن الجداول الثلاثة عند مستوى ارتباط$(\gamma = .99)$،أن أداء مقدرات PCR أفضل من أداء بعض مقدرات RR، وتوضح هذه النتيجة صحة الفرضية الثانية القائلة بأفضلية أداء طريقة PCR مقابل طرق RR في حالة التعدد الخطي التام أو شبه التام، وعند مستوى الارتباط $(\gamma = .9)$ أظهرت

معظم مقدرات RR أداءًا أفضل من أداءمقدرات PCR، وعند مستوى الارتباط $(\gamma = .7)$ كذلك كانت معظم مقدرات RRذات أداء أفضل من أداء بعض مقدرات PCR.

عند المقارنة بين مقدرات RR المختلفة وفقا لطرق تقدير عامل التحيز أو معلمة الحافة K، نجد أن معظم المقدرات تحت عمود (2 متغيرين) كان أدائها يكاد يكون متطابق، وبصورة إجمالية يمكن القول إن الطرق (K_{KM12}و K_{KM2}، K_Dو K_{NHSL} و K_{AS} و K_j و K_{arith}^{KS}و K_{KS}) لها أداء أفضل مع مختلف أحجام العينات، وعدد المتغيرات التفسيرية، وعند كل مستويات الارتباط أو شدة التعدد الخطي. بينما الطرق (K_{HK} و K_{HKB} و K_{LW} و K_{HSL} و K_{MED}) كانت أفضل أداءًا مع تزايد عدد المتغيرات التفسيرية (ثمانية متغيرات واثني عشر متغيرا)،وعند أحجام العينات المختلفة، وكل مستويات الارتباط أو شدة التعدد الخطي.

4 – 5 – 4: أداء المُقدرات كدالة في $\sigma = 5$:

الجداول (4 – 12 ، 4 – 13 ، 4 – 14) توضح قيم MSEعند مستويات الارتباط المختلفة $(\gamma = .7, .9, .99)$، وتحت جميع الأعمدة (متغيرات)، عند دراسة هذه النتائج نجد أن مقدرات PCR، وجميع مقدرات RRكانت الأفضل مقارنةبمقدرات OLS. وكذلك تتفق هذه النتيجة مع فرضية البحث الأولى في أن أداء طرق RRأفضل من أداءOLSعند وجود حالة التعدد الخطي.

عند المقارنة بين مقدرات PCR، ومقدرات RR، تحت الأعمدة (متغيرات)، ضمن الجداول الثلاثة عند مستوى ارتباط$(\gamma = .99)$، نجدأن أداء مقدرات PCR أفضل من أداء جميع مقدرات RR، وتوضح هذه النتيجة صحة الفرضية الثانية بأن أداء طريقة PCR أفضل من أداء طرق RR في حالة التعدد الخطي التام أو شبه التام. وعند مستوى الارتباط $(\gamma = .9)$نجد أن معظم مقدرات RRكانت لها الأداء الأفضلمقارنة بـ PCR،باستثناء أدائها عند تزايد عدد المتغيرات التفسيرية حيث قدمت أداء أفضل من RR، وكذلك عند مستوى الارتباط $(\gamma = .7)$ معظم مقدرات RRكان لها أداء أفضل من أداء بعض مقدرات PCR، باستثناء أدائها عند تزايد عدد المتغيرات التفسيرية.

عند إجراء المقارنة بين مقدرات انحدار الحافة المختلفة بصورة إجمالية، نجد أن الطرق (K_{KM12}وK_{KM2}،K_DوK_{NHSL} و K_{AS} و K_j و K_{arith}^{KS}وK_{KS}) لها أداء أفضل مع مختلف أحجام العينات، وعدد المتغيرات التفسيرية، وعند كل مستويات الارتباط ، وبصفة خاصة عند وجود متغيرين تفسيريين أو عند استخدام أربعة متغيرات تفسيرية. في حين قدمت الطرق (K_{HK} و K_{HKB} و K_{LW} و K_{HSL} و K_{MED}) أداءًا أفضل عند زيادة عدد المتغيرات التفسيرية (ثمانية متغيرات واثني عشر متغيراً)،ومع أحجام العينات المختلفة، وكل مستويات شدة التعدد الخطي.

4 – 5 – 5: أداء المُقدرات كدالة في التشتت σ:

هنا تنصب الدراسة حول تأثير مستويات التباين المختلفة،وبصورة عامة في أداء الطرق المختلفة PCR و RR و OLS، وبصورة أخص تأثير التباين بين قيم المتغيرات على أداء المقدرات،حيث تم اتخاذ قيم متوسط مربعات الخطأ MSEكمعيار للأداء. يلاحظ على الأشكل البيانية (1 – أ)، (1 – ب)، (1 – ج) الواردة بالملحق (1)، أن طريقة PCR، ومقدرات طرق RR المختلفة، وطريقة OLSيتقارب أدائها تبعا لتناقص مستويات التباين داخل المتغيرات التفسيريةوتثبت هذه النتيجة صحة الفرضية الثالثة للبحث القائلة بتقارب أداء المقدرات الثلاث كلما قل مستوى التابين داخل المتغيرات التفسيرية. يلاحظ أيضا أن قيمMSE تتزايد تبعا لتزايد مستوى التباين بين مشاهدات المتغيرات التفسيرية والعكس صحيحا، وهذه النتيجة أيضا تؤيد ما ذهبت إليه الفرضية الرابعة للبحث التي تدعي تزايد قيم MSE مع تزايد مستوى التابين داخل المتغيرات التفسيرية.

4 – 5 – 6: أداء المُقدرات كدالة في حجم العينة n:

لدراسة أداء مُقدرات PCR و RR و OLS كدالة في حجم العينة تم اعداد أشكال بيانية لقيم MSE كدالة في أحجام العينات، مع ثبات شدة التعدد الخطي (γ = 0.99)، وعند مستويات مختلفة من التباين تمثلها الانحرافات المعيارية (σ = 5 ,.5 ,.3 ,.1) . ويلاحظ على الأشكل البيانية (2 – أ)، (2 – ب)، (2 – ج)، (2 – د) الواردة بالملحق (2)، أن طريقة PCR ومقدرات طرق RR المختلفة لها أداء شبه متجانس عند مختلف أحجام العينات،ويلاحظ أن حجم العينة يلعب دوراً مؤثراً في أداء طرق PCR وRRو OLSحيث تبين الرسوم البيانية إلى حد ما كلما زاد عدد

المشاهدات،كلما قدمت هذه الطرق أداءًا أفضل، أي كلما زاد عدد المشاهدات صغرت قيم MSE، وهذه النتيجة تتفق مع فرضية البحث الخامسة القائلة بأن أداء طريقة PCR، وطرق RR وطريقة OLS تتأثر بعدد المشاهدات عند تعرض النموذج لمشكلة التعدد الخطي.

4 – 5 – 7: أداء المُقدرات كدالة في مستوى شدة التعدد الخطي

الأشكال البيانية الواردة بالملحق الثاني (3– أ)، (3– ب)، (3– ج)، (3– د) توضح أداء مُقدرات PCR وRR وOLS كدالة في مستويات معامل الارتباط المختلفة أو مستويات شدة التعدد الخطي المختلفة. عند مستويات مختلفة من التباين تمثلها الانحرافات المعيارية 1. , 3. , 5. , 5 = σ. يلاحظ تراجع أداء طريقة OLSعندما يقارن بأداء طريقةPCR وأداء مقدرات طرقRR عند مختلف عدد المتغيرات ومستويات التباين، فيما لوحظ أن أداء طريقة PCR كان أكثر إنسجاما مع طرق RR. أيضا اتضح من الرسوم البيانية أنهكلما زادت شدة التعدد الخطي زادت قيمة MSEتبعا لها بغض النظر عن مستوى التباين أو حجم العينة وذلك مع جميع الطرق المستخدمة OLS وPCR وRR، وتؤيد هذه النتيجة الفرضية السادسة القائلة بتزايد قيم MSEمع تزايد شدة التعدد الخطي.

الفصل الخامس

النتائج والتوصيات

النتائج والتوصيات

5 – 1: تمهيد

يتنــاول هــذا الفصــل أهــم مــا توصــل إليــه البحــث مــن نتــائج تتعلــق باســتخدام طريقــة انحــدار المكونــات الرئيســية، وطــرق انحــدار الحافــة المختلفــة، فــي حــال تعــرض نمــوذج الانحــدار الخطــي إلــى مشــكلة التعــدد الخطــي. كمــا يتنــاول هــذا الفصــل أيضــا توصــيات البحــث في ضوء الإستنتاجات.

5 – 2: النتائج

توصل البحث إلى عدد من النتائج نستعرضها فيما يلي:

1. قدمت كل من مُقدرات طريقة PCR، وRR أداء أفضل من أداء طريقة OLS، عند وقوع النموذج تحت تأثير مشكلة التعدد الخطي.
2. قدمت طريقة المكونات الرئيسية PCR أداءًا أفضل من أداء طريقةOLS وRR، في حال وجود ارتباط خطي تام أو شبه تام بين المتغيرات التفسيرية.
3. عند مستويات شدة التعدد الخطي $(\gamma = .9\ ,\ .7)$ قدمت مُقدرات RR المختلفة أداء أفضل من أداء طريقة PCR.
4. تتزايد قيم MSE تبعا لتزايد التباين في المتغيرات التفسيرية والعكس صحيحا، وذلك في ظل تأثر النموذج بمشكلة التعدد الخطي.
5. يقترب أداء طريقة OLSمن أداء كل من طريقة PCR، وطرق RR، كلما قل التباين داخل المتغيرات التفسيرية، أي أن أداء طريقة PCR، وطرق RRو OLSتتأثر بالتباين في ظل تأثر النموذج بمشكلة التعدد الخطي.
6. كلما زادت شدة التعدد الخطي زادت قيمة MSE تبعا لها بغض النظر عن مستوى التباين، أو حجم العينة، وذلك مع جميع الطرق المستخدمة OLS وPCR وRR.

7. توصل البحث إلى أن حجم العينة يلعب دوراً مؤثراً في أداء RRوPCR حيث كلما زاد عدد المشاهدات كلما قدمت PCR وRRو OLSأداءًا أفضل، أي كلما زاد عدد المشاهدات كلما صغرت قيم MSE.
8. حازت طريقة تقدير معلمة الحافة $\mathrm{K}_{\mathrm{KM12}}$ على الأداء الأفضل بين طرق تقدير معلمة الحافة، وتليها$\mathrm{K}_{\mathrm{KHB}}$ و K_{LW} وK_{KM2} وK_{MED} وK_j،عند جميع مستويات التعدد الخطي، ومستوى التشتت(σ = .5)، وعند تضمين متغيرين تفسيريين.
9. أحرزت طريقة تقدير معلمة الحافة K_{AS} أفضل أداء وتليها $\mathrm{K}_{\mathrm{KM12}}$ و $\mathrm{K}_{\mathrm{KM2}}$ و K_{NHSL} وK_D، وذلك عند جميع مستويات التعدد الخطي، ومستوى التشتت (σ = .5)، وعند تضمين أربعة متغيرات تفسيرية.
10. أظهرت طريقة تقدير معلمة الحافة $\mathrm{K}_{\mathrm{KM12}}$ أفضل أداء تليها طريقة K_{LW} و K_D و K_{HSL} و$\mathrm{K}_{\mathrm{KM2}}$ وK_{arith}^{KS} ، وذلك ضمن مستويات التعدد الخطي المختلفة، ومستوى التشتت (σ = .5)، وعند تضمين النموذج ثمانية متغيرات تفسيرية.
11. قدمت طريقة تقدير معلمة الحافة $\mathrm{K}_{\mathrm{KM2}}$ وK_{AS} أفضل أداء تليهما طريقة K_{HK} وK_{HKB} و K_{LW} و K_D و K_{NHSL} وK_{KS} وK_{arith}^{KS} وK_j، وذلك ضمن مستويات التعدد الخطي المختلفة، ومستوى التشتت (σ = .5)، وعند تضمين النموذج إثني عشر متغيرات تفسيرية.
12. عند مستويات التشتت (σ = 0.3 , 0.1 , 5)، وضمن مستويات التعدد الخطي المختلفة أظهرت الطرق K_{KM12} وK_{KM2} ، K_D و K_{NHSL} وK_{AS} وK_j و K_{arith}^{KS} وK_{KS}أفضل أداء، وذلك عند تضمين النموذج متغيرين تفسيرين، أو أربعة متغيرات تفسيرية، بينما أظهرت الطرق K_{HK} وK_{HKB} وK_{LW} وK_{HSL} وK_{MED} في حال تضمين النموذج ثمانية أو إثني عشر متغيرات تفسيرية.

5 – 3: التوصيات.

في ضوء النتائج التي توصل إليها الباحث من خلال الدراسة النظرية والتطبيقية خرجت الدراسة بعدد من التوصيات هي:

1. ضرورة استخدام انحدار المكونات الرئيسية PCR في حال وقوع النموذج تحت تأثير التعدد الخطي التام أو شبه التام.
2. ضرورة استخدام نحدار الحافة RR عندما يكون النموذج واقعاً تحت تأثير مشكلة التعدد الخطي العالي بحيث يكون الارتباط في مستوى لا يزيد عن 90% $(\gamma \leq 0.9)$.
3. يفضل استخدام طرق انحدار الحافة K_{KM12} و K_{KM2} ، K_D و K_{NHSL} و K_{AS} و K_j و K_{arith}^{KS} و K_{KS} عندما يكون عدد المتغيرات التفسيرية في حدود أربعة متغيرات، وإستخدام الطرق K_{HK} و K_{HKB} و K_{LW} و K_{HSL} و K_{MED} في حال زاد عدد المتغيرات التفسيرية عن أربعة متغيرات.
4. يوصي الباحث باجراء بحوث أخرى لدراسة PCR و RR و OLS ، والعلاقة بين مستوى تباين المتغيرات التفسيرية وقيم MSE.
5. ضرورة اجراء بحوث أخرى لدراسة PCR و RR و OLS ، والعلاقة بين حجم العينة وقيمة MSE.
6. ضرورة إجراء بحوث أخرى لدراسة PCR و RR و OLS ، والعلاقة بين مستويات أخرى لشدة التعدد الخطي وقيم MSE.
7. يوصي الباحث باجراء بحوث أخرى لدراسة العلاقة بين أداء الطرق PCR و RR و OLS ، ومستوى التباين داخل المتغيرات التفسيرية.
8. يوصي الباحث بإجراء بحوث أخرى حول استخدام RR في تخطي تأثيرات التعدد الخطي في نماذج الانحدار السوقي المتعدد Multinomial Logistic Regression، والانحدار غير الخطي المتعدد Multiple Nonlinear Regression.
9. يوصي الباحث باجراء بحوث أخرى لدراسة العلاقة بين عدد المتغيرات التفسيرية المضمنة بالنموذج وأداء بعض طرق RR عند مستويات مختلفة من شدة التعدد الخطي.

المراجع العربية:

[1] أبو عمه، محمد عبد الرحمن (مترجم)، تأليف بريان ج مانلي،الأساسفيالطرقالاحصائيةالمتعددةالمتغيرات، 2001م،الرياض،السعودية: النشر العلميوالمطابع،جامعةالملكسعود.

[2] إسماعيل،محمد عبد الرحمن،تحليلالانحدارالخطي2001م،الرياضالمملكةالعربيةالسعودية: معهدالادارةالعامة.

[3] سرور، إبراهيم سرور (مترجم) تأليف زافين أ. كاريان وإدوارد ج. دوديويز، المحاكاةالاحصائيةالحديثةومحاكاةالنظمونظامالمحاكاةمتعددةالأغراض، 1999م،الرياض،السعودية: النشرالعلميوالمطابع،جامعةالملكسعود.

[4] سودان، محمد عال (مترجم) تأليف جلبرت سترنج،الجبرالخطيوتطبيقاته2000م،الرياض،السعودية: النشرالعلميوالمطابع،جامعةالملكسعود.

[5] عناني، محمد عبد السميع،التحليلالقياسيوالاحصائيللعلاقاتالإقتصادية : مدخلحديثباستخدامSPSSطبعة أولى2009م ، الاسكندرية: الدارالجامعية.

[6] كنجو، أنيس اسماعيل وآخرون (مترجم)،نماذج احصائية تطبيقية: انحدار، تحليل تباين وتصاميم تجريبية الجزء الأول، 2000م، الرياض: النشر العلمي والمطابع جامعة الملك سعود.

[7] النشواتي، هشام مصطفى وآخرون،التحليلالاحصائيمتعددالمتغيراتباستخدامحزمساس2011م،الرياض،السعودية: النشرالعلميوالمطابع،جامعةالملكسعود.

[8] يحيى، مزاحم محمد،استخدامالمكوناتالرئيسيةوانحدارالحرففيتقديرمعادلةالسعرالعالميللقمحللفترةمن (1961-2002). مجلةتكريتللعلومالاداريةوالاقتصادية، 2005م، 1(1): ص 146 – 156.

[9] يوسف، هيثم يعقوب وآخرون، استخدامالأساليبالاحصائيةفيمعالجةمشكلةالتعددالخطي،مجلةديالىللعلومالزراعية، 2010م، 2(2): ص 162 – 176.

المراجع الإنجليزية:

[10] Abdallah, H.Y., *A Simulation Study of Ridge Regression Method with Autocorrelated Errors.* Shendi University Journal, 2009(6): p. 1 - 19.

[11] Acock, Alan C. "SAS, Stata, SPSS: A Comparison". Journal of Marriage and Family, 2005. 67 (4): 1093–1095.

[12] Adnan, N., M.H. Ahmad, and R. Adnan, *A comparative study on some methods for handling multicollinearity problems.* Matematika, 2006. **22**(2): p. 109-119.

[13] Alabi, O., K. Ayinde, and T. Olatayo, *Effect of multicollinearity onpower rates of the ordinary least squares estimators.* Journal of Mathematics and Statistics, 2008. **4**(2): p. 75-80.

[14] Al-Hassan, Y.M.,Al-Kassab, M. M.,*A Monte Carlo Comparison between Ridge and Principal Components Regression Methods.* Applied Mathematical Sciences, 2009. **3**(42): p. 2085-2098.

[15] Al-Hassan, Y.M., *A Monte Carlo Evaluation of Some Ridge Estimators.* Jordan Jornal for Applied Sciences, 2008. **10**(2): p. 101-110.

[16] Al-Hassan, Y.M., *Performance of a New Ridge Regression Estimator.* Journal of The Association of Arab Universities for Basic and Applied Sciences, 2010. **9** :p. pp. 43 - 50.

[17] Alheety, M.I. and B.G. Kibria, *On the Liu and almost unbiased Liu estimators in the presence of multicollinearity with heteroscedastic or correlated errors.* Surveys in Mathematics and its Applications, 2009. **4**: p. 155-167.

[18] Alin, A., *Multicollinearity.* Wiley Interdisciplinary Reviews: Computational Statistics, 2010. **2**(3): p. 370-374.

[19] Alkhamisi, M.A. and G. Shukur, *A Monte Carlo Study of Recent Parameters.* Communication in Statistics - Simulation and Computation, May 2007. **36**(3): p. pp. 535-547.

[20] Al-Nueimy, A.M.T., *Mechanism of Missing Data and Estimating them by Principal Component Regression.* Paper Presented on Second Conference for Mathematics and Statistics and Informatics Dec. , 2009: p. pp. 312 - 322.

[21] An, S., W. Liu, and S. Venkatesh, *Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression.* Pattern Recognition, 2007. **40**(8): p. 2154-2162.

[22] Arnold, S.F., *The Theory of Linear Models & Multivariate Analysis*. Wiley Series in Probability and Mathematical Statistics1981, New York: John Wiley & Sons, Inc.

[23] Askin, R.G., *Multicollinearity in regression: review and examples.* Journal of Forecasting, 1982. **1**(3): p. 281-292.

[24] Bashtian, M.H., M. Arashi, and S. Tabatabaey, *Using improved estimation strategies to combat multicollinearity.* Journal of Statistical Computation and Simulation, 2011. **81**(12): p. 1773-1797.

[25] Baye, M.R. and D.F. Parker, *Combining ridge and principal component regression: a moneydemand illustration.* Communications in Statistics-Theory and Methods, 1984. **13**(2): p. 197-205.

[26] Belsley, D., *Multicollinearity: diagnosing its presence and assessing the potential damage it causes least squares estimation.* NBER Working Paper, 1976.

[27] Belsley, D.A., E. Kuh, and R.E. Welsch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley Series in Probability and Mathematical 1980, USA: John Wiley & Sons, Inc.

[28] Birkes, D. and Y. Dodge, *Alternative Methods of Regression*1993, New York: John Wiley & Sons, Inc.

[29] Budden, M. and P.H.a.L. Hoffman, *On The Generation of Correlation Matrices.* Applied Mathematics E-Note, 2008. **8**: p. pp. 279 - 282.

[30] Chatterjee, S. and A.S. Hadi, *Regression Analysisby Example*. 4th ed 2006 Hoboken, New Jersey John Wiley & Sons, Inc.

[31] De Veaux, R.D. and L.H. Ungar, *Multicollinearity: A tale of two nonparametric regressions*, in *Selecting Models from Data*1994, Springer. p. 393-402.

[32] Der, G. and B.S. Everitt, *A Handbook of Statistical Analyses Using SAS* 2002 Boca Raton London: Chapman and Hall/CRC.

[33] Dorugade, A.V. and D.N. Kashid, *Alternative Method for Choosing Ridge Parameter for Regression.* Applied Mathematical Sciences, 2010. **4**(9): p. pp. 447 - 456.

[34] Ehsanes Saleh, A.M. and B. Golam Kibria, *Performance of some new preliminary test ridge regression estimators and their properties.* Communications in Statistics-Theory and Methods, 1993. **22**(10): p. 2747-2764.

[35] Encyclopedia of Research Design Encyclopedia of research design. 2010.http://en.wikipedia.org/wiki/SAS_(software), on 13/08/2014.

[36] Fabrycy, M.Z., Multicollinearity caused by specification errors. Applied Statistics, 1975: p. 250-254

[37] Farkas, O. and K. Héberger, *Comparison of ridge regression, partial least-squares, pairwise correlation, forward-and best subset selection methods for prediction of retention indices for aliphatic alcohols.* Journal of chemical information and modeling, 2005. **45**(2): p. 339-346.

[38] Farrar, D.E. and R.R. Glauber, *Multicollinearity in regression analysis: The problem revisited.* The Review of Economics and Statistics, 1967. **49**(1): p. 92-107.

[39] Fekedulegn, B.D., et al., *Coping with Multicollinearity: An Example on Application of Principal Components Regression in Dendroecology*. USDA Forest Service, Sep. 2002.

[40] Feldstein, M.S., *Multicollinearity and the mean square error of alternative estimators.* Econometrica: Journal of the Econometric Society, 1973: p. 337-346.

[41] Fitrianto, A. and L.C. Yik, *Performance of Ridge Regression Estimator Methods on Small Sample Size by Varying Correlation Coefficients: A Simulation Study*. Journal of Mathematics and Statistics, 2014. **10**(1): p. 25 - 29.

[42] Fox, J., *Regression Diagnostics*. Quantitative Applications in the Social Sciences ed. M.S. Lewis-Beck 1991: Sara Miller McCune, Sage Publications, Inc.

[43] Friedman, L. and M. Wall, *Graphical views of suppression and multicollinearity in multiple linear regression.* The American Statistician, 2005. **59**(2): p. 127-136.

[44] Garson, G.D., *Creating Simulated Datasets.* Statistical Associates Publishing, 2012: p. pp. 1 - 15.

[45] Guilkey, D.K. and J.L. Murphy, *Directed ridge regression techniques in cases of multicollinearity*. Journal of the American Statistical Association, 1975. **70**(352): p. 769-775.

[46] Gujarati, D.N. and Sangeetha, *Basic Econometrics* 4 th ed 2007, New Delhi: Tata McGraw-Hill.

[47] Gujarati, D.N. and D.C. Porter, *Essentials of econometrics.* 1999.

[48] Gunst, R.F. and R.L. Mason, *Biased Estimation in Regression: An Evaluation Using Mean Squared Error*. Journal of the American Statistical Association, Sep. , 1977. **72**(359): p. pp. 616-628.

[49] Gupta, V., *Regression Explained in Simple Terms*. VJBooks.net 2000: VJBooks.

[50] Hadi, A.S. and R.F. Ling, *Some Cautionary Notes on the Use of Principal Components Regression.* American Statistician, 2012. **52**(1): p. 15 - 19.

[51] Hocking, R.R ,.*Methods and Applications of Linear Models: Regression and the Analysis of Variance*. 2nd ed. Wiley Series in Probability and Statistics, ed. D.J. Balding, et al.2003, Hoboken: John Wiley and Sons, Inc. .

[52] Hoerl, A.E. and R.W. Kennard, *Ridge regression iterative estimation of the biasing parameter*. Communications in Statistics-Theory and Methods, 1976. **5**(1): p. 77-88.

[53] Hoerl, A.E. and R.W. Kennard, *Ridge Regression, 1980: Advances, Algorithms and Applications*1981: American Sciences Press.

[54] Hoerl, A.E. and R.W .Kennard, *Ridge Regression: Applications to Nonorthogonal Problems.* Technometrics, 1970. **12**(1): p. pp. 69-82.

[55] Hoerl, A.E. and R.W. Kennard, *Ridge regression: Biased estimation for nonorthogonal problems.* Technometrics, 1970. **12**(1): p. 55-67.

[56] Hoerl, A.E. and R.W. Kennard, *Ridge regression: biased estimation for nonorthogonal problems.* Technometrics, 2000. **42**(1): p. 80-86.

[57] Hoerl, A.E., R.W. Kannard, and K.F. Baldwin, *Ridge regression: some simulations.* Communications in Statistics-Theory and Methods, 1975. **4**(2 :(p. 105-123.

[58] Hoerl, A.E., R.W. Kennard, and R.W. Hoerl, *Practical use of ridge regression: a challenge met.* Applied Statistics, 1985: p. 114-120.

[59] Hoerl, R.W., J.H. Schuenemeyer, and A.E. Hoerl, *A Simulation of Biasesd Estimation and Subset Selection Regression Techniques.* Technometrics, Nov. 1986. **28**(4): p. pp. 369 - 380.

[60] Inc., S.I., *Base SAS ® 9.2 Procedures Guide*2009: Cary, NC: SAS Institute Inc.

[61] Jackson, J.E., *A User's Guide to Principal Components*. A Wiley-Interscience Publication1991, New York: John Wiley & Sons, Inc.

[62] Khalaf, G., *A Proposed Ridge Parameter to Improve the Least Square Estimator*. Journal of Modern Applied Statistical Methods, 2012. **11**(2 Article 15.): p. pp. 443 - 449.

[63] Kibria, B.G ,.*Performance of some new ridge regression estimators.* Communications in Statistics-Simulation and Computation, 2003. **32**(2): p. 419-435.

[64] Koop, G., *Analysis of Economic Data* 2 nd ed 2005, New York: John Wiley & Sons, Inc.

[65] Kovács, P., T. Petres, and L. Tóth, *A new measure of multicollinearity in linear regression models.* International Statistical Review, 2005. **73**(3): p. 405-412.

[66] Kozak, A., *Effects of multicollinearity and autocorrelation on the variable-exponent taper functions.* Canadian Journal of Forest Research, 1997. **27**(5): p. 619-629.

[67] Kutner, M.H., et al., *Applied Linear Statistical Models*. McGraw-Hill International Edition 5th ed. McGraw-Hill/Irwin Series 2005, Printed in Singapore: McGraw-Hill Companies, Inc.

[68] Lafi, S. and J. Kaneene, *An explanation of the use of principal-components analysis to detect and correct for multicollinearity.* Preventive Veterinary Medicine, 1992. **13**(4): p. 261-275.

[69] Lee, T.-S. and D.B. Campbell, *Selecting the optimum k in ridge regression.* Communications in Statistics-Theory and Methods, 1985. **14**(7): p. 1589-1604.

[70] Lockridge, J., *Stepwise analyses should never be used by researchers* in *Paper Presented at the annual meeting of the Southwest Educational Research Association* Jan. 1997: Austin, TX.

[71] Mansfield, E.R. and B.P. Helms, *Detecting multicollinearity.* The American Statistician, 1982. **36**(3a): p. 158-160.

[72] Mansson, K., G. Shukur, and B.G. Kibria, *On Some Ridge Regression Estimators: A Monte Carlo Simulation Study Under Different Error Variances.* Journal of Statistics, 2010. **17**: p. pp. 1-22.

[73] Marquardt, D.W. and R.D. Snee, *Ridge regression in practice.* The American Statistician, 1975. **29**(1): p. 3-20.

[74] Martin, D., *A Spreadsheet Tool for Learning the Multiple Regression F-test, t-tests, and Multicollinearity.* Journal of Statistics Education, 2008. **16**(3): p. n3.

[75] Mayes, T.R., *Generating Correlated Normally Distributed Random Numbers in Excel.* The Journal of Financial Modeling and Eduacational Technology, 2010. **1**(1): p. 59 - 66.

[76] McDonald, G.C. and D.I. Galarneau, *A Monte Carlo Evaluation of Some Ridge-Type Estimators.*Journal of the American Statistical Association, Tun. , 1975. **70**(350): p. pp. 407-416.

[77] Midi, H. and M. Zahari, *A Simulation Study On Ridge Regression Estimators In The Presence Of Outliers And Multicollinearity.* Jurnal Teknologi, Dis. 2007. **47**(C): p. 59–74.

[78] Montgomery, D.C. and E.A. Peck, *Introduction to Linear Regression Analysis*. Wiley Series in Probability and MathematicalStatistics1982, USA: John Wiley and Son, Inc.

[79] Muniz, G ,.et al., *On Developing Ridge Regression Parameters: a graphical Investigation.* SORT July-Dec. , 2012. **36**(2): p. pp. 115-138.

[80] Naes, T. and B.-H. Mevik, *Understanding the Collinearity Problem in Regression and Discriminant Analysis.* Journal of chemometrics, 2001(15): p. 413-426.

[81] O'Rourke, N., L. Hatcher, and E.J. Stepanski, *A Step-by-Step Approach to Using SAS for Univariate and Multivariate Statistics* 2nd ed 2005 Cary, NC: SAS Institute Inc. .

[82] Pasha, G.R. and M.A.A. Shah, *Application of Ridge Regression to Multicllinearity Data.* Journal of Research (Science), June, 2004. **15**(1): p. pp. 97-106.

[83] Paulson, D.S., *Handbook of Regression and Modeling: Applications for the Clinical and Pharmaceutical Industries*2007, USA: Chapman & Hall/CRC.

[84] Rao, C.R. and H. Toutenburg, *Linear Models: Least Squares and Alternatives*. 2nd ed1999, New York: Springer-Verlag New York, Inc.

[85] Rawlings, J.O., S.G. Pantula, and D.A. Dickey, *Applied Regression Analysis: A Research Tool*. 2 ed1998, New York: Sppringer-Verlag Inc.

[86] Rencher, A.C .and G.B. Schaalje, *Linear Models in Statistics*. 2nd ed 2008, Hoboken John Wiley & Sons, Inc.

[87] Schroeder, M.A., J. Lander, and S. Levine-Silverman, *Diagnosing and dealing with multicollinearity.* Western Journal of Nursing Research, 1990. **12**(2): p. 175-187.

[88] Sen, A. and M. Srivastava, *Regression Analysis: Theory, Methods, and Applications* Springer Texts in Statistics ed. G. Casella, S. Fienberg, and I. Olkin, 1990, New York: Sprnger-Verlag Inc.

[89] Shieh, Y.-Y. and R.T. Fouladi, *The effect of multicollinearity on multilevel modeling parameter estimates and standard errors.* Educational and Psychological Measurement, 2003. **63**(6): p. 951-985.

[90] Spanos, A. and A. McGuirk, *The problem of near-multicollinearity revisited: erratic vs systematic volatility.* Journal of econometrics, 2002. **108**(2): p. 365-393.

[91] Sufian, A.J.M., *Analyzing Collinear Data by Principal Component Regression Approach - An Example from Developing Countries.* Journal of Data Science, 2005. **3**: p. pp. 221 - 232.

[92] Tabachnick, B.G. and L.S. Fidell, *Using Multivariate Statistics* 5 th ed 2007, Boston USA: Pearson Education, Inc.

[93] Troskie, C. and D. Chalton. *Detection of outliers in the presence of multicollinearity*. in *Multidimensional statistical analysis and theory of random matrices, Proceedings of the Sixth Lukacs Symposium, eds. Gupta, AK and VL Girko*. 1996.

[94] Tutz, G. and H. Binder, *Boosting ridge regression.* Computational Statistics & Data Analysis, 2007. **51**(12): p. 6044-6059.

[95] Vaughan, T.S. and K.E. Berry, *Using Monte Carlo techniques to demonstrate the meaning and implications of multicollinearity.* Journal of Statistics Education, 2005. **13**(1): p. n1.

[96] Vinod, H.D., *A survey of ridge regression and related techniques for improvements over ordinary least squares.* The Review of Economics and Statistics, 1978. **60**(1): p. 121-131.

[97] Wang, S.-G., S.-K. Tse, and S.-C. Chow, *On the measures of multicollinearity in least squares regression.* Statistics & probability letters, 1990. **9**(4): p. 347-355.

[98] Walker, E. and J.B. Birch ,*Influence measures in ridge regression.* Technometrics, 1988. **30**(2): p. 221-227

[99] Weissfeld, L.A., *A multicollinearity diagnostic for models fit to censored data.* Communications in Statistics-Theory and Methods, 1989. **18**(6): p. 2073-2085.

[100] Wicklin, R., *Simulating Data with SAS* 2013, Cary , North Carolina , USA: SAS Institute Inc.

[101] Willan, A.R. and D.G. Watts, *Meaningful multicollinearity measures.* Technometrics, 1978. **20**(4): p. 407-412.

[102] Yan, X. and X.G. Su, *Linear Regression Analysis: Theory and Computing*2009, Printed in Singapore: World Scientific Publishing Co. Pte. Ltd.

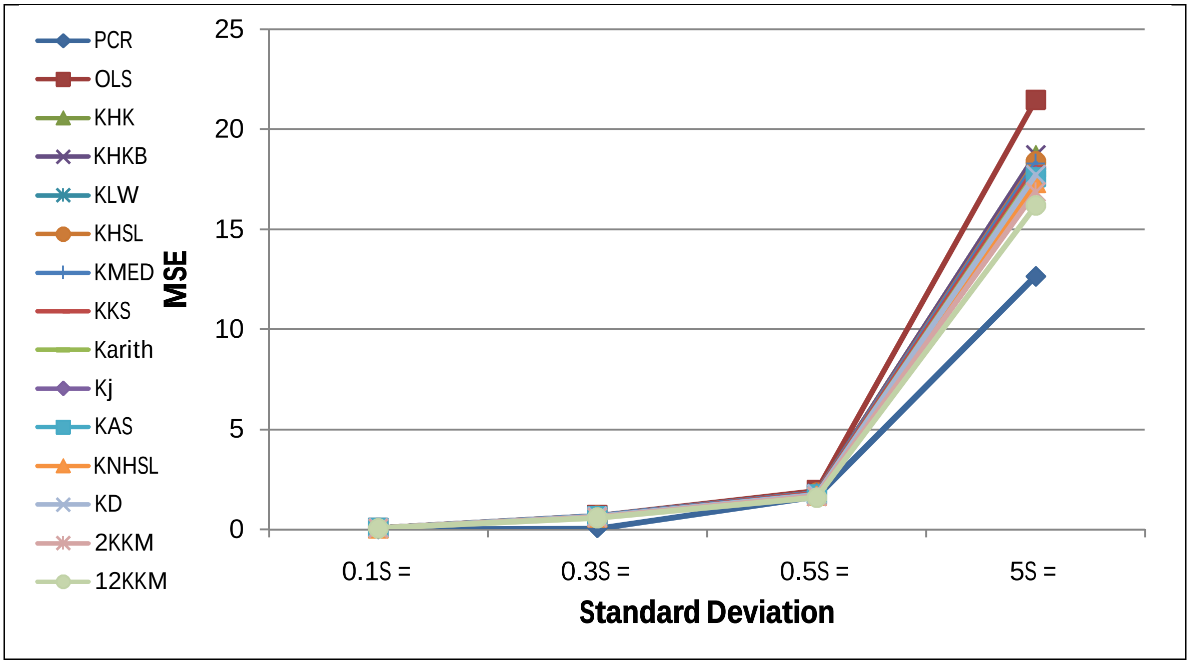
[103] Yong-wei, G., et al., *A method to measure and test the damage of multicollinearity to parameter estimation [J].* Science of Surveying and Mapping, 2008. **2**: p. 044.

[104] Yoshioka, S., *Multicollinearity and Avoidance in Regression Analysis.* Behaviormetrika, 1986(19): p. 103-120.

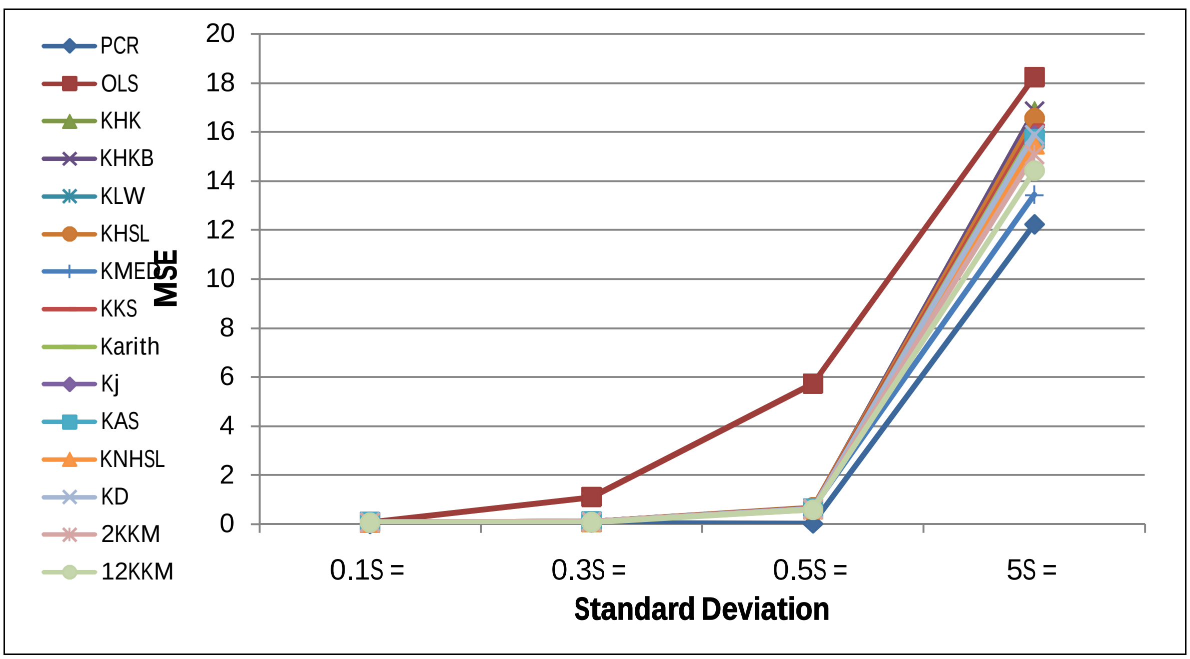
[105] Zhang, J. and M. Ibrahim, *A Simulation Study on SPSS Ridge Regression and Ordinary Least Squares Regression Procedures fpr Multicollinearity Data.* Journal of Applied Statistics, Aug. , 2005. **32**(6): p. pp. 571 - 588.

الملاحق

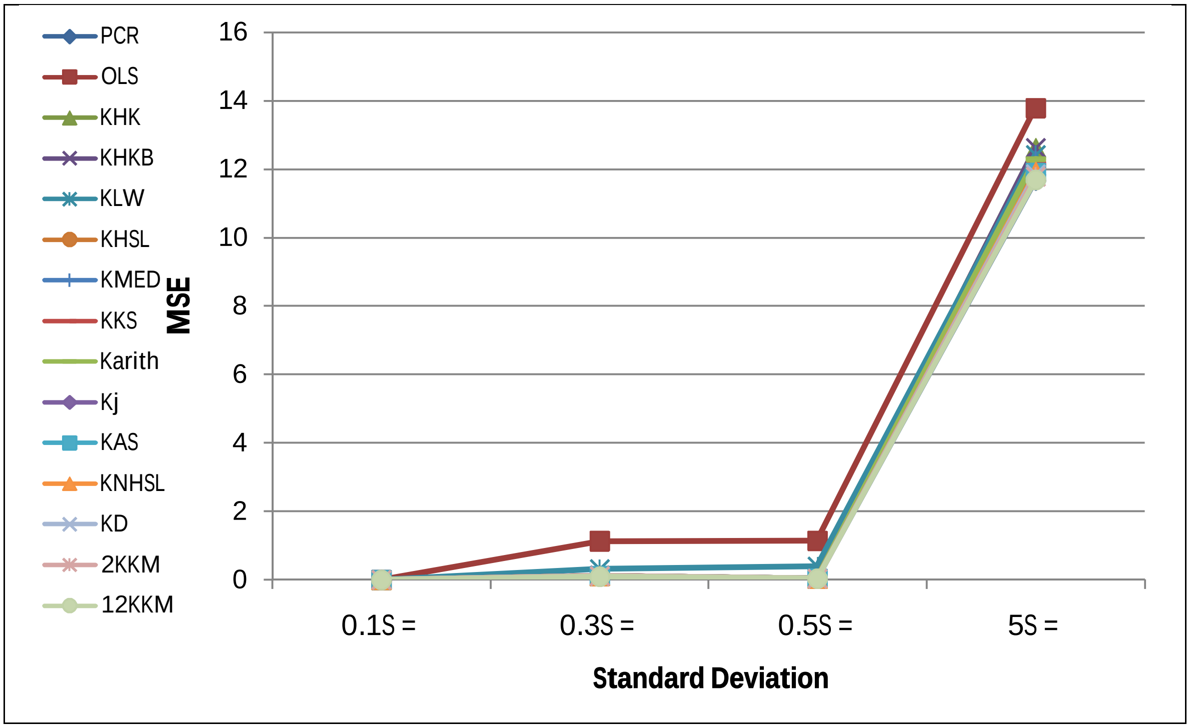
ملحق (1): أداء مُقدرات PCR و RR و OLS كدالة في مستوى التباين.



ملحق (1– أ): أداء مُقدرات PCR و RR و OLS كدالة في مستوى التباين، عند مستوى ارتباط r = 0.99 وحجم عينة n = 10

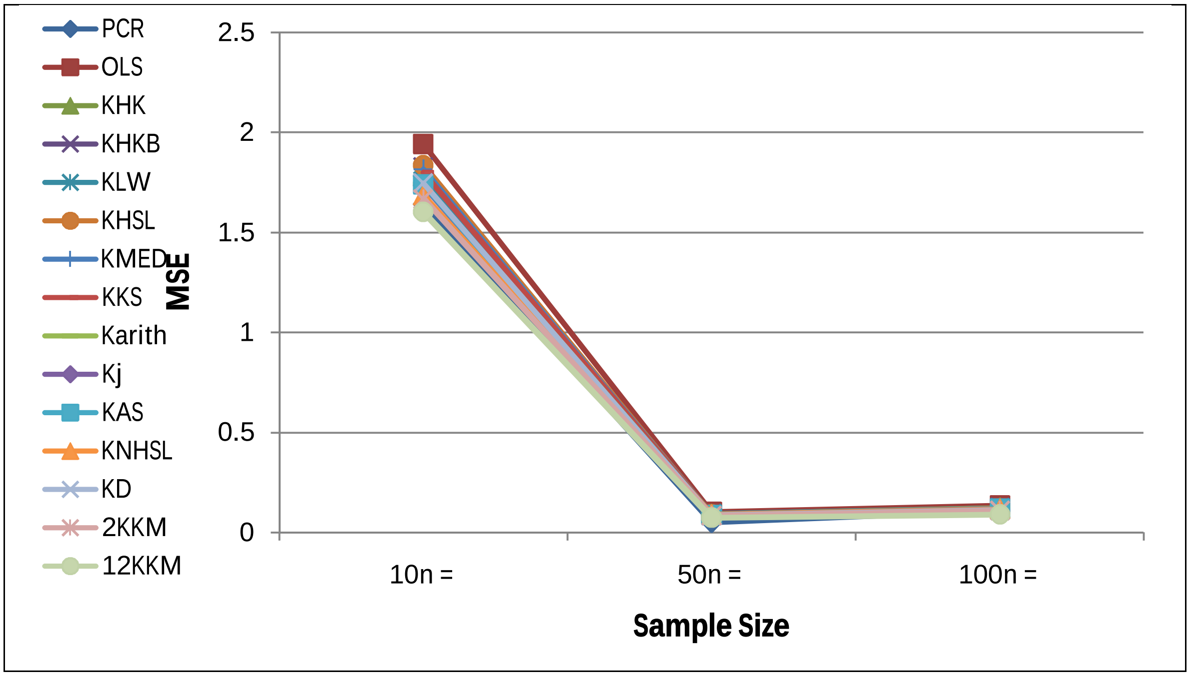


ملحق (1–ب): أداء مُقدرات PCR و RR و OLS كدالة في مستوى التباين، عند مستوى ارتباط r = 0.99 وحجم عينة n = 50

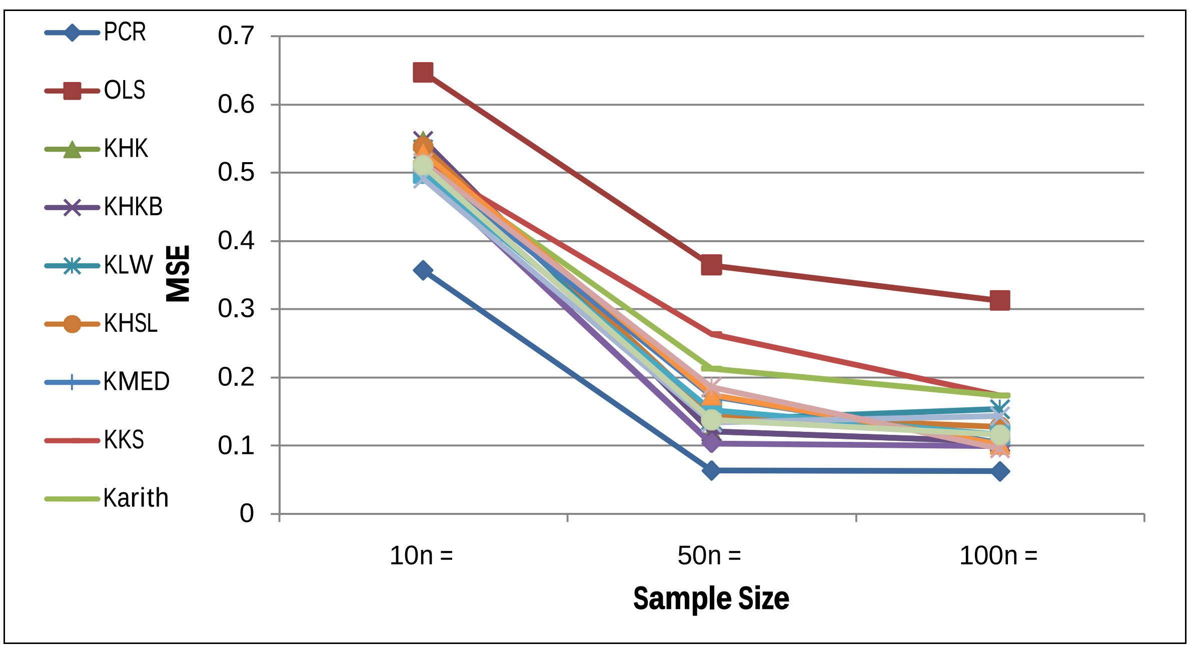


ملحق (1–ج): أداء مُقدرات PCR و RR و OLS كدالة في مستوى التباين، عند مستوى ارتباط $r = 0.99$ وحجم عينة $n = 100$

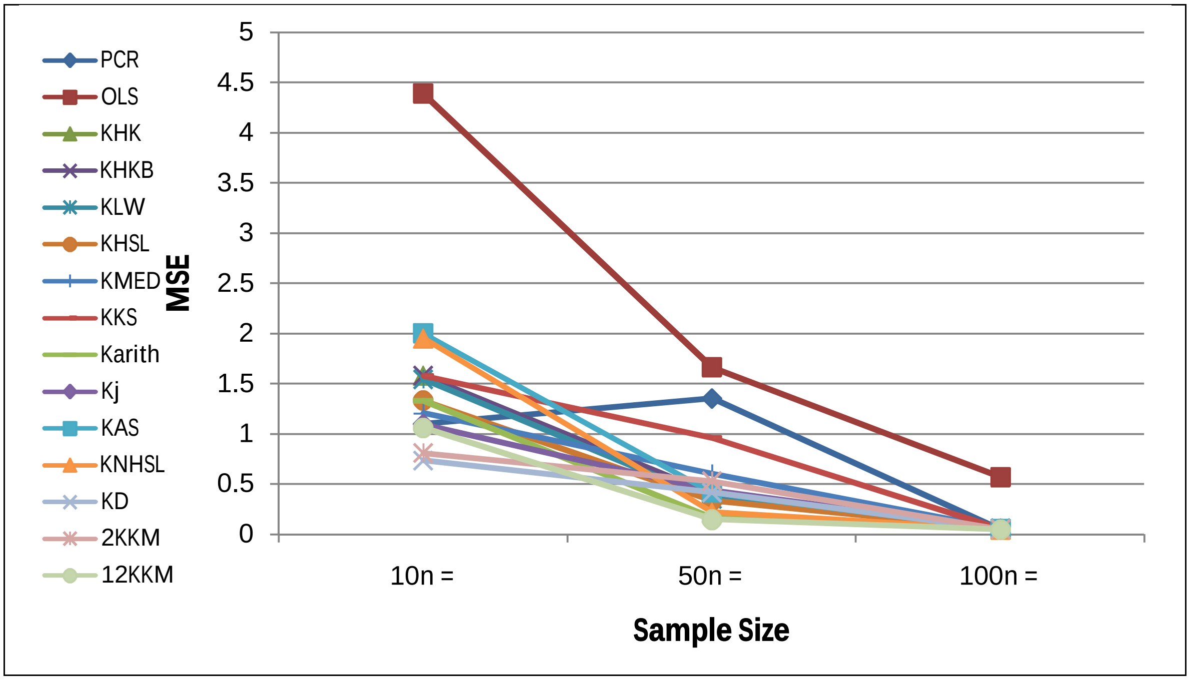
ملحق (2): أداء مُقدرات PCR و RR و OLS كدالة في حجم العينة.



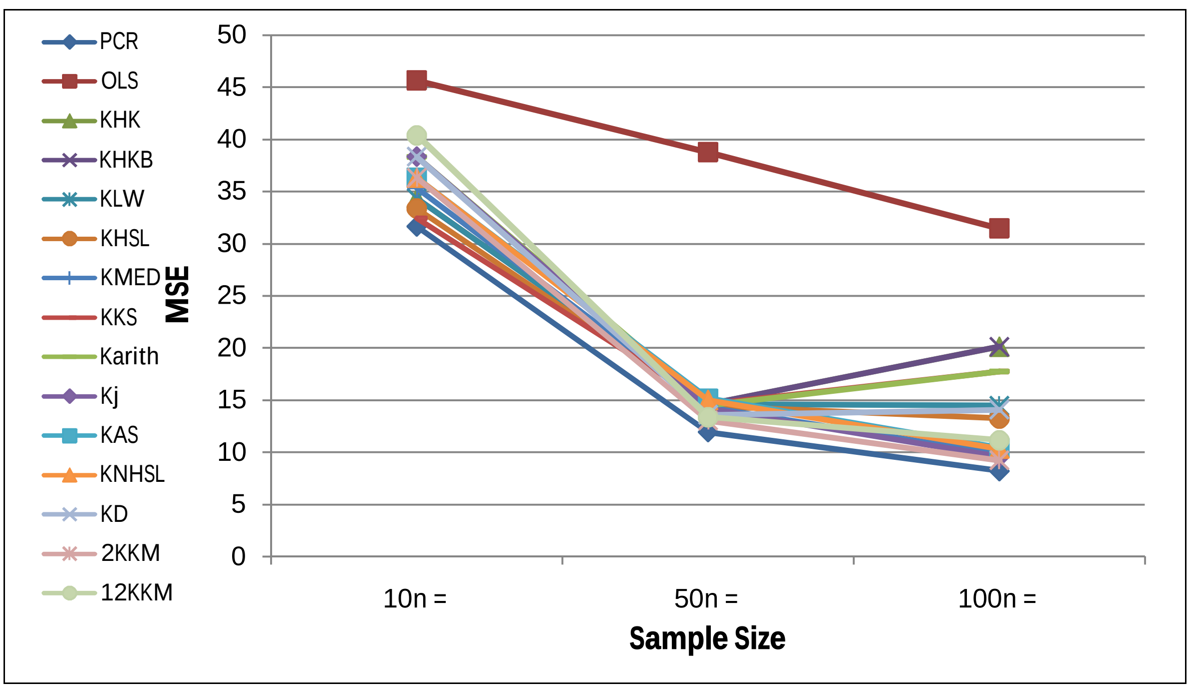
ملحق (2– أ): أداء مُقدرات PCR و RR و OLS كدالة في حجم العينة ، وعند مستوى ارتباط r = 0.99 وانحراف معياري σ = .1



ملحق (2– ب): أداء مُقدرات PCR و RR و OLS كدالة في حجم العينة، عند مستوى ارتباط r = 0.99 وانحراف معياري σ = .3

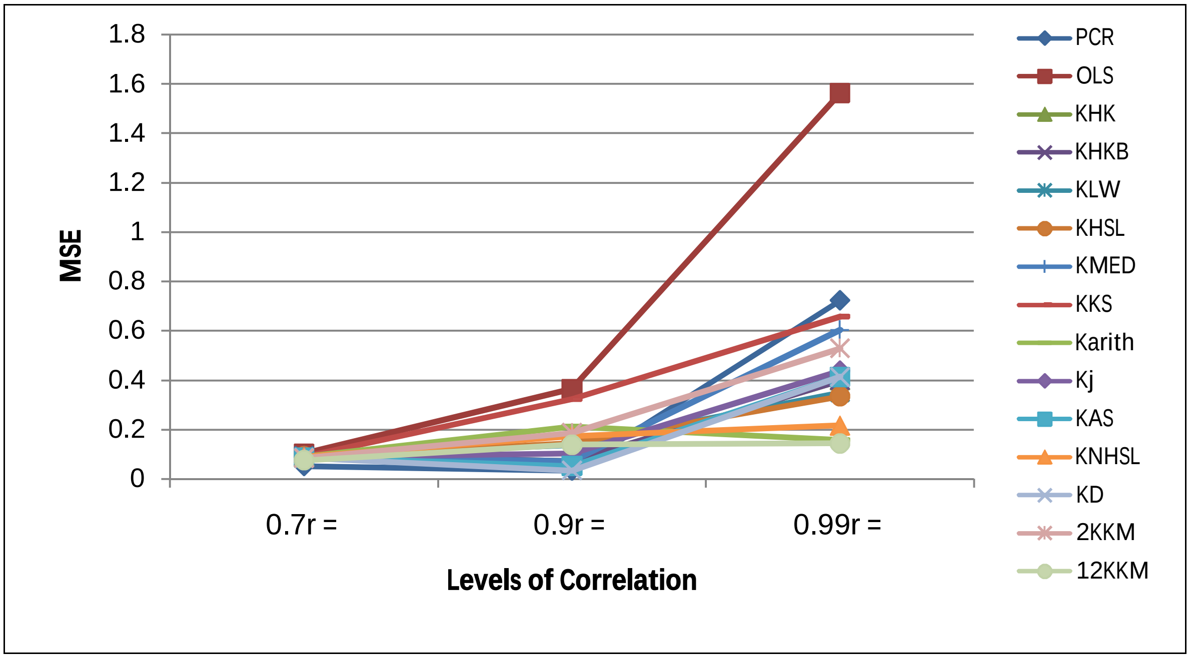


ملحق (2–ج): أداء مُقدرات PCR و RR و OLS كدالة في حجم العينة، عند مستوى ارتباط r
= 0.99 وانحراف معياري $\sigma = .5$

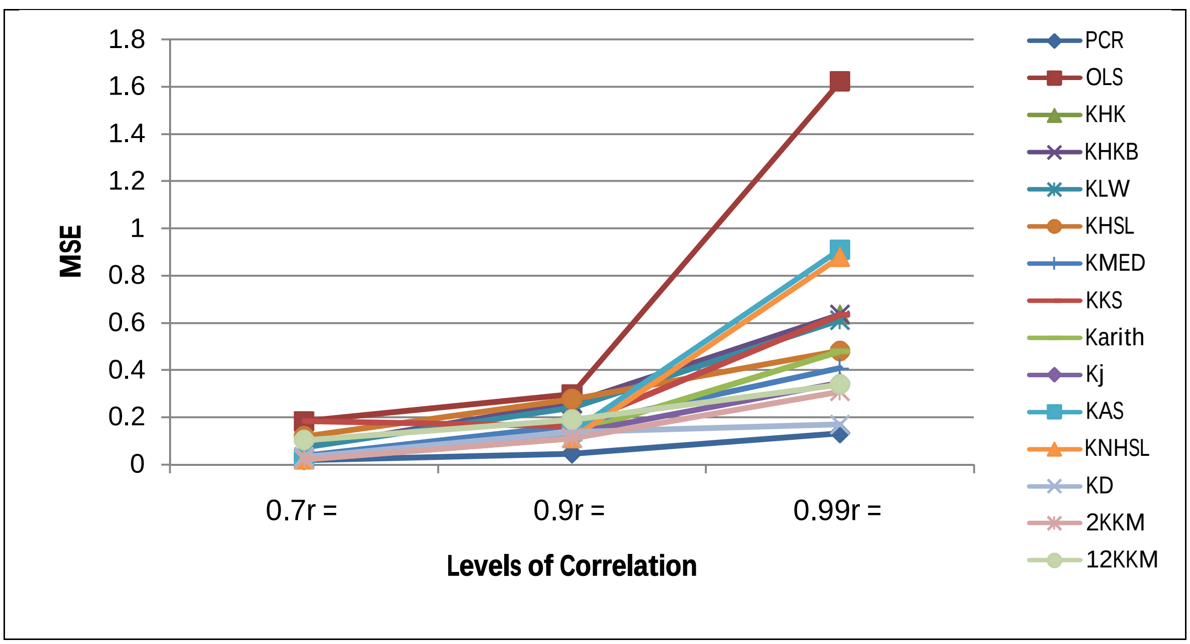


ملحق (2–د): أداء مُقدرات PCR و RR و OLS كدالة في حجم العينة، عند مستوى ارتباط r
= 0.99 وانحراف معياري $\sigma = 5$

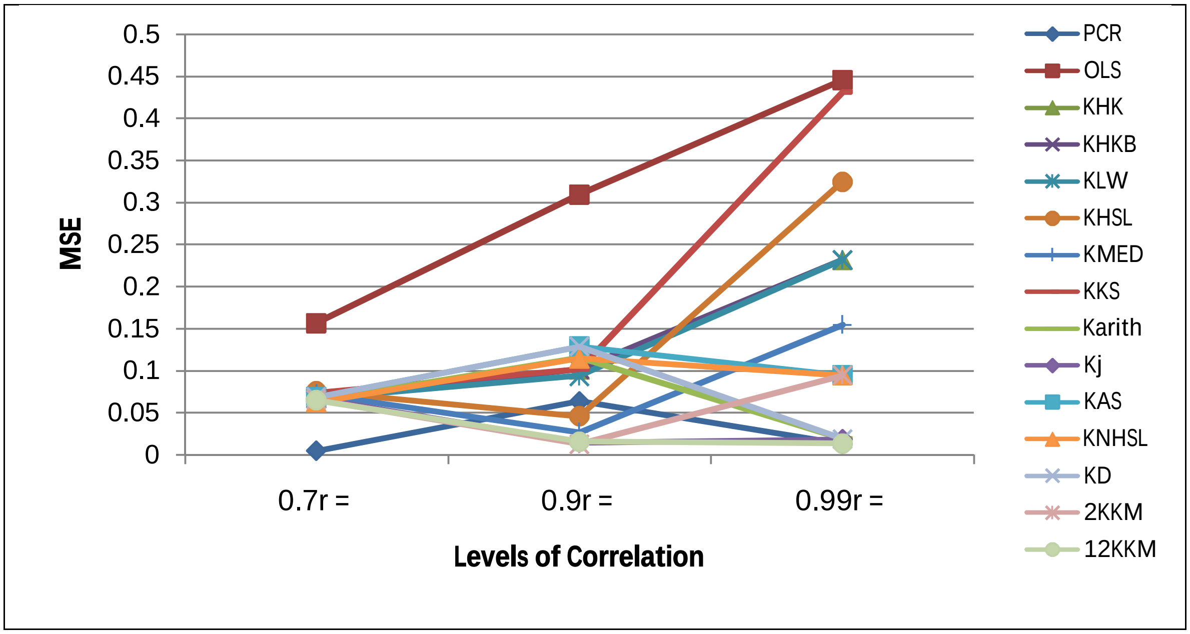
ملحق (3): أداء مُقدرات PCR و RR و OLS كدالة في مستوى معامل الارتباط γ.



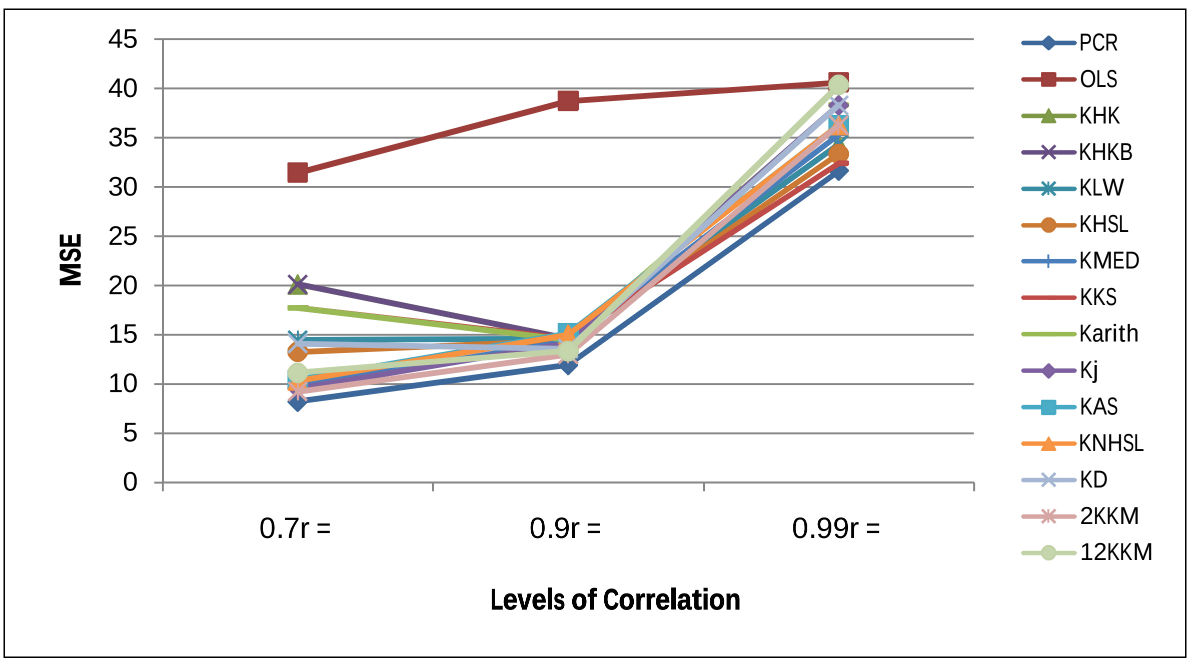
ملحق (3–أ): أداء مُقدرات PCR و RR و OLS كدالة في مستوى معامل الارتباط γ، عند مستوى ارتباط $r = 0.99$ وانحراف معياري $\sigma = .1$



ملحق (3–ب): أداء مُقدرات PCR و RR و OLS كدالة في مستوى معامل الارتباط γ، عند مستوى ارتباط $r = 0.99$ وانحراف معياري $\sigma = .3$

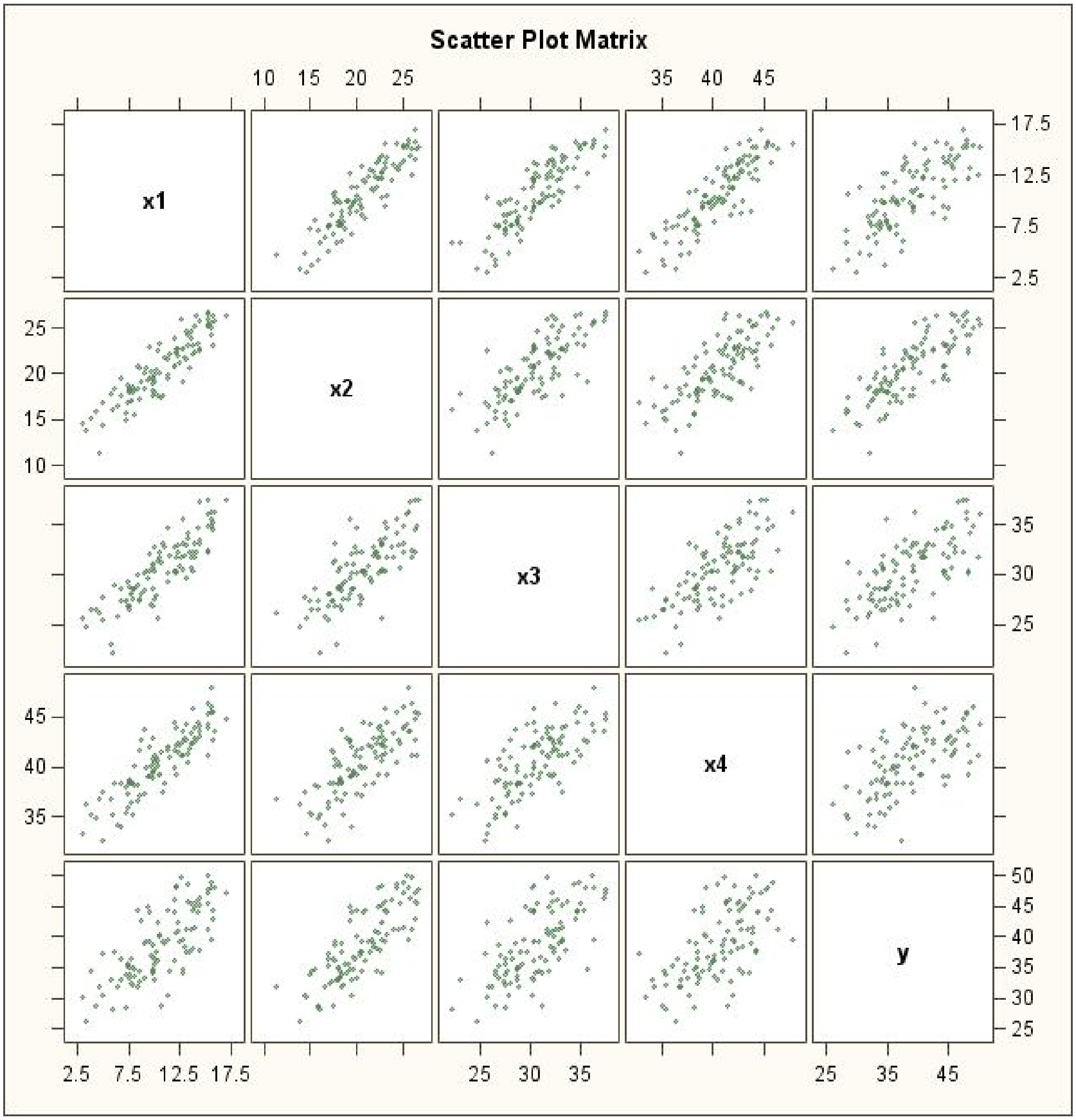


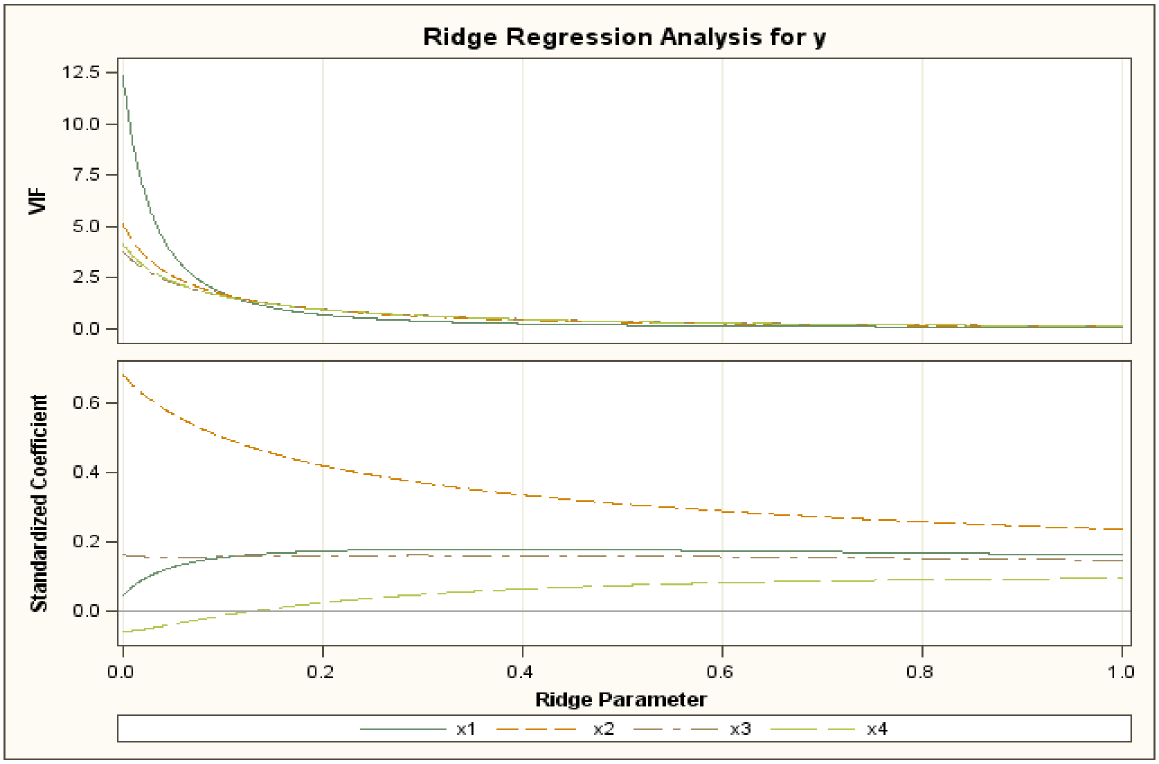
ملحق (3–ج): أداء مُقدرات PCR و RR و OLS كدالة في مستوى معامل الارتباط، γ، عند مستوى ارتباط $r = 0.99$ وانحراف معياري $\sigma = .5$

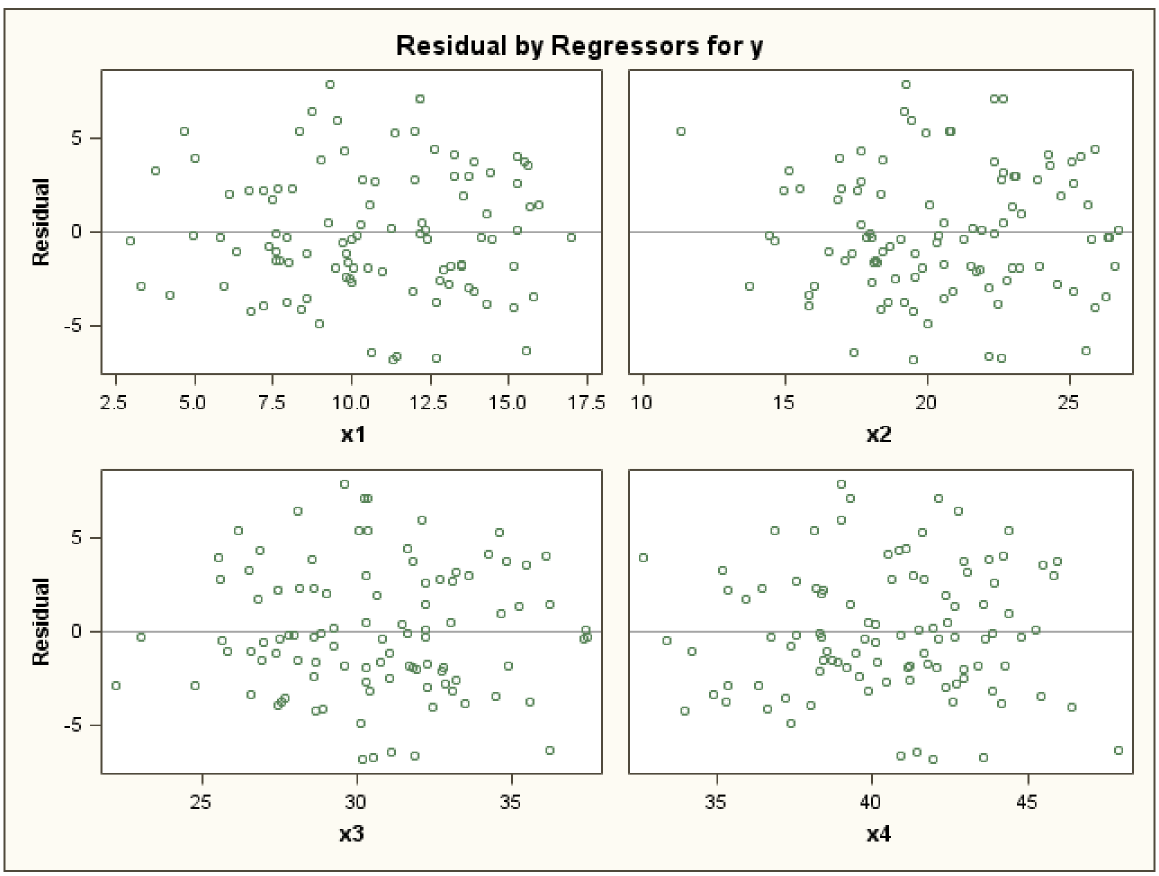


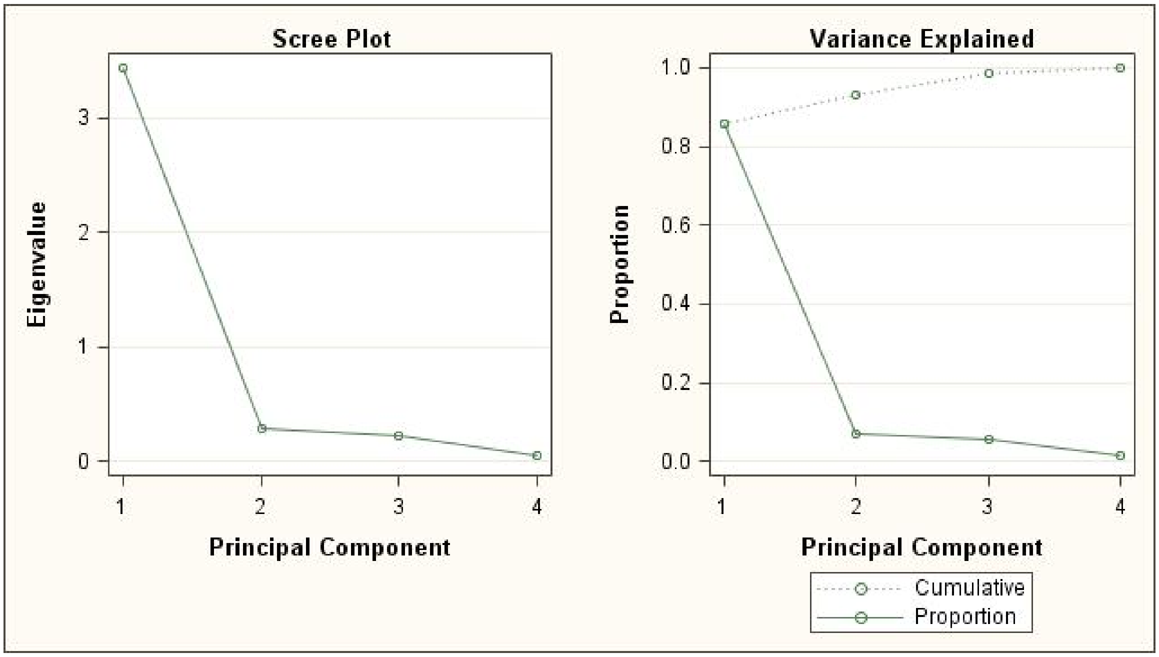
ملحق (3–د): أداء مُقدرات PCR و RR و OLS كدالة في مستوى معامل الارتباط، γ، عند مستوى ارتباط $r = 0.99$ وانحراف معياري $\sigma = 5$

ملحق (4): عينة من مخرجات برنامج SAS V. 9.2 .

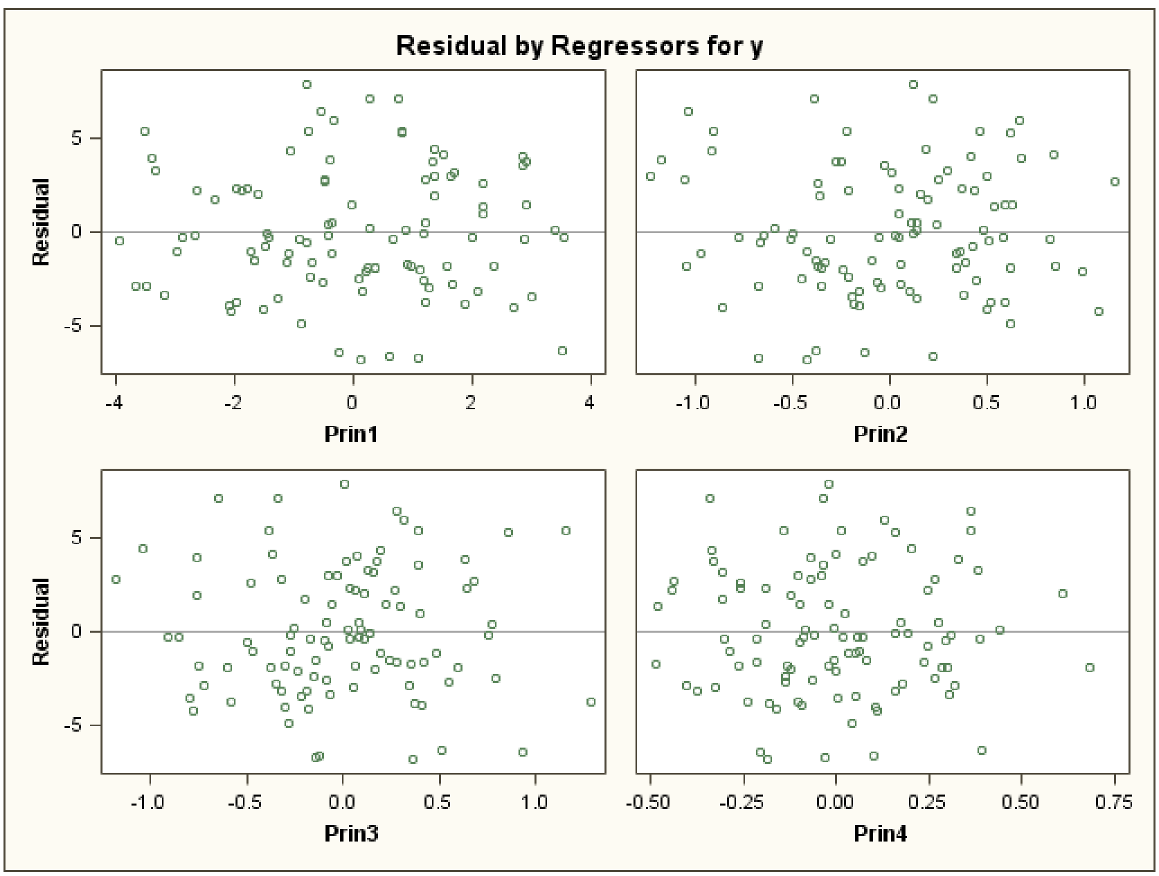



Ridge Regression Analysis for y
VIF
12.5
10.0
7.5
5.0
2.5
0.0
Standardized Coefficient
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Ridge Parameter
x1
x2
x3
x4

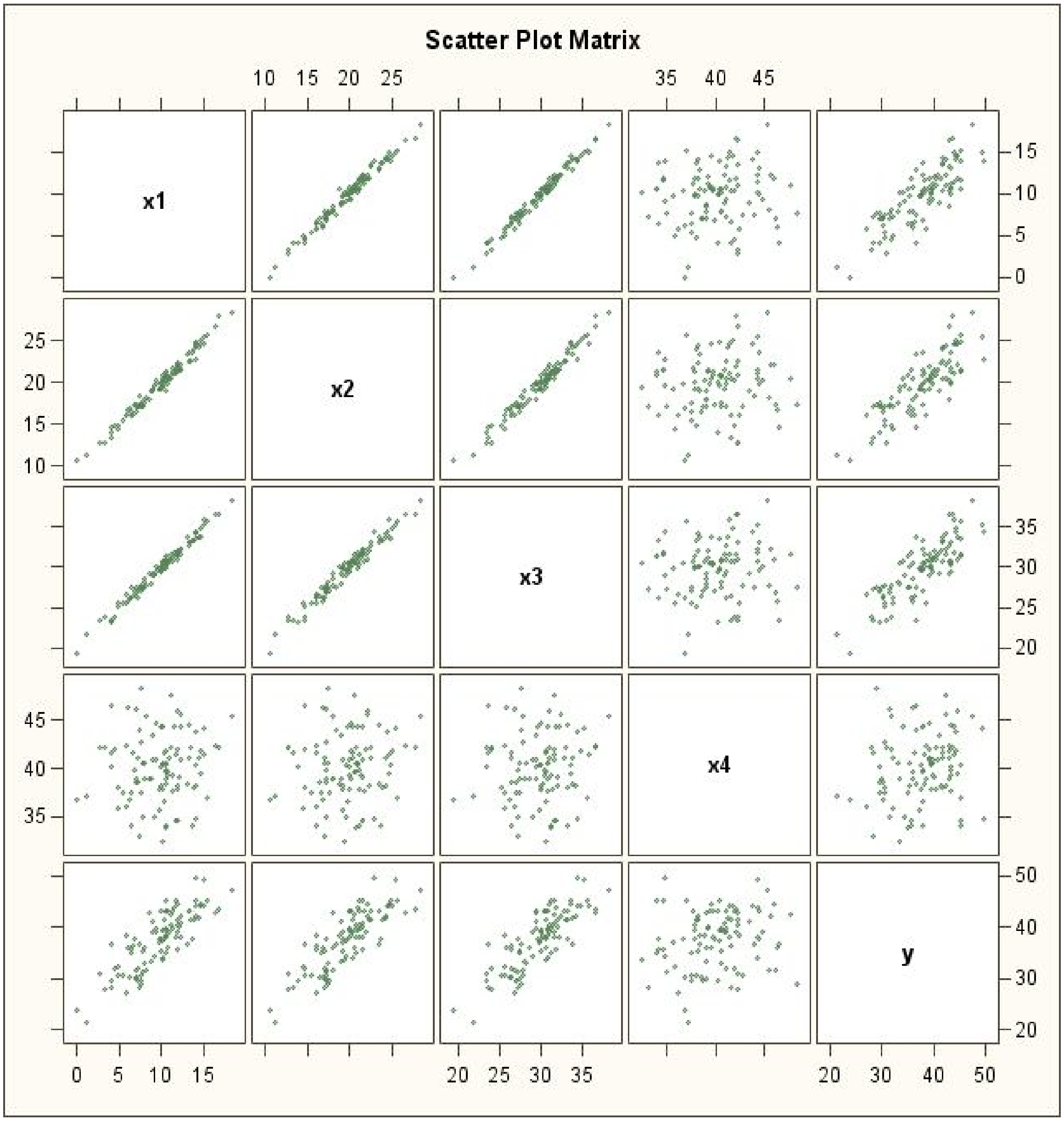

Residual by Regressors for y
Residual
5
0
-5
2.5
5.0
7.5
10.0
12.5
15.0
17.5
x1
10
15
20
25
x2
Residual
25
30
35
x3
35
40
45
x4

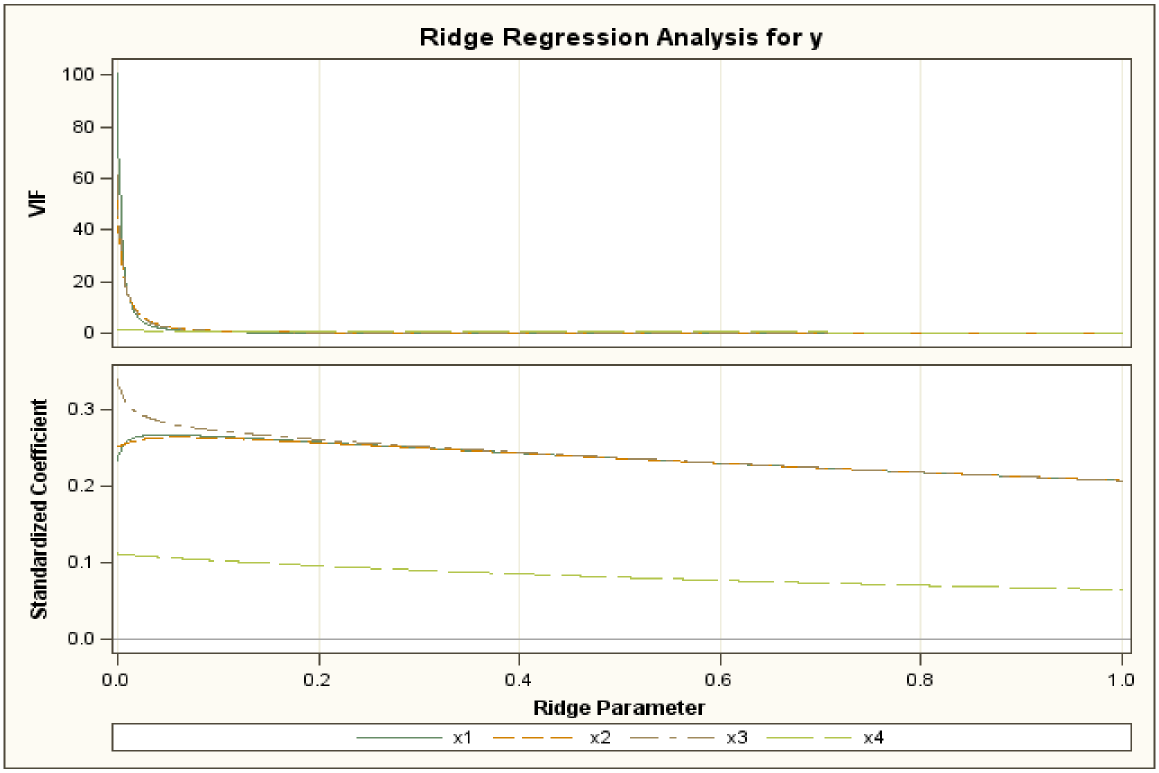


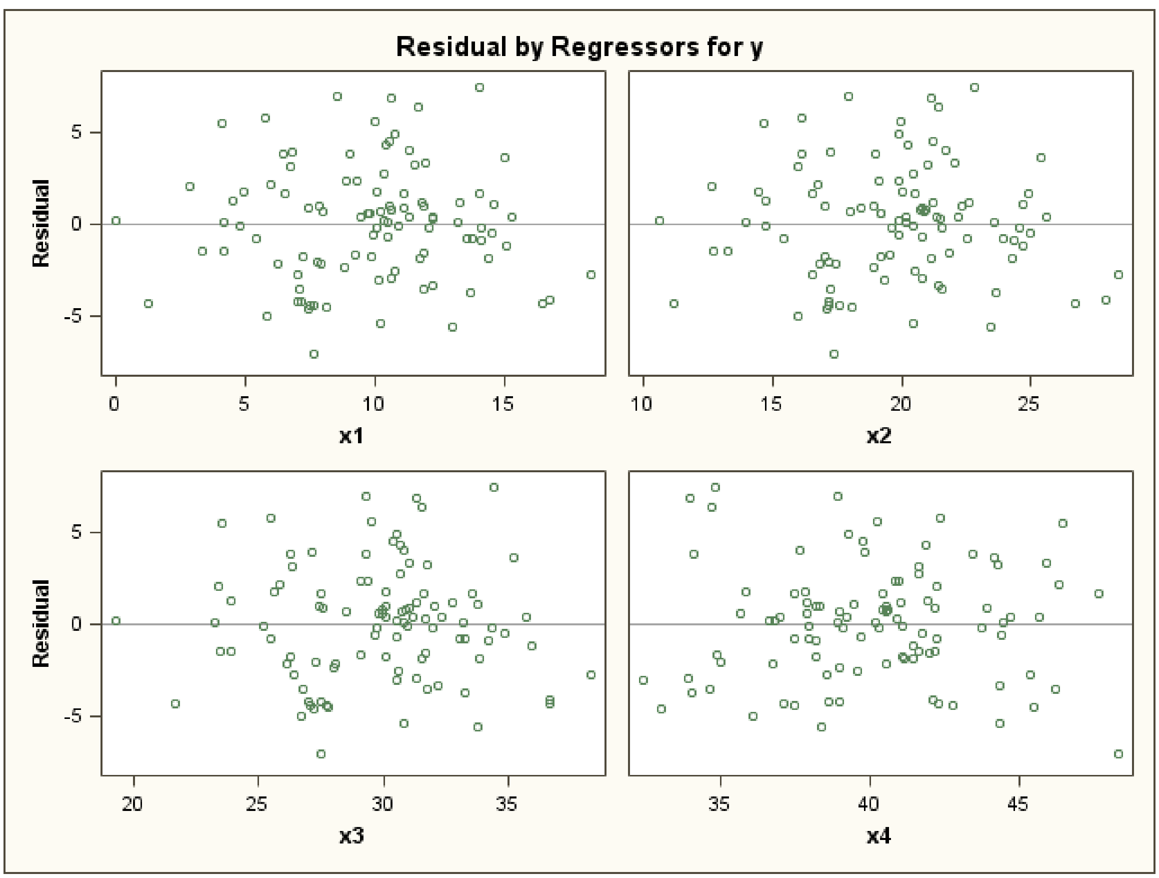
Scree Plot
Eigenvalue
Principal Component
Variance Explained
Proportion
Principal Component
Cumulative
Proportion

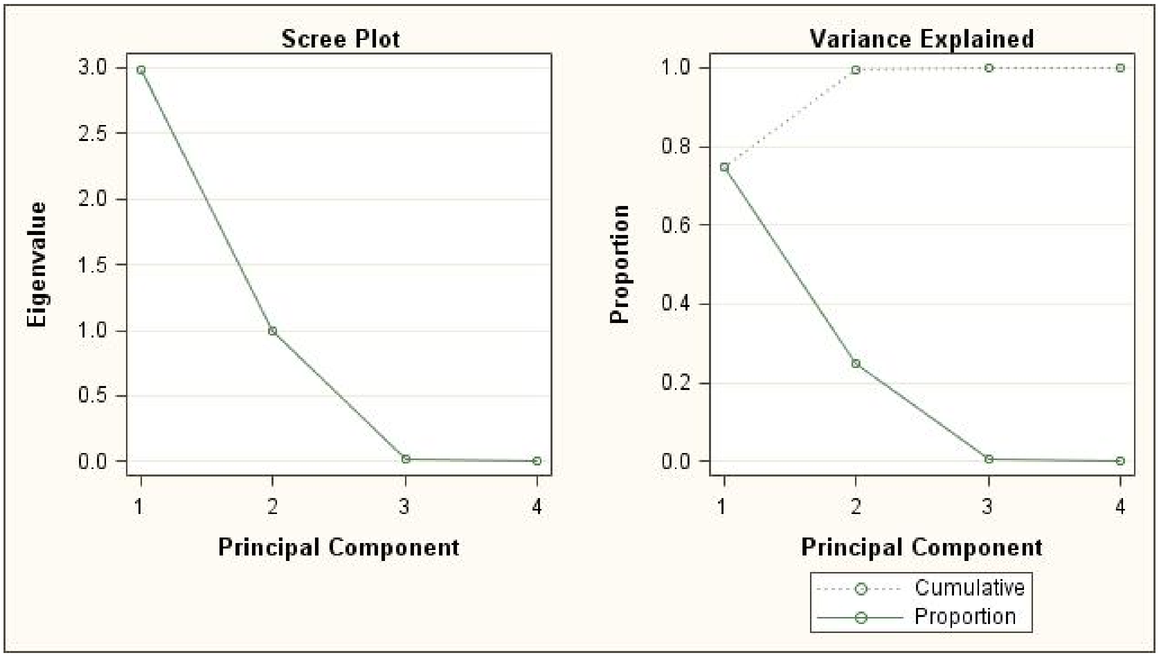


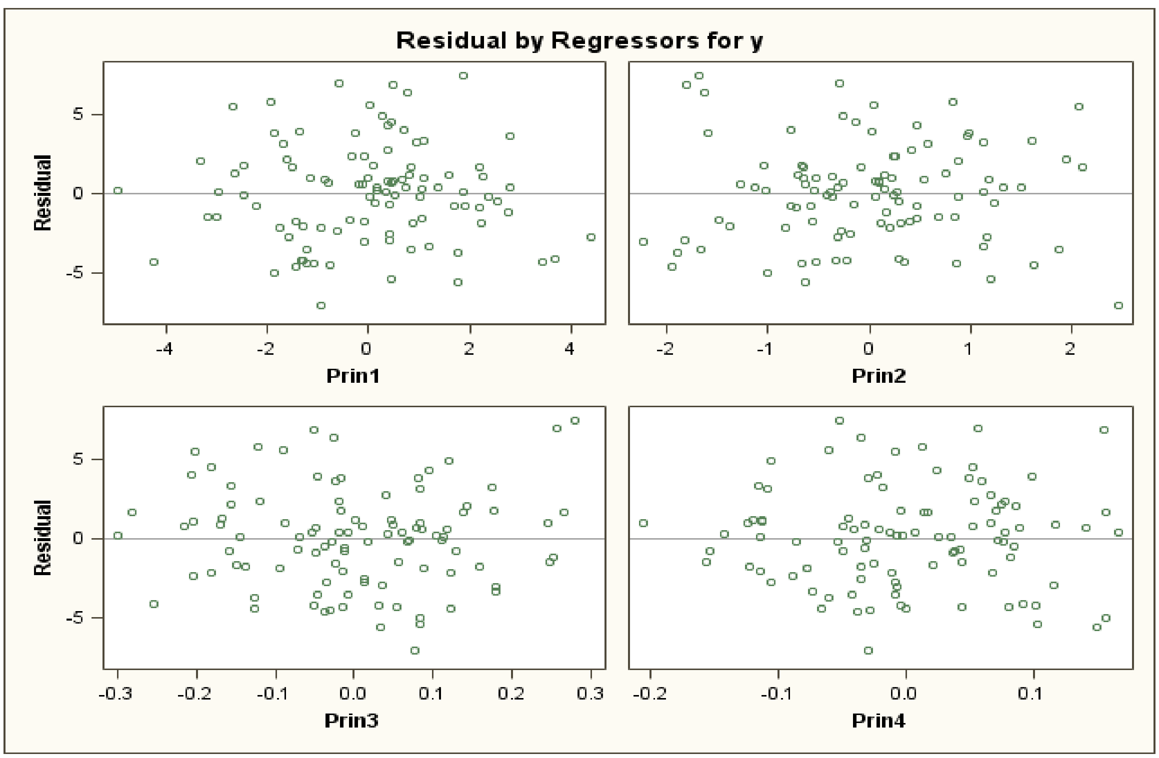
Residual by Regressors for y
Residual
Prin1
Prin2
Residual
Prin3
Prin4


Scatter Plot Matrix
x1
x2
x3
x4
y
10 15 20 25
35 40 45
15
10
5
0
25
20
15
10
35
30
25
20
45
40
35
50
40
30
20
0 5 10 15
20 25 30 35
20 30 40 50


Ridge Regression Analysis for y
VIF
100
80
60
40
20
0
Standardized Coefficient
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
0.8
1.0
Ridge Parameter
x1
x2
x3
x4


Residual by Regressors for y
Residual
5
0
-5
0
5
10
15
x1
10
15
20
25
x2
Residual
20
25
30
35
x3
35
40
45
x4


Scree Plot
Eigenvalue
Principal Component
Variance Explained
Proportion
Principal Component
Cumulative
Proportion


Residual by Regressors for y
Residual
Prin1
Prin2
Residual
Prin3
Prin4