طريقة ومعيار جديدان للكشف عن مشكلة التعدد الخطي ومعالجتها واختيار متغيرات الانحدار

د. صباح فرج عبد الحسين*

المستخلص:

تعد مشكلة التعدد الخطي من المشاكل الكبرى التي قد تواجه الباحث عند وصفه للعلاقة الخطية بين ظاهرة عشوائية معينة والعوامل المؤثرة فيها بمعادلة انحدار خطي متعدد، وذلك لما تؤدي اليه من تشويه لهذه العلاقة على صعيد كل عامل من هذه العوامل بحيث تبدو هذه العوامل وكأنها غير مهمة لتفسير الظاهرة... وقد اقترحت حلول متعددة لمعالجة هذه المشكلة الا ان هذه الحلول- كما يرى العديد من الباحثين- كان لها آثار جانبية تمثلت بالأضرار باحدى خصائص التقدير الخطي الأفضل لمعامل الانحدار وفقدان القدرة على تحقيق النموذج الوافي (الصحيح) فضلاً عن التفسير البسيط لمعاملات الانحدار. ولذلك اقترح الباحث معياراً وطريقة جديدة للاستدلال على وجود مشكلة التعدد الخطي ومعالجتها بدون آثار جانبية تقريباً. وقد اشتق الباحث هذا المعيار الذي أسماه RUF عامل عدم الاستقرار النسبي باستخدام طريقة المربعات الصغرى الاعتيادية في تقدير معلمات الانحدار، وأثبت رياضياً ان هذا العامل هو نفسه عامل تغير الأهمية النسبية لمتغيرات الانحدار، وعليه فقد استخدمه في الوصول الى أفضل نموذج جزئي للانحدار الخطي وفقاً لمعايير المفاضلة المعروفة وبالمقارنة مع الطرائق الأخرى.

Abstract:

Multi co linearity is regarded as a big problem which may face a researcher while studying a linear relation between some random phenomenon and the factors that affect it using multiple linear regression equation because it leads to disfigure that relation for each factor such that the factors look like as if they where not important to explain the

---

* مدرس / معهد الإدارة / الرصافة

مقبول للنشر بتاريخ 2010/12/21

phenomenon. Several solutions are proposed to treat the problem but all these solutions- as some researchers believe- have side-effects such a disadvantaging at least one of (BLUE) properties and disabling to accomplish the adequate model as well as the simple explanation of the regression- coefficients. So the researcher proposed a new criterion and procedure to infer and treat the multicollinearity problem without side-effects approximately. He derived his criterion named RUF using OLS method to estimate the regression- parameters and he proved that criterion itself is the factor of changing the relative importance of regression variables more over he used it to achieve the best sub- regression model according to standard criteria and by comparing with the other procedures.

Key- words:

Multicollinearity: التعدد الخطي; Adequate Model: النموذج الوافي (الصحيح); VIF: عامل تضخم التباين; RUF: عامل عدم الاستقرار النسبي.

1. المقدمة والهدف:

إن تفسير معادلة الانحدار الخطي المتعدد يعتمد ضمنياً على الفرضية التي تقول ان المتغيرات التفسيرية لا تكون مرتبطة مع بعضها ارتباطاً قوياً، فمعامل الانحدار عادة يفسر على انه مقياس للتغير في المتغير المعتمد عندما يزداد المتغير التفسيري المناظر بمقدار وحدة واحدة وان كل المتغيرات التفسيرية الأخرى تعد ثابتة، وهذا التفسير يمكن ان لا يكون صحيحاً اذا وجدت علاقة خطية قوية بين المتغيرات التفسيرية تنعكس في التأثير على علاقة هذه المتغيرات بالمتغير المعتمد.

ان الذي يهم الباحث في هذه العلاقة بين المتغيرات التفسيرية هو انعكاسها السلبي على علاقة كل متغير تفسيري بالمتغير المعتمد والتي يمكن أن تأخذ واحداً أو أكثر من الأعراض الآتية:

1. عدم أهمية المتغير التفسيري في نموذج الانحدار الخطي المتعدد حيث يبدو معامل انحداره غير معنوي بفعل علاقته الخطية القوية مع غيره من المتغيرات التفسيرية التي تؤدي الى تضخم الخطأ المعياري لتقديره.
2. تشوه طبيعة العلاقة ما بين المتغير التفسيري والمتغير المعتمد من خلال التغير الكبير في قيمة معامل انحدار المتغير التفسيري والتي قد تصل في بعض الأحيان الى تغير اشارته التي تدل على كون العلاقة طردية أم عكسية، مما يعطي تفسيراً مضللاً لهذه العلاقة الفردية بين المتغير التفسيري والمتغير المعتمد. وبذلك تكون تقديرات معاملات الانحدار غير دقيقة وعملية اختبار الفرضيات حول هذه المعاملات، مربكة ومشكوكاً فيها. ولحل هذه المشكلة اقترحت طرائق

مختلفة، جرى استعراضها في هذا البحث، ثم اقترح الباحث طريقة جديدة تحقق الأهداف الآتية معاً:

1. حل مشكلة العلاقات الخطية المتعددة في معادلة الانحدار الخطي المتعدد وزوال كافة مظاهرها وآثارها على تقديرات معاملات الانحدار.
2. تحقيق أفضل نموذج جزئي للانحدار الخطي المتعدد على وفق معايير المفاضلة المعروفة.

2. الجانب النظري Theoretical part.

2-1 مفهوم التعدد الخطي وآثاره على تقديرات معاملات الإنحدار.

2-1-1 مفهوم التعدد الخطي Multicollinearity - concept:

وردت في الأدب الاحصائي صياغات مختلفة للتعبير عن مفهوم التعدد الخطي ما بين المتغيرات التفسيرية في نموذج الانحدار الخطي المتعدد، نذكر منها على سبيل المثال لا الحصر، ما يلي:

- يرى الباحثان Samprit Chatter Jee & Bertram Price ان مشكلة التعدد الخطي ((هي حالة عدم التعامد الخطيرة ما بين المتغيرات التفسيرية)) [1].
- ويرى Bowerman & O'connell بأنها ((حالة اعتماد كل متغير تفسيري على المتغيرات التفسيرية الأخرى)) [2].
- بينما يرى George O. Wesolowsky بأنها: ((حالة ارتباط أحد المتغيرات المستقلة بمتغير مستقل آخر أو مجموعة من المتغيرات المستقلة ذات إتجاه خطي)) [3].
- أما الدكتور دومينيك سالفاتور أستاذ الاقتصاد في جامعة فوردهام فأنه يرى بأن مشكلة التعدد الخطي هي: ((الحالة التي يكون فيها بين اثنين أو أكثر من المتغيرات التفسيرية في نموذج الانحدار ارتباط قوي مما يجعل من الصعب أو المستحيل عزل تأثيراتها الفردية على المتغير التابع (المعتمد))) [4].
- وأما Smith & Drapper فانهما يريان بأنها ((الحالة المرضية للمتغيرات التفسيرية التي تكون فيها أعمدة هذه المتغيرات في مصفوفتها $X'X$ تعتمد احداها على الأخرى بصورة تقريبية)) [5].

ويلاحظ ان كل هذه الصياغات الأدبية للتعبير عن مفهوم التعدد الخطي تكاد تكون واحدة في جوهرها فهي تجمع على انها حالة تكون فيها المتغيرات التفسيرية معتمدة بعضها على البعض الآخر، والباحث لا يختلف مع غيره في تشخيص هذه الحالة الا أنه يرى أن التعبير عن مفهوم التعدد الخطي بهذه الصياغات العامة قد يوحي- وهو ما نجده في العديد من الأدبيات الاحصائية- بطريقة خاطئة لمعالجة مشكلة التعدد الخطي في نموذج الانحدار الخطي المتعدد، مثال ذلك:

1. حذف المتغير الذي له أكبر معامل ارتباط بسيط مع غيره من المتغيرات التفسيرية.
2. حذف المتغير الذي له أكبر مربع معامل ارتباط متعدد مع غيره من المتغيرات التفسيرية.
3. حذف المتغير الذي تقدير معامل انحداره له اكبر عامل تضخم VIF(i).

والحقيقة ان هذه الطريقة في المعالجة وان كانت تؤدي الى زوال مظاهر مشكلة التعدد الخطي الا إنها وفي أغلب الحالات تؤدي الى تحيز كبير في نموذج الانحدار بحيث ان قيمة احصاءه Cp تكون كبيرة فضلاً عن عدم مطابقة النموذج للبيانات بسبب كبر MSE النسبي مما يجعل المعالجة على حساب اختيار النموذج الأفضل وفق المعايير المعروفة كما سنرى ذلك في الجانب التطبيقي.

أضف الى ذلك ان مشكلة التعدد الخطي قد لا تتجسد في معاملات الارتباط البسيط ما بين المتغيرات التفسيرية، أي انها قد تكون موجودة بالرغم من عدم وجود قيم كبيرة لمعاملات الارتباط البسيط ما بين المتغيرات التفسيرية، وهو ما لاحظه العديد من الباحثين [6]. لذلك فان الباحث يرى أن التعريف الدقيق لمشكلة التعدد الخطي يكمن في تحليل المعادلات الطبيعية وتشخيص المركبة التي تسببها والتي أسميناها ((مركبة الانحدار الخطي غير المباشر للمتغير التفسيري)) وهو ما سنعرضه بالتفصيل في مبحث لاحق. وعلى هذا الأساس فان الباحث يعرف مشكلة التعدد الخطي بأنها: ((الحالة التي يكون فيها واحد في الأقل من المتغيرات التفسيرية في نموذج الانحدار الخطي المتعدد يعاني من عدم استقرار تقدير معامل انحداره وتضخم تباينه في آن معاً بما ينعكس في تغير كبير لأهميته الابتدائية نتيجة قوة علاقته الخطية النسبية غير المباشرة بالمتغير المعتمد)). ويعني الباحث بـ((أهميته الابتدائية)) قوة علاقته الخطية بالمتغير المعتمد التي يمثلها معامل الارتباط البسيط ما بين كل متغير تفسيري والمتغير المعتمد ابتداءً عند اختياره والمتغيرات التفسيرية الأخرى لصياغة نموذج الانحدار الوافي (الصحيح) adequate model.

2-1-2 آثار التعدد الخطي على تقديرات المربعات الصغرى لمعاملات الانحدار:

يرى الباحث أنه يمكن تحديد أثر التعدد الخطي على تقديرات معاملات الانحدار بنقطتين هما:

2-1-2-1 تضخم الخطأ المعياري لتقديرات معاملات الانحدار (Sbi) وهو ما ينعكس في لا معنوية هذه التقديرات ومن ثم عدم أهمية المتغيرات التفسيرية بالرغم من أن العلاقة الخطية لكل منها مع المتغير المعتمد هي علاقة قوية تشير إليها معاملات ارتباطها البسيط مع المتغير المعتمد. وتظهر عدم الأهمية هذه في اختبار t حيث الاحصاءة t تساوي:

$$|t| = \left|\frac{bi}{sbi}\right| \text{ under } H0: Bi = 0$$

وتصغر قيمتها كلما تضخم الخطأ المعياري Sbi فيؤدي ذلك الى لا معنوية معامل الانحدار المقدر وعدم أهمية المتغير التفسيري في النموذج. والحقيقة ان Sb_i يتضخم كلما كان مربع معامل الارتباط المتعدد للمتغير X_i مع المتغيرات التفسيرية الاخرى قويا أي كلما كان معامل التحديد $R^2(i)$ كبيرا للنموذج:

$$\hat{X}i = b_o + \sum_{\forall j \neq i} b_j x_j \,; \quad i,j = 1,2,3,\ldots,m$$

ذلك لأن:

$$S_{bi}^2 = \frac{MSE}{sii[1 - R_{(i)}^2]} \text{ where } 0 \le R_{(i)}^2 \le 1$$

وواضح ان المقام يصغر كثيرا كلما اقتربت احدى الكميتين S_{ii} او $1 - R_{(i)}^2$ من الصفر واذا افترضنا ان العينة قد اظهرت التباينات في قيم المتغير X_i فان تضخم تباين معامل انحداره انما يعود الى الكمية الثانية $[1 - R_{(i)}^2]$ عندما تقترب من الصفر وبعبارة اخرى عندما يقترب $R_{(i)}^2$ من الواحد الصحيح حيث عند ذاك يقترب تباين تقدير معامل الانحدار من المالانهاية.

2-1-2-2 عدم استقرار تقديرات معاملات الانحدار:

ان عدم استقرار تقدير معامل الانحدار لاي متغير تفسيري يفقد هذا المتغير اهميته في تفسير الاختلافات في قيم المتغير المعتمد حتى وان كان اثر تضخم تباين تقدير معامله بسيطا، لذلك فان هذين الاثرين -تضخم تقدير معامل الانحدار وعدم استقراره- متلازمان للدلالة على اهمية المتغير التفسيري في النموذج وعلى وجود مشكلة التعدد الخطي. وهكذا فان هذين الاثرين للتعدد الخطي تؤديان بالباحث الى استدلال خاطئ عن اهمية المتغيرات التفسيرية في النموذج.

2 -2 طرائق الكشف عن وجود مشكلة التعدد الخطي:

توجد طرائق عديدة للكشف عن التعدد الخطي في بيانات متغيرات الانحدار (المتغيرات التفسيرية) بعضها يعطي نتائج مؤكدة وبعضها الاخر غير مؤكدة نستعرضها فيما يلي:

2- 2- 1 حساب مصفوفة معاملات الارتباط البسيط بين المتغيرات التفسيرية:

اذ يعتقد بعض الباحثين انه اذا كانت قيم معاملات الارتباط بين المتغيرات التفسيرية كبيرة فذلك مؤشر على وجود مشكلة تعدد خطي والعكس صحيح حيث لا توجد مشكلة تعدد خطي عندما تكـــون تلك القيم صغيرة في مدى الارتباط المعروف. ولكن ثبت بالتطبيق كما اشرنا انفا ان هـــذا المؤشـــر ليس دائما صحيحا اذ قد تكون هناك مشكلة تعدد خطي بالرغم من ان قيم معاملات الارتباط صـــغيرة وذلك لان المشكلة في هذه الحالة لا تتجسد في العلاقة الخطية بين كل متغيـــرين تفســـيريين وانمـــا متغير تفسيري ومجموعة المتغيرات التفسيرية الاخرى معا، مما دفع الباحثين للنظر الى الموضـــوع من هذه الزاوية.

2- 2- 2 حساب مربع معامل الارتباط الخطي المتعدد ما بين المتغيرات التفســـيرية:

ان الفشل الذي واجه الباحثين في الكشف عن التعدد الخطي في بعض التطبيقات باستخدام معـــاملات الارتباط البسيط بين المتغيرات التفسيرية قد دفعهم كما اشرنا للنظر الى الموضوع من زاوية تعـــدد العلاقات الخطية بين كل متغير تفسيري والمتغيرات الأخرى، ولذلك توجه الباحثون الى حساب معامل الارتباط الخطي المتعدد (بدلاً من الارتباط البسيط) لكل متغير تفســـيري مـــع مجموعـــة المتغيـــرات التفسيرية الأخرى، وحددوا باستخدام معامل التحديد والذي هو مربع معامل الارتباط الخطي المتعدد- قيمة الاختلافات المفسرة في قيم كل متغير تفسيري وحيث ان تضخم تباين تقدير أي متغير تفسيري مثل Xi يعتمد على قيمة معامل تحديده $R^2(i)$ كما أشرنا آنفـــاً لـــذلك فـــان $R^2(i)$ أو $1-R^2(i)$ أو $\frac{1}{1-R^2(i)}$ (وهذه الاخيرة يرمز لها باختصار VIF(i) أي عامل تضخم تبـــاين تقـــديرات معـــاملات الانحدار) كلها هي مؤشرات أو مقاييس للتضخم في تقديرات معاملات الانحـــدار ومـــن ثـــم فهـــي مؤشرات على وجود مشكلة التعدد الخطي. واستطراداً في هذا الموضوع فان البعض يحسب متوسط عوامل التضخم في تباينات معاملات الانحدار $\overline{VIF}$ للاستدلال على وجود مشكلة التعدد الخطـــي فـــي النموذج، ذلك لأن عدم وجود تضخم فيها يعني أن قيمة معامل تحديد كـــل متغيـــر تفســـيري $R^2(i)$ تساوي صفر مما يجعل قيمة هذا المتوسط تساوي واحد في هذه الحالة ما دام ان عامـــل التضـــخم يحسب من العلاقة:

$$VIF(i) = \frac{1}{1-R^2(i)} \text{ where } 0 \le R^2 \le 1 \dots\dots\dots (1)$$

فلو فرضنا ان عدد المتغيرات التفسيرية في النموذج يساوي m فيكون متوســـط عوامـــل التضخم يساوي:

لا يوجد تضخم فالمتغيرات متعامدة $\overline{VIF} = \frac{\sum_{i=1}^{m} VIF(i)}{m} = 1$

يوجد تضخم (المتغيرات غير متعامدة) > 1

والحقيقة ان هذا التحليل قد أوحى بحساب مؤشر (معيار) جديد لتقييم تقديرات معاملات الانحدار في النموذج سمي ((متوسط مربع الخطأ الكلي في تقديرات معاملات الانحدار)) الذي صيغته الرياضية:

$$E\left[\left(\hat{\underline{B}} - \underline{B}\right)'\left(\hat{\underline{B}} - \underline{B}\right)\right]$$

والتي يمكن برهان انها تساوي $\hat{\sigma}^2 \sum_{i=1}^{m} VIF(i)$ حيث:

$\hat{\underline{B}}$: هي متجه القيم التقديرية لمعاملات الانحدار بعد تحويل قيم المتغيرات في النموذج الى قيم قياسية (معيارية **Standard**) بصورة X_i^* و Y_i^*.

$$\therefore E\left[\left(\hat{\underline{B}} - \underline{B}\right)'\left(\hat{\underline{B}} - \underline{B}\right)\right] = E[(X^{*\prime}X^*)^{-1}X^{*\prime}][Y^* - E(Y)^*][Y^{*\prime} - E(Y^*)][X^*(X^{*\prime}X^*)^{-1}]$$

$$= E[e'e(X^{*\prime}X^*)^{-1}] = \hat{\sigma}^2 \, Trace\,(X^{*\prime}X^*)^{-1} = \hat{\sigma}^2 \, Trace\, R^{-1}$$

حيث R^{-1} هي معكوس مصفوفة معاملات الارتباط ما بين المتغيرات التفسيرية.

$$\therefore E[(\hat{\underline{B}} - \underline{B})'(\hat{\underline{B}} - \underline{B})] = \hat{\sigma}^2 \sum_{i=1}^{m} VIF(i)$$

وذلك لأن عناصر قطر المصفوفة R^{-1} هي قيم عوامل تضخم تباينات تقديرات معاملات الانحدار، وهذا يعني ان متوسط مربع الخطأ الكلي في تقديرات معاملات الانحدار المحسوبة بطريقة المربعات الصغرى يكون أقل ما يمكن عندما يكون $\sum_{i=1}^{m} VIF(i) = m$ حيث **m** هي عدد المتغيرات في النموذج أي عندما تكون المتغيرات متعامدة **orthogonal** ومن ثم تكون تقديرات المربعات الصغرى لمعاملات الانحدار أكثر دقة وأقرب الى قيمها الحقيقية (المجهولة). ويمكن الاستفادة من هذه الحقيقة في تحديد قيمة نسبية لقياس انحراف تقدير المربعات الصغرى لمعلمة الانحدار عن قيمتها الحقيقية كما يلي:

$$\because E\left[\left(\hat{\underline{B}} - \underline{B}\right)'\left(\hat{\underline{B}} - \underline{B}\right)\right] = \hat{\sigma}^2 \sum_{i=1}^{m} VIF(i) = m\hat{\sigma}^2$$

المتغيرات متعامدة ولا يوجد تضخم

$$\therefore \frac{\hat{\sigma}^2 \sum_{i=1}^{m} VIF(i)}{\hat{\sigma}^2 m} = \frac{\sum_{i=1}^{m} VIF(i)}{m} = \overline{VIF}$$

وبذلك يكون $\overline{VIF}$ متوسط عوامل تضخم تباينات تقديرات معاملات الانحدار في النموذج ليس مؤشراً أو معياراً لوجود مشكلة التعدد الخطي وحسب، وانما هو ايضاً مؤشـرا أو معيـار لدقـة تقـديرات المربعات الصغرى لمعلمات إنحدار النموذج وأساس جيد للمفاضلة ما بين نماذج الانحـدار الخطـي المختلفة. وسوف نعتمده في المفاضلة كمعيار الى جانب المعايير الأخرى المعروفة وذلك في الجانب التطبيقي من هذا البحث.

3-2-2 حساب الجذور المميـزة أو الكامنـة characteristic (Latent) roots:

وتسمى أيضاً **Eigen Values** وهي قيم تباينات لمتغيرات جديدة تستحدث باعتبار كل منها توليفـة خطية **linear combination** من المتغيرات التفسيرية الأصلية، تدعى هذه المتغيـرات الجديـدة بالعوامل أو المكونات الرئيسية **principal components** وتتميز بكونها متعامدة **orthogonal** أي ان احلالها بدلاً من المتغيرات التفسيرية الأصلية من شأنه تجـاوز مشـكلة العلاقـات الخطيـة المتعددة بين قيم المتغيرات التفسيرية. ان قيم تباينات هذه العوامل أو المكونات يرمز لها عادة $\lambda i's$ ويمكن الحصول عليها بخطوتين:

1. ايجاد مصفوفة الارتباط ما بين المتغيرات التفسيرية الأصلية **R**.
2. حل المعادلة المميزة الآتية بالنسبة الى λ:

$|R - \lambda I| = 0$ وهي معادلة متعددة الحدود **polynomial** في λ من الدرجة **m** أي انها:

$$\lambda^m + Cm - 1\lambda^{m-1} + \dots\dots + C1\lambda + C0 = 0$$

ان حل هذه المعادلة يعطينا هذه القيم $\lambda i's$ المسماة بالجذور المميزة وعددها **m** بحيث ان:

$\lambda_1 > \lambda_2 \dots\dots > \lambda m$ وان $\lambda i \geq 0$، وان $\sum_{i=1}^{m} \lambda i = m$ وان $\sum_{i=1}^{m} VIF(i) = \sum_{i=1}^{m} \frac{1}{\lambda_i}$ وهذا يعنـي ان أية قيمة لـ λi تقترب من الصفر من اليمين تجعل المجموع والمتوسط كبير جداً مما يدل علـى وجود تضخم في تباينات تقديرات معاملات الانحدار ومن ثم فهي اشارة واضحة على وجود مشـكلة التعدد الخطي. ولذلك تستخدم هذه الجذور المميزة $\lambda i's$ في الكشف عن وجود مشكلة التعدد الخطـي بصورة قاطعة وأكيدة.

4-2 معالجة أو كيفية التعامل مع مشكلة التعدد الخطـي Multicollineartiy-Treatment:

توجد طرائق مختلفة لمعالجة مشكلة العلاقات الخطية المتعددة بين قيم المتغيرات التفسيرية، رائدها جميعاً

هو اضعاف هذه العلاقات الخطية بحيث تصبح غير مؤثرة في أهمية المتغيرات التفسيرية عندما تكون معاً في نموذج الانحدار الخطي، وأهم هذه الطرائق ما يلي:

2-4-1 زيادة حجم العينة العشوائية: في بعض الأحيان تكون مشكلة العلاقات الخطية المتعددة ناجمة عن نقص في بيانات العينة، لذلك تكون المعالجة بزيادة حجم العينة وذلك بجمع بيانات اضافية تؤدي الى اضعاف العلاقات الخطية بين المتغيرات التفسيرية ولكن ذلك ليس ممكناً في أغلب الأحيان بسبب القيود على الميزانية المعدة للبحث، وعامل الوقت المحدد لانجازه، والجهد الإضافي المبذول فضلاً عن الكادر الذي يقوم بذلك.

2-4-2 استبدال المتغير التفسيري الذي له معامل ارتباط خطي متعدد عال مع المتغيرات التفسيرية الأخرى: ويتم اللجوء الى هذا الخيار (في حالة توفر المتغير البديل) عندما تكون علاقات هذا المتغير الخطية مع المتغيرات التفسيرية الأخرى غير ناشئة عن نقص في بيانات العينة وانما هي علاقات خطية داخلية متأصلة في طبيعة هذا المتغير، ولذلك يلجأ الباحث في هذه الحالة الى البحث عن متغير يعتقد بأن ليس له علاقات خطية قوية مع المتغيرات التفسيرية الأخرى وفي الوقت نفسه يرتبط بعلاقة خطية قوية مع المتغير المعتمد وذلك لكي يحافظ نموذج الانحدار على قوته التفسيرية ومطابقته للبيانات. ولكن مع الأسف فان اللجوء الى هذا الخيار ليس متاحاً في الغالب لأنه يتطلب الحصول على متغير بديل بمواصفات معينة وإمكانية جمع البيانات عنه وفق الحجم المحدد للعينة العشوائية.

2-4-3 حذف المتغير التفسيري الذي له مربع معامل ارتباط خطي متعدد $R^2(i)$ أو عامل تضخم VIF(i) كبير: وهذه هي أسهل طريقة لمعالجة آثار التعدد الخطي على معنوية تقديرات المربعات الصغرى لمعاملات الانحدار (المتبقية في النموذج)، ولذلك يلجأ إليها العديد من الباحثين عندما تواجههم مشكلة العلاقات الخطية المتعددة.. ولكن هذه الطريقة وهي تعالج هذه المشكلة تخلق مشكلة جديدة تتمثل بتحيز معادلة الانحدار المتحققة وعدم كفاية نموذجها Inadequate model بعد ان حذف منه متغير أو متغيرات تفسيرية (مهمة) ذات ارتباط خطي قوي ومن ثم عدم قدرته على تفسير ومطابقة البيانات بسبب كبر متوسط مربعات البواقي فيه وكبر احصاءة Mallow التي من المفترض أن تكون أصغر ما يمكن بسبب حذف عدد من متغيرات النموذج الكلي وبعبارة أخرى أن قيمتها يجب ان تكون قريبة جداً من عدد المعلمات المتبقية في النموذج كما هو واضح من صيغة حساب إحصاءة Cp-Mallow:

$$Cp = \frac{SSEp}{S^2} - (n - 2p) \ldots\ldots\ldots (3)$$

S^2: متوسط مربعات الخطأ (MSE) للنموذج الكلي- بجميع المتغيرات التفسيرية.

p: عدد معلمات النموذج.

SSEp: مجموع مربعات البواقي (الخطأ) للنموذج الذي يحتوي P من المعلمات.

n: حجم العينة.

2-4-4 استخدام تحليل المكونات الرئيسية Principal Components Analysis:

حيث ان المشكلة هي في عدم تعامد قيم متغيرات الانحدار (المتغيرات التفسيرية) لذلك فان الحل يكون- كما يراه Hotelling- باستحداث متغيرات انحدار جديدة تتميز بتعامدها orthogonal variables ويتحقق ذلك عندما يكون كل متغير تفسيري جديد هو توليفه خطية Linear Combination من المتغيرات التفسيرية الأخرى [7]. فلو فرضنا أن المتغيرات التفسيرية الأصلية هي X_1 وX_2 وX_3 وX_4 والمتغيرات الجديدة هي Z_1 وZ_2 وZ_3 وZ_4 فان:

$$Z_1 = a_{11}X_1 + a_{21}X_2 + a_{31}X_3 + a_{41}X_4 \; ; \; Z_2 = a_{12}X_1 + a_{22}X_2 + a_{32}X_3 + a_{42}X_4 \; ; \; Z_3 = a_{13}X_1 + a_{23}X_2 + a_{33}X_3 + a_{43}X_4 \; ; \; Z_4 = a_{14}X_1 + a_{24}X_2 + a_{34}X_3 + a_{44}X_4$$

ان معاملات المتغيرات الجديدة أو المكونات الرئيسة (aij) يتم تقديرها بحيث تجعل تباين المكون الرئيسي الأول Z_1 هو الأكبر، وهذا يتحقق عندما يكون مجموع مربعات aij يساوي واحد أي عندما يتحقق الشرط: $a'a = a_1^2 + a_2^2 + a_3^2 + a_4^2 = 1$ وحيث أن لكل جذر مميز λi مرافق هو متجه قياسي مميز عناصره هي العناصر aij التي تحقق الشرط المذكور لذلك فان تقدير قيم هذه العناصر يكون بعد حساب الجذور المميزة وذلك بحل المعادلة $|R - \lambda I|a = 0$ بالنسبة الى aij. وحيث ان هذه الطريقة تكشف عن التعدد الخطي من خلال صغر قيمة واحد في الأقل من الجذور المميزة في المدى (1،0) فان أحد متغيراتها الجديدة في الأقل يعد مرادفاً للصفر لذلك يمكن حذفه وحساب دالة الانحدار بدلالة المتغيرات الجديدة المتبقية وهي متغيرات متعامدة لا تعاني من مشكلة التعدد الخطي. ولكن يعاب على هذه الطريقة اندثار ((هوية)) المتغيرات التفسيرية الأصلية- اذا صح التعبير- وذلك لحلول متغيرات جديدة مكانها لا يمكن انتسابها الى متغير تفسيري بعينه لكونها توليفة خطية من جميع المتغيرات التفسيرية الأصلية، وبذلك يصعب تفسير تقديرات معاملات انحدارها باعتبار كل منها هو الميل الحدي لتغير المتغير التفسيري الجديد وليس الأصلي، ولذلك فان أغلب الباحثين يستخدم هذه الطريقة لغرض الكشف عن مشكلة التعدد الخطي وحسب.

2-4-5 استخدم تحليل انحدار الــ Ridge (Ridge Regression Analysis):

في العام 1970 اقترح Hoerl & Kennard [8] طريقة جديدة لتقدير معاملات الانحدار في نموذج الانحدار الخطي المتعدد عند وجود مشكلة التعدد الخطي، وذلك باضافة مصفوفة قطرية بثابت تتحدد قيمته في المدى [1، 0] الى مصفوفة القيم المعيارية للمتغيرات التفسيرية بمـــا يـــؤدي الـــى تحولها الى مصفوفة متعامدة orthogonal وبذلك يتم تجاوز مشكلة العلاقات الخطية المتعددة. فاذا فرضنا ان Z هو المتغير التفسيري (المعياري) وان y هو المتغير العشوائي (المعتمد) وان K هــو الثابت المضاف فان معاملات انحدار الــ Ridge تقدر بالعلاقة:

$$\underline{b}(k) = (Z'Z + KI)^{-1}Z'y \text{ where } E\left(\underline{b}(k)\right) \neq \underline{B}$$

وواضح من هذه العلاقة انه عندما [k = 0] نحصل علـــى تقـــديرات المربعـــات الصـــغرى لمعاملات الانحدار

$$\underline{b}(0) = (Z'Z)^{-1}Z'y \text{ where } E\left(\underline{b}(0)\right) = \underline{B}$$

أي ان تقديرات الــ Ridge تكون (متحيزة) بمقدار قيمة الثابت k المضاف، وانها توليفة خطية من تقديرات المربعات الصغرى (غير المتحيزة) لقيم المتغيرات التفسيرية المعيارية نفسها حيث [9]:

$$\underline{b}(k) = [I + k(Z'Z)^{-1}]^{-1}\underline{b}(0)$$

وربما يجد الكثيرون في طريقة انحدار الــ Ridge حلاً مقبـــولاً لمشـــكلة العلاقـــات الخطيـــة المتعددة والحصول في الوقت نفسه على تقديرات دقيقة لمعاملات الانحدار في النمـــوذج وان كانـــت (متحيزة) وذلك لتميز النموذج بأن له أصغر متوسط لمربعات الخطأ وبذلك يكون تفسيره ومطابقتـــه للبيانات جيدة. وهذا صحيح- في رأي الباحث- لو كانت مصـــفوفة معـــاملات الارتبـــاط مـــا بـــين المتغيرات التفسيرية المستخدمة في ايجاد تقديرات المربعات الصغرى (المعيارية) هي نفسها- دون تغيير- المصفوفة المستخدمة في ايجاد تقديرات الــ Ridge، ولكن الحقيقة هي ان هذه المصـــفوفة قد تغيرت باضافة مصفوفة التحيز KI اليها بحيث أصبح معامل ارتباط المتغير مع نفسه لا يســـاوي 1 وانما يساوي k + 1 ؟! ولذلك فالتحيز في انحدار الــ Ridge يختلف عن التحيز فـــي انحـــدار المربعات الصغرى للنموذج الجزئي، من حيث كون الأول (تحيز الــ Ridge) هو عامل خارجي قدتم ادخاله على مصفوفة معاملات الارتباط ما بين المتغيرات التفسيرية بحيث لا يمكن ازالته حتى لو تم اختيار المتغيرات (الصحيحة) في نموذج الانحدار، بينما (التحيز) في تقديرات المربعـــات الصـــغرى للنموذج الجزئي هو نتيجة عرضية داخلية لاسقاط أحد المتغيرات (المهمة) من النموذج، وهو يمكن

قياسه باستخدام احصاءة Cp- Mallow كما يمكن تقليله الى حد كبير في المــدى (1، 0) بحيــث يمكن اهماله- كما سنرى في تطبيق طريقتنا المقترحة.

2-5 المعيار والطريقة المقترحان للكشف عن التعدد الخطي وتجاوز آثاره:

2-5-1 المعيار المقترح للكشف عن التعدد الخطي وأهميته: نفرض ان لدينا متغيرين تفسيريين X_1 وX_2 ومتغير معتمد هو y يرتبطان بعلاقة الانحدار الخطي:

$$Yi = B_0 + B_1X_1i + B_2X_2i + ei \; ; \; i = 1, 2, \ldots\ldots, n$$

فأنه- وكما هو معلوم- باستخدام طريقة المربعات الصغرى الاعتيادية OLS تتحقــق لنــا المعادلات الطبيعية الآتية:

$$\sum Y = nbo + b_1\sum X_1 + b_2\sum X_2$$
$$\sum X_1 Y = b_0\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_2$$
$$\sum X_2 Y = b_0\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2$$

ومن المعادلة الأولى يكون:

$$b_0 = \bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2$$

وبتعويض قيمة b_0 في المعادلة الثانية يكون:

$$\sum X_1 Y = n\bar{X}_1\left(\bar{Y} - b_1\bar{X}_1 - b_2\bar{X}_2\right) + b_1\sum X_1^2 + b_2\sum X_1X_2$$

وبفتح القوس وتجميع الحدود المتشابهة يكون:

$$\underset{\displaystyle S_1Y}{\underset{\downarrow}{\sum X_1 Y - n\bar{X}_1\bar{Y}}} = \underset{\displaystyle b_1S_{11}}{\underset{\downarrow}{b_1\left(\sum X_1^2 - n\bar{X}_1^2\right)}} + \underset{\displaystyle b_2S_{12}}{\underset{\downarrow}{b_2\left(\sum X_1X_2 - n\bar{X}_1\bar{X}_2\right)}}$$

$$\therefore b_1 = \frac{S_1Y - b_2S_{12}}{S_{11}} \Rightarrow b_1 = \frac{S_1Y}{S_{11}} - b_2\frac{S_{12}}{S_{11}}$$

وبنفس الطريقة عند تعويض قيمة b_0 في المعادلة الطبيعية الثالثة نحصل على:

$$S_2Y = b_2S_{22} + b_1S_{12}$$

$$\therefore b_2 = \frac{S_2Y - b_1S_{12}}{S_{22}} \Longrightarrow b_2 = \frac{S_2Y}{S_{22}} - b_1\frac{S_{12}}{S_{22}}$$

وبصورة عامة لو كان لدينا m من التغيرات التفسيرية فان:

$$S_1Y = b_1S_{11} + b_2S_{12} + b_3S_{13} + \cdots + b_mS_{1m}$$

$$S_2Y = b_1S_{12} + b_2S_{22} + b_3S_{23} + \cdots + b_mS_{2m}$$

$$S_3Y = b_1S_{13} + b_2S_{23} + b_3S_{33} + \cdots + b_mS_{3m}$$

$$\downarrow$$

$$S_mY = b_1S_{1m} + b_2S_{2m} + b_3S_{3m} + \cdots + b_mS_{mm}$$

أي انه ولكل متغير تفسيري Xi يكون:

$$SiY = biSii + \sum_{\forall j \neq i} bjSij \; ; i, j = 1, 2, \ldots \ldots \ldots, m$$

$$\therefore b_i = \frac{Siy - \sum_{\forall j \neq i} bjsij}{sii} \Rightarrow bi = \frac{siy}{sii} + \frac{-\sum_{\forall j \neq i} bjsij}{Sii} \quad i, j = 1, 2, \ldots, m \_\_\_\_\_\_\_\_ (4)$$

ويتضح من العلاقة الأخيرة (4) ان علاقة الانحدار الخطي ما بين أي متغير تفسيري Xi ومتغير معتمد Y في نموذج انحدار خطي متعدد يمكن تجزئتها الى مركبتين هما:

1. مركبة العلاقة الخطية المباشرة ما بين Xi و Y بمعامل انحدار خطي بسيط وليكن bi^* حيث:

$$bi^* = \frac{Siy}{Sii} \; ; i = 1, 2 \ldots \ldots, m$$

$$SiY = \sum XiY - \frac{\sum Xi \sum Y}{n} \; ; Sii = \sum X^2 i - \frac{(\sum Xi)^2}{n}$$

2. مركبة العلاقة الخطية غير المباشرة ما بين Xi و Y وذلك عن طريق علاقة Xi بـ Xj حيث Xj تمثل المتغيرات التفسيرية الأخرى في النموذج $(j \neq i)$ التي لها علاقة خطية بالمتغير المعتمد Y، ولتكن di حيث:

$$di = \frac{-\sum_{j \neq i} bjSij}{Sii} = -\sum_{\forall j \neq i} bjbij \; ; i, j = 1, 2, \ldots \ldots, m$$

وبذلك يكون:

$$bi = b^*i + di \Rightarrow di = bi - b^*i \ldots \ldots (5)$$

حيث:

bi: معامل الانحدار الجزئي للمتغير Xi في نموذج الانحدار الخطي المتعدد لY على المتغيرات التفسيرية.

b^*i: معامل الانحدار البسيط (المباشر) للمتغير التفسيري Xi في نموذج انحدار Y على X.

bij: معامل الانحدار البسيط (المباشر) للمتغير التفسيري Xi في نموذج انحدار Xj على Xi.

وواضح أنه عندما لا توجد علاقات خطية ما بين المتغيرات التفسيرية أي عندما تكون هذه المتغيرات متعامدة orthogonal فان [di = 0] اذ لا توجد علاقة انحدار غير مباشرة للمتغير المعتمد على أي من المتغيرات التفسيرية وبذلك يكون $R^2(i)$ لكل متغير تفسيري يساوي صفر فلا يوجد تضخم في تباين تقديرات معاملات انحدار هذه المتغيرات، هذا من جهة، ومن جهة أخرى فان [di = 0] يعني ان $bi = bi^*$ أي ان المتغير التفسيري يكون مستقراً تماماً. ولكن عدم استقرار تقدير معامل انحدار المتغير التفسيري عند تحوله من معامل انحدار بسيط الى معامل انحدار جزئي قد لا يكون بسبب تضخم تباينه أي لا يكون بسبب كبر قيمة معامل تحديده $R^2(i)$ وانما بسبب خطأ التحيز في التقدير نتيجة عدم التعيين الصحيح لنموذج الانحدار، لذلك يجب التفريق ما بين هذين النوعين من عدم الاستقرار، فعدم الاستقرار الناجم عن خطأ التحيز هو مؤشر (صحي) وليس مرضي لأنه لا يؤدي الى تضخم تباين تقدير معاملات الانحدار. ولذلك فان أعراض ((مرض)) العلاقات الخطية المتعددة multicollinearity هي في تلازم ظهور عدم استقرار معامل انحدار المتغير التفسيري وتضخم تباينه وهو ما ينعكس في اضعاف أهمية المتغير التفسيري التي قد تصل الى حد عدم معنويته في نموذج الانحدار كما قد يشير الى ذلك اختبار t. ان أهمية المركبة di (مركبة علاقة الانحدار غير المباشرة) هي في كونها مجموعاً جبرياً لحدود مشتركة من علاقة الانحدار ما بين المتغيرات التفسيرية bij من جهة وبينها وبين المتغير المعتمد bj من جهة أخرى، لذلك فما يبدو من (ضعف) مثلاً في العلاقات الخطية ما بين المتغيرات التفسيرية التي تكشف عنها مصفوفة قيم معاملات الارتباط البسيط والتي توحي بعدم وجود مشكلة التعدد الخطي قد يتحول الى (قوة) في النتيجة النهائية للمجموع الجبري مما يعطي إشارة واضحة ولكن معاكسه تفيد بأن عدم استقرار معامل الانحدار هو مؤشر على وجود التعدد الخطي واذا انعكس ذلك في عدم معنويته فيكون هذا المتغير مرشحاً للحذف لعدم اهميته في النموذج. وهذا هو السبب الذي يجعل مؤشر قيم معاملات الارتباط البسيط ما بين المتغيرات التفسيرية مؤشراً خادعاً في بعض الأحيان على وجود أو عدم وجود مشكلة التعدد الخطي. أضف الى ذلك ان المركبة di مترافقة مع $R^2(i)$ تصبح مؤشراً على تغير

(تناقص) الأهمية الابتدائية للمتغير التفسيري اذا ما وضعت في صيغة نسبية تسمح بمقارنة التغير في أهمية المتغيرات التفسيرية وذلك كما يلي:

$$\frac{bi - bi^*}{bi^*} = \frac{di}{bi^*}$$

وبذلك يكون هذا التغير النسبي لتقدير معامل الانحدار هو عامل عدم الاستقرار النسبي لمعامل الانحدار وليكن رمزه اختصاراً Relative Unstability Factor :RUF وهو في الوقت نفسه يكون مؤشراً على التغير النسبي للأهمية الابتدائية للمتغير التفسيري وذلك لأن الأهمية الابتدائية لأي متغير تفسيري مثل Xi هي في قوة علاقته الخطية بالمتغير المعتمد Y أي في قيمة معامل ارتباطه البسيط riy وعندما يكون في نموذج انحدار متعدد تتمثل هذه الأهمية المقارنة بمعامل انحداره المعياري $\hat{Bi}$ (Bettacoefficient) لذلك يكون التغير النسبي لأهمية المتغير التفسيري يساوي:

$$\frac{\hat{Bi} - riy}{riy}$$

وحيث ان:

$$\hat{Bi} = bi\left(\frac{sii}{syy}\right)^{\frac{1}{2}} \;\&\; riy = bi^*\left(\frac{sii}{syy}\right)^{\frac{1}{2}}$$

$$\therefore \frac{Bi - riy}{riy} = \frac{\left(\frac{sii}{syy}\right)^{\frac{1}{2}}[bi - bi^*]}{\left(\frac{sii}{syy}\right)^{\frac{1}{2}} bi^*} = \frac{bi - bi^*}{bi^*} = \frac{di}{bi^*}$$

وبذلك يكون RUF ليس معياراً أو مؤشراً لعدم الاستقرار النسبي وحسب وإنما هو معيار أو مؤشر لتغير الأهمية الابتدائية للمتغير التفسيري أي ان:

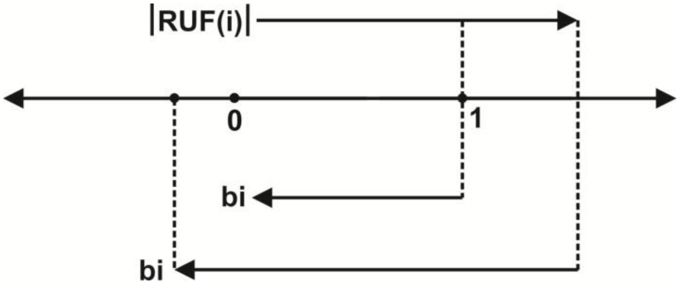
$$RUF(i) = \frac{\hat{Bi} - riy}{riy} = \frac{di}{bi^*} \dots\dots(6)$$

وحيث ان أغلب حزم البرامج الاحصائية المتوفرة حالياً مثل SPSS وSAS تعرض الى جانب قيم معاملات الانحدار الجزئية، قيم معاملات الانحدار المعيارية ومصفوفة معاملات الارتباط البسيط ما بين متغيرات النموذج لذلك لا حاجة لحساب syy و Siy ∀ i = 1, 2, 3,....., m تمهيداً لحساب bi^* وانما استخدام قيم معاملات الانحدار المعيارية وقيم معاملات ارتباط كل متغير تفسيري مع المتغير المعتمد والمتوفرة عادة لحساب عامل الاستقرار النسبي في العلاقة (6).

وبالرجوع الى مركبتي معامل الانحدار الجزئي في العلاقة (5) نلاحظ أن تناقص أهمية المتغير التفسيري في النموذج والذي تحدده قيمة عامل عدم الاستقرار النسبي RUF(i) السالبة في العلاقة (6) يتحقق عندما تكون المركبتان (المباشرة وغير المباشرة) باتجاهين (متعاكسين) حيث يظهر عامل عدم الاستقرار النسبي RUF(i) سالباً لهذا السبب وتزداد سالبيته عن الواحد الصحيح عندما تكون القيمة المطلقة لمركبة الانحدار غير المباشر di (أكبر) من القيمة المطلقة لمركبة الانحدار المباشر (البسيط) bi^* حيث يؤدي ذلك الى تغير اشارة تقدير معامل انحدار المتغير التفسيري في نموذج الانحدار الخطي المتعدد. ان هذا التغير في اشارة معامل انحدار المتغير التفسيري يجعل من وجوده في النموذج سبباً في زيادة مجموع مربعات البواقي وكذلك متوسط مربعات البواقي عن طريقين:

1. حذف مجموع مربعات انحداره الجزئي من مجموع مربعات الانحدار الكلية، وذلك لأن مجموع مربعات انحداره الجزئي [bisiy] أو [iriyB̂] سيكون سالباً.
2. نقصان درجات حرية تقدير متوسط مربعات البواقي.

ويترتب على ذلك زيادة الأخطاء المعيارية لتقدير معاملات الانحدار فينعكس ذلك في اختبار t بعدم معنويتها وعدم أهمية متغيراتها في النموذج. ويمكن تمثيل العلاقة العكسية بين معامل الانحدار الجزئي وعامل عدم استقراره النسبي على مستقيم الأعداد كما يلي:



حيث يلاحظ في المدى [0، 1] ان قيمة |RUF(i)| ⟵ 1 فأن قيمة bi ⟵ صفر وبذلك فان المتغير التفسيري Xi يصبح غير مهم في نموذج الانحدار الخطي المتعدد كلما اقترب |RUF(i)| من الواحد الصحيح. بينما في المدى [0، ∞] فان قيمة |RUF(i)| ⟵ +∞ فان قيمة bi ⟵ − ∞ وبذلك يكون:

عندما |RUF(i)| > 1 فان bi < صفر أي باتجاه معاكس لاتجاه أصل علاقة المتغير التفسيري Xi بالمتغير المعتمد مما يجعله قطعاً غير مهم بل (ومعرقل) لمطابقة نموذج الانحدار الخطي للبيانات المشاهدة. وهكذا يكون عامل عدم الاستقرار النسبي RUF(i) هو العامل الحاسم في

بقاء أو حذف المتغير التفسيري من النموذج وصولاً الى النموذج الجزئي الأفضل وفق المعايير المعروفة، وليس عامل تحديد المتغير التفسيري $R^2(i)$ ولا عامل تضخم تباين تقدير معامل انحداره $VIF(i)$ كما يعتقد البعض ويوصي بمعالجة مشكلة التعدد الخطي على أساسه، وهو ما سنبينه تطبيقياً في الجانب التطبيقي من هذا البحث. نخلص من ذلك الى ان معيار التعدد الخطي هو معيار تغير الأهمية الابتدائية للمتغير التفسيري في نموذج الانحدار الخطي المتعدد عندما تكون متغيراته غير متعامدة، وهو نفسه عامل عدم الاستقرار النسبي RUF المتلازم مع معامل التحديد للمتغير التفسيري $R^2(i)$ أو عامل تضخم التباين $VIF(i)$ والذي ينعكس في اختبار (t) بعدم أهمية ذلك المتغير. وهذا يعني ان $R^2(i)$ أو $VIF(i)$ لكل متغير تفسيري X_i يمكن قياس أثرهما على عدم استقرار معامل انحداره وعدم معنويته بحساب $RUF(i)$ واختبار $|t|$ معاً.

ولكي نضع $RUF(i)$ متلازماً مع اختبار $|t|$ بصورة معيار للتعدد الخطي وأداة تستخدم للوصول الى النموذج الجزئي الأفضل ينبغي تحديد نقطة حرجة تمثل الحد الأدنى المطلوب للاشارة الى وجود مشكلة العلاقات الخطية المتعددة. وحيث ان $RUF(i)$ الناجم عن التعدد الخطي في المدى [0، 1] يشير الى عدم أهمية المتغير التفسيري كلما ابتعدت قيمته المطلقة عن الصفر باتجاه الواحد الصحيح، لذلك يقترح الباحث أن تكون قيمة الحد الأدنى لـ $|RUF(i)|$ هي 0.5 وبذلك يكون معيار الكشف عن مشكلة التعدد الخطي وأداة الباحث في تجاوز آثارها وصولاً لنموذج الانحدار الأفضل هي الآتية:

$$|RUF(i)| \geq 0.5 \;\&\; |t| < t\left(\frac{\alpha}{2}, n-p\right) \text{-------------(7)}$$

حيث α: مستوى المعنوية الذي يحدده الباحث لغرض اختبار فرضية عدم معنوية كل متغير تفسيري.

P: عدد معلمات النموذج.

n: حجم العينة العشوائية.

2-5-2 الطريقة المقترحة لتجاوز آثار التعدد الخطي وتحقيق أفضل نموذج انحدار خطي (جزئي):

The procedure to overcoming the multicollinearity-impact and achieving the best sub-set regression model:

لا بد من الاشارة أولاً الى أن هذه الطريقة في تجاوز آثار التعدد الخطي واختيار نموذج الانحدار الأفضل انما تقوم على أساس أن متغيرات الانحدار (المتغيرات التفسيرية) هي في الأساس متغيرات (مهمة) أي ان معاملات ارتباطها البسيط مع المتغير المعتمد ((قوية)) نسبياً وهي التي أسميناها

بالأهمية الابتدائية للمتغيرات التفسيرية، وأن ما يظهر من عدم أهميتها في نموذج الانحدار المتعدد الذي يكشف عنه اختبار |t| سببه تضخم تباين تقديرات معاملات انحدارها أي ان سببه هو مشكلة العلاقات الخطية المتعددة. لقد تم تصميم خطوات تجاوز آثار التعدد الخطي واختيار متغيرات الانحدار ونموذجها الأفضل على أساس توفر كل المعلومات المطلوبة في حزمة البرامج الاحصائية المعروفة كبرنامج SPSS الموجود في جميع الجامعات العراقية وعلى أساس سهولة وسرعة حساب المعيار المقترح في (2-5-1) من هذا البحث، وهذه الخطوات هي كما يلي:

<u>الدورة الأولى:</u>

1. **من النموذج الكلي (بجميع متغيرات الانحدار) ندخل البيانات في برنامج الانحدار الخطي لـ SPSS مثلاً، فنحصل على قيم معاملات الانحدار المعيارية وقيم مصفوفة معاملات الارتباط البسيط بين كافة المتغيرات في النموذج، التي نحسب منها المعيار المقترح RUF(i) لكل متغير تفسيري Xi وفق العلاقة (6) من البحث (2−5−1).**
2. **نحدد مستوى المعنوية لاختبار معامل انحدار كل متغير تفسيري وليكن 0.05 ونضع النتائج الآتية في الجدول كما يلي:**

المتغيرات التفسيرية في النموذج	معاملات انحدارها bi	أهميتها الابتدائية riy	معاملاتها المعيارية Bi	عوامل عدم استقرارها RUF(i)	عوامل تضخم تبايناتها VIF(i)*	معنوية معاملات الانحدار
x_1	b_1	r_1y	B_1	$RUF_{(1)}$	$VIF_{(1)}$	
x_2	b_2	r_2y	B_2	$RUF_{(2)}$	$VIF_{(2)}$	
↓	↓	↓	↓	↓	↓	
x_m	bm	r_my	B_m	$RUF_{(m)}$	$VIF_{(m)}$	

3. **نبحث عن المتغيرات التي تحقق معيار وجود مشكلة التعدد الخطي المشار اليه في العلاقة (7) ونحذف من بينها المتغير الذي له (أكبر) قيمة مطلقة لعامل عدم الاستقرار النسبي [RUF(i) السالبة].**

---

* ان ادراج قيم هذا العمود أو عدم ادراجها لا يؤثر على خطوات العمل وانما هي للتأكد من تضخم تباين تقديرات معاملات الانحدار.

الدورة الثانية:

نكرر الخطوتين (1) و(2) للنموذج الجديد (الجزئي) ونبحث عن المتغيرات التي تحقق معيار وجود مشكلة التعدد الخطي كما في الدورة الأولى فان لم نجد نتوقف ونعتبر هذا النموذج هو النموذج الأفضل وقد زالت منه كل آثار التعدد الخطي. وبخلاف ذلك نستمر ونحذف المتغير الذي له أكبر قيمة مطلقة لعامل عدم الاستقرار النسبي... وهكذا نستمر في حذف المتغيرات حتى نصل الى مجموعة جزئية من المتغيرات التفسيرية لا تحقق معيار وجود مشكلة التعدد الخطي فنتوقف عند ذاك ونعتبر نموذجها هو النموذج الأفضل.

3. الجانب التطبيقي Applied Part:

سوف نطبق المعيار المقترح لتحقيق أفضل نموذج جزئي للانحدار – في حالة وجود علاقات خطية متعددة – على المسألة التي عرضها A. hald في كتابه الموسوم Statistical Theory with Engineering Applications للأسباب الآتية [10]:

ان هذه المسألة تعرض صعوبات نموذجية عادة ما تحدث في تحليل الانحدار – كما يصفانها Drapper & Smith في الصفحة 296 من كتابهما Applied Regression Analysis وان هذه المسألة قد استخدم في حلها مختلف طرائق تقدير معاملات الانحدار ومختلف طرائق الحصول على أفضل نموذج جزئي للانحدار وفي أكثر من مصدر احصائي، مما يوفر فرصة ذهبية للمقارنة ما بين الطريقة التي يقترحها الباحث وبين هذه الطرائق، من حيث النتائج وسهولة الوصول اليها وسرعة انجازها. ولكي لا يتوسع هذا الجانب أكثر مما يجب ومع عدم الاخلال بالهدف من اختيار هذه المسألة نقارن طريقتنا المقترحة بالطرائق الثلاثة المعروفة وهي [11]:*

1. طريقة الاختيار المتقدم (الأمامي) Forward Selection Procedure.
2. طريقة الحذف التراجعي Backward Elimination Procedure.
3. طريقة النهج الحكيم Step-Wise Procedure.

* قيم F– الجزئية للمتغيرات التفسيرية في تطبيق الطرائق الثلاثة أخذت من كتاب خاشع الراوي ((المدخل الى تحليل الانحدار)).

فضلاً عن طريقة حذف المتغير التفسيري على أساس قوة ارتباطه الخطي مع المتغيرات التفسيرية الأخرى أو تضخم تباين تقدير معامله. وتكون معايير المفاضلة اضافة الى سهولة وسرعة الوصول الى النتائج النهائية باستخدام برامج الانحدار المتوفرة أو بدونها، هي الآتية:
MSE: متوسط مربعات الخطأ (البواقي)؛ R^2: معامل التحديد؛ R^2a: معامل التحديد المعدل؛ **Cp**: احصاءة **Mallow**؛ **Cp-p**: مقدار التحيز في تقديرات المربعات الصغرى؛ $\overline{VIF}$: معيار انحراف تقديرات المربعات الصغرى لمعاملات الانحدار عن قيمها الحقيقية بسبب تضخم تبايناتها (وجود مشكلة التعدد الخطي).

جدول رقم (1)

بيانات مسألة Hald

المشاهدات	y	x_1	x_2	x_3	x_4
1	78.5	7	26	6	60
2	74.3	1	29	15	52
3	104.3	11	56	8	20
4	87.6	11	31	8	47
5	95.9	7	52	6	33
6	109.2	11	55	9	22
7	102.7	3	71	17	6
8	72.5	1	31	22	44
9	93.1	2	54	18	22
10	115.9	21	47	4	26
11	83.8	1	40	23	34
12	113.3	11	66	9	12
13	109.4	10	68	8	12

3-1 تطبيق طريقة الاختيار المتقدم Forward selection:

نلخص الخطوات والنتائج بالجدول الاتي:

مراحل التطبيق	F- الجزيئة للمتغيرات	FIN	المتغير الداخل الى النموذج
الدورة الاولى: (الانحدار على كل متغير تفسيري)	$F_{(1)}=12.60$ $F_{(2)}=21.96$ $F_{(3)}=4.40$ $F_{(4)}=22.80$	F(0.05, 1, 11)=4.48	X_4
الدورة الثانية: الانحدار على المتغيرين (x_j,x_4) حيث j= 3, 2,1	$F_{(1)}=108.22$ $F_{(2)}=0.1725$ $F_{(3)}=40.29$	F(0.05, 1, 10)=4.96	X_1

الـــدورة الثالثـــة: الانحـــدار علـــى المتغيرات (x_j,x_4,x_1) حيث j= 2, 3	$F_{(2)}=5.02$ $F_{(3)}=4.23$	F(0.05, 1, 9)=5.12	—

∴ نموذج الانحدار الجزئي الافضل هو بدلالة x_4,x_1 حيث معادلة الانحدار الخطي المتعدد في نهائيـة الدورة الثانية هي:

$$\hat{y} = 103.097 + 1.440X_1 - 0.614X_4 ;\ MSE = 7.476 ;\ R^2 = 0.972 ;\ R_a^2 = 0.964$$

3-2 تطبيق طريقة الحذف التراجعي Backward Elimination:

نلخص الخطوات والنتائج بالجدول الاتي:

مراحل التطبيق	F- الجزئية للمتغيرات	Fout	المتغير الخارج من النموذج
الـــدورة الاولـــى: الانحـــدار علـــى جميـــع المتغيرات التفسيرية معا.	$F_{(1)}=4.43$ $F_{(2)}=0.50$ $F_{(3)}=0.02$ $F_{(4)}=0.04$	F(0.05, 1, 8)=5.32	X_3
الـدورة الثانيـة: الانحـدار علـى المتغيـرات الباقية (x_4,x_2,x_1) معا.	$F_{(1)}=154.01$ $F_{(2)}=5.03$ $F_{(4)}=1.00$	F(0.05, 1, 9)=5.12	X_4
الـدورة الثالثـة: الانحـدار علـى المتغيـرات (x_2,x_1) معا.	$F_{(1)}=146.52$ $F_{(2)}=208.58$	F(0.05, 1,10)=4.96	—

∴ نموذج الانحدار الجزئي الافضل هو بدلالة x_2, x_1 حيث معادلة الانحدار الخطي:

$$\hat{y} = 52.577 + 1.468X_1 + 0.662X_2 ;\ MSE = 5.7905 ;\ R^2 = 0.979 ;\ R_a^2 = 0.975$$

3-3 تطبيق طريقة النهج الحكيم Step-wise procedure:

نلخص الخطوات والنتائج في الجدول الاتي:

مراحل التطبيق	F- الجزئية للمتغيرات	FIN = Fout	المتغير الداخل الى النموذج	المتغير الخارج
الدورة الاولى: الانحدار على كل متغير تفسيري.	$F_{(1)}=12.60$ $F_{(2)}=21.96$ $F_{(3)}=4.40$ $F_{(4)}=22.80$	F(0.05, 1,11)=4.48	X_4	—
الدورة الثانية: الانحدار على المتغيرين (x_j,x_4) حيث j=1,2,3.	$F_{(1)}= 108.22$ $F_{(2)}=0.1725$ $F_{(3)}=40.29$ $F_{(4)}=159.930$	F(0.05, 1,10)=4.96	X_1	—
الدورة الثالثة: الانحدار على المتغيرات (x_j,x_4,x_1) حيث j= 2,3.	$F_{(2)}=5.02$ $F_{(3)}=4.23$	F(0.05, 1,9)=5.12	—	—

∴ نموذج الانحدار الجزئي الافضل هو بدلالة x_4, x_1 حيث معادلة الانحدار الخطي المتعدد هي:

$$\hat{y} = 103.097 + 1.440X_1 - 0.614X_4 \; ; \; MSE = 7.476 \; ; \; R^2 = 0.972 \; ; \; R_a^2 = 0.964$$

3-4 تطبيق طريقة الحذف على اساس اكبر قوة مربع ارتباط خطي متعدد ما بين المتغيرات التفسيرية $R^2_{(i)}$:

مراحل التطبيق	$R^2(i)$	المتغير الخارج من النموذج
الدورة الاولى: لانحدار على جميع المتغيرات التفسيرية معا. حدار كل متغير تفسيري على المتغيرات التفسيرية الاخرى.	$R^2_{(1)} = 0.974$ $R^2_{(2)} = 0.996$ $R^2_{(3)} = 0.979$ $R^2_{(4)} = 0.996$	X_2 or X_4 نفرض X_2
الدورة الثانية: لانحدار على بقية المتغيرات التفسيرية معا (x_4,x_3,x_1). حدار كل متغير تفسيري في هذه الدورة على غيره من المتغيرات التفسيرية معا.	$R^2_{(1)} = 0.728$ $R^2_{(3)} = 0.711$ $R^2_{(4)} = 0.153$	X_1
الدورة الثالثة: لانحدار على المتغيرات التفسيرية (x_4,x_3) معا. حدار x_3 على x_4 وبالعكس.	$R^2_{(3)} = 0.001$ $R^2_{(4)} = 0.001$	—

∴ النموذج الجزئي الافضل للانحدار هو النموذج بدلالة (x_4,x_3) ومعادلة الانحدار الخطي المتعدد:

$$\hat{y} = 131.282 - 1.200X_3 - 0.725X_4 \; ; MSE = 17.574 \; ; \; R^2 = 0.935 \; ; \; R_a^2 = 0.9$$

22

وجدير بالذكر انه اذا اختار الباحث وفق هذه الطريقة حذف x_4 بدلا من x_2 في الدورة الاولى فانه سيضطر ايضا الى حذف x_1 في الدورة الثانية وبذلك يكون النموذج الجزئي الافضل هو بدلالة x_3, x_2 وتكون معادلة الانحدار الخطي المتعدد هي:

$$\hat{y} = 72.075 + 0.731X_2 - 1.008X_3 \; ; \; MSE = 41.5443 \; ; \; R^2 = 0.84703 \cong 0.85$$

3-5 تطبيق الطريقة المقترحة:

طريقة عامل الاستقرار النسبي RUF-procedure.

<u>الدورة الأولى</u>: First cycle:

1. نحسب انحدار y على جميع المتغيرات التفسيرية معا (x_1 x_2 x_3 x_4) باستخدام أي برنامج انحدار مثل SPSS فتكون معادلة الانحدار الناتجة هي:

$$\hat{y} = 62.405 + 1.551X_1 + 0.510X_2 + 0.102X_3 - 0.144X_4 ;\ MSE = 5.983 ;\ R_y^2 = 0.982 ;R_a^2 = 0.974 ;F = 111.479$$

2. نحسب قيم عامل عدم الاستقرار النسبي للمتغيرات التفسيرية بتطبيق العلاقة (6):

$$RUF_{(i)} = \frac{\bar{B}_i - riy}{riy} \ ;i = 1, 2, 3, 4$$

حيث: $\bar{B}_i$: معاملات Betta (معاملات الانحدار المعيارية)

riy: معاملات الارتباط البسيط بين كل متغير تفسيري x_i والمتغير المعتمد y.

3. نحدد مستوى المعنوية لرفض فرضية العدم: Ho: Bi=0 والتي هي عادة [α=0.05] وتكون الجدول الاتي:

Xi المتغيرات التفسيرية	bi معاملات انحدارها	$\bar{B}i$ معاملات Betta (المعيارية)	riy معاملات الارتباط البسيطة	RUF (i) عامل عدم الاستقرار النسبي	VIF (i) عامل تضخيم التباين	Sign معنوية معامل الانحدار
X_1	1.551	0.607	0.731	-0.170	38.462	0.501
X_2	0.510	0.528	0.816	-0.353	250	0.896
X_3	0.102	0.043	-0.535	-1.080	47.619	0.071
X_4	-0.144	-0.160	-0.821	-0.805	250	0.844

نحدد من الجدول المتغيرات التفسيرية التي معاملاتها انحدارها غير معنوية وعامل عدم استقرارها اكبر او يساوي 0.5 كقيمة مطلقة أي التي تحقق معيار عدم الاستقرار النسبي ووجود مشكلة التعدد الخطي فنجدها جميع هذه المتغيرات وعندها نحذف منها المتغير x_3 لان القيمة المطلقة لعامل عدم استقراره النسبي هي الاكبر.

<u>الدورة الثانية</u> second cycle: نكرر الخطوات نفسها فنحصل على معادلة الانحدار والجدول الاتي:

$$\hat{y} = 71.648 + 1.452X_1 + 0.416X_2 - 0.237X_4 \ ;\ MSE = 5.330 ;\ R_y^2 = 0.982 ;R_a^2 = 0.976 ;\ F = 166.832$$

Xi المتغيرات التفسيرية	bi معاملات انحدارها	$\bar{B}i$ معاملات Betta (المعيارية)	riy معاملات الارتباط البسيطة	RUF (i) عامل عدم الاستقرار النسبي	VIF (i) عامل تضخيم التباين	Sign معنوية معامل الانحدار
X_1	1.452	0.568	0.731	-0.222	1.066	0.000
X_2	0.416	0.430	0.816	-0.473	18.868	0.052
X_4	-0.237	-0.263	-0.821	-0.680	18.868	0.205

يلاحظ من الجدول في هذه الدورة ان X_4 فقط يحقق معيار عدم الاستقرار النسبي ووجود مشكلة التعدد الخطي لذلك نحذف X4 من النموذج.

الدورة الثالثة Third cycle:

نكرر الخطوات كما في الدورتين الاولى والثانية فنحصل على معادلة الانحدار والجدول الاتي:

$$\hat{y} = 52.577 + 1.468\, X_1 + 0.662\, X_2$$

$$; MSE = 5.790\ ; R_y^2 = 0.979\ \cong 0.98\ ; R_a^2 = 0.974\ ; F = 229.504$$

Xi المتغيرات التفسيرية	bi معاملات انحدارها	$\bar{B}i$ معاملات Betta (المعيارية)	riy معاملات الارتباط البسيطة	RUF (i) عامل عدم الاستقرار النسبي	VIF (i) عامل تضخيم التباين	Sign معنوية معامل الانحدار
X_1	1.468	0.574	0.731	-0.214	1.055	0.000
X_2	0.662	0.685	0.816	-0.161	1.055	0.000

وحيث ان معاملات انحدار المتغيرات الباقية $(X_2,\ X_1)$ في النموذج قد اصبحت معنوية لـذلك نتوقف عن الحذف ونعتبر نموذج الانحدار بدلالـة (X_2, X_1) هـو النمـوذج الافضـل فـيلاحظ ان المتغيرات الباقية في نموذج الانحدار وفق هذه الطريقة تتميز بما يلي:

1. قيمة عامل عدم استقرارها النسبي صغيرة في المدى [1 , 0] أي انها مستقرة نسبيا.
2. قيمة عامل تضخم معاملات انحدارها صغيرة جدا في المدة نفسه [1 , 0].
3. قيم معاملات Betta (المعيارية) كبيره نسبيا في المدة نفسه [1 , 0] مما يدل علـى اهميتهـا فـي النموذج.

4. تقييم نتائج تطبيق طرائق اختيار متغيرات الانحدار لأفضل نموذج جزئي: كما اشرنا فـي بدايـة الجانب التطبيقي من هذا البحث فان معايير المفاضلة بين الطرائق المستخدمة ستكون – اضافة الـى سهولة وسرعة الوصول الى النتائج النهائية – ما يلي:

MSE: متوسط مربعات البواقي – R_a^2 معامل التحديد – R_y^2 معامل التحديد المعدل – C_p- احصاءة Mallow ومقدار التحيز C_p-p – VIF: انحراف التقديرات بسبب تضخم تبايناتها.

ولحساب قيمة احصاءة Mallow ومقدار التحيز في كل نموذج جزئي نتج عـن تطبيـق الطرائـق المذكورة، نعوض بالعلاقة رقم (3) في المبحث 2-4. واما حساب قيم VIF فيتطلب حسـاب قيمـة $R^2(i)$ لانحدار كل متغير تفسيري على المتغيرات التفسيرية الاخرى الباقية فـي النمـوذج الجزئـي النهائي ثم تطبيق العلاقة رقم (1) المذكورة في المبحث 2–2. والجدول الاتي يجمع نتائج حساب كل مقاييس المفاضلة لمختلف الطرائق.

جدول رقم (2)

قيم معايير المفاضلة ما بين النماذج الجزئية للانحدار التي افرزتها

مختلف طرائق اختيار متغيرات الانحدار

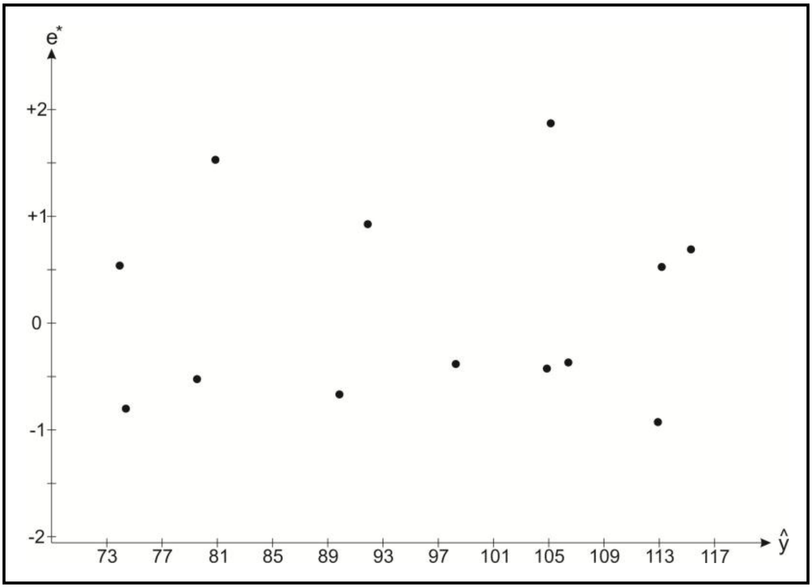
الطريق للطبقة	متغيرات النموذج	MSE	R_y^2	R_a^2	C_p	C_p-p	$\overline{VIF}$

	الجزئي						
1. Forward selection	$X_1\&X_4$	7.48	0.97	0.97	5.50	2.5	1.064
2. Rack ward elimination	$X_1\&X_2$	5.79	0.98	0.97	2.68	0.32	1.055
3. Step wise	$X_1\&X_4$	7.48	0.97	0.97	5.50	2.5	1.064
4. R_i^2[VIFi] Elimination	$X_3\&X_4$	17.57	0.94	0.92	22.37	19.37	1.001
5. RUF Procedure	$X_1\&X_2$	5.79	0.98	0.97	2.68	0.32	1.055

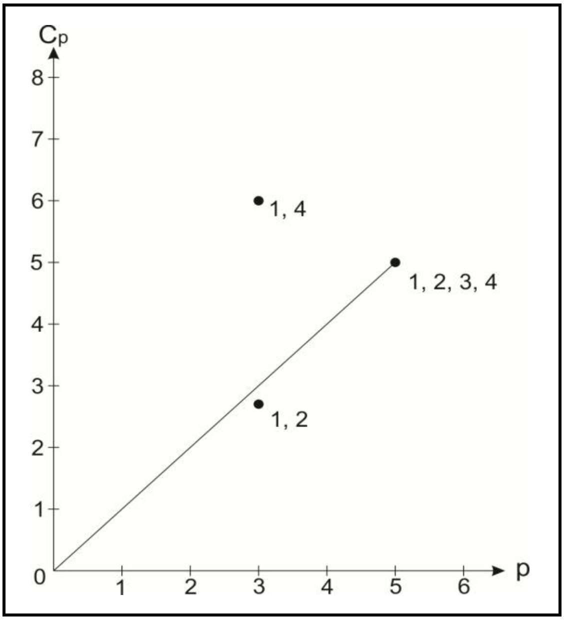
وواضح من هذا الجدول ان النموذج الجزئي الافضل هو النموذج المتحقق بطريقتنا المقترحة طريقة عامل عدم الاستقرار النسبي RUF، وكذلك طريقة الحذف التراجعي Back ward elimin وذلك لتميز ذلك النموذج باصغر MSE واصغر C_p واكبر R_y^2 و R_a^2 فضلا عن اقل تحيز في المدى [1 , 0] وعدم تضخم في تباين تقديرات معلمات انحداره مما يجعلها اقرب الى قيمها الحقيقية في المجتمع، فضلا عن ذلك فان طريقتنا المقترحة تتميز عن طريق الحذف التراجعي وكل الطرائق الاخرى بسهولة وسرعة الوصول الى نتائجها. كما يتضح من الجدول ان النموذج الجزئي (الاسوء) للانحدار هو الذي حققته طريقة الحذف على اساس قوة ارتباط المتغير التفسيري بالمتغيرات التفسيرية الاخرى [VIF(R_i^2)-Elimin.] حيث كان له اكبر MSE واكبر C_p ومن ثم اكبر تحيز وابتعاد عن النموذج الوافي adequate-model. وباستخدام نموذج الانحدار الجزئي الافضل وفق معايير المفاضلة – وهو النموذج الذي تحقق باستخدام الطريقة المقترحة للباحث – نحسب الاخطاء المعيارية e^*_i وقيم y المتنبأ بها $(\hat{y})$ – لاحظ الجدول رقم (3) ونرسم النقاط المشتركة $(e_i^*, \hat{y}i)$، كما نرسم C_p – لاحظ الشكلين (1) و (2) على التوالي فنزداد تأكداً من ان هذا النموذج الجزئي للانحدار هو النموذج الصحيح لانه لا يخرق الفرضيات الاساسية المتعلقة بحد الخطأ ويحقق في الوقت نفسه من خلال النقطة (C_p, P): (2.68, 3) اقصى اقتراب من الخط $C_p = p$.

جدول رقم (3): القيم التنبؤية لـــ (y) والاخطاء المعيارية لنموذج الانحدار $\hat{y} = f(X_1, X_2)$

المشاهدة	$\hat{y}i$	$e_i^* = \frac{ei}{s}$	المشاهدة	$\hat{y}i$	$e_i^* = \frac{ei}{s}$
1	80.074	-0.654	8	74.575	-0.862
2	73.251	0.436	9	91.275	0.758
3	105.815	-0.629	10	114.538	0.566
4	89.258	-0.689	11	80.536	1.357
5	97.293	-0.579	12	112.437	0.359
6	105.152	1.682	13	112.293	-1.202
7	104.002	-0.541			



شكل رقم (1) الرسم البياني لــ e_i^* مقابل $\hat{y}i$ لمسألة Hald في النموذج الجزئي $\hat{y} = f(X_1, X_2)$



شكل رقم (2) الرسم البياني لــ C_p من بيانات Hald ويلاحظ عدم ظهور نقطة الدالة $\hat{y} = f(X_3, X_4)$ وذلك لتحيزها العالي جدا

جدول رقم (4)

قيم معايير المفاضلة لجميع معادلات الانحدار المتحققة بكل الانحدارات الممكنة(*)

عدد المعلمات المقدرة بالمعادلة	معادلة الانحدار	MSE	R_y^2	C_p	$\lvert C_p - p \rvert$	$\overline{VIF}$
1	$\hat{y} = b_0$	226.3136	0	442.92	441.92	0
2	$\hat{y} = b_0 + b_1X_1$	115.0624	0.53395	202.55	200.55	0
	$\hat{y} = b_0 + b_2X_2$	82.3942	0.66627	142.49	140.49	0
	$\hat{y} = b_0 + b_3X_3$	176.3092	0.28587	315.16	313.16	0
	$\hat{y} = b_0 + b_4X_4$	80.3515	0.67459	138.73	136.73	0
3	$\hat{y} = b_0 + b_1X_1 + b_2X_2$	5.7904	0.97868	2.68	0.32	1.055
	$\hat{y} = b_0 + b_1X_1 + b_3X_3$	122.7073	0.54817	198.10	195.10	3.115
	$\hat{y} = b_0 + b_1X_1 + b_4X_4$	7.4762	0.97247	5.50	2.5	1.064
	$\hat{y} = b_0 + b_2X_2 + b_3X_3$	41.5443	0.84703	62.44	59.44	1.019
	$\hat{y} = b_0 + b_2X_2 + b_4X_4$	86.8880	0.68006	138.23	135.23	18.868
	$\hat{y} = b_0 + b_3X_3 + b_4X_4$	17.5738	0.93529	22.37	19.37	1.001
4	$\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3$	5.3456	0.98228	3.04	0.96	2.485
	$\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_4X_4$	5.3303	0.98234	3.02	0.98	12.934
	$\hat{y} = b_0 + b_1X_1 + b_3X_3 + b_4X_4$	5.6485	0.98128	3.50	0.50	2.774
	$\hat{y} = b_0 + b_2X_2 + b_3X_3 + b_4X_4$	8.2017	0.97282	7.34	3.34	16.500
5	$\hat{y} = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4$	[illegible]982	0.98238	5.00	0	146.520

والان ماذا يحدث اذا غير الباحث مستوى المعنوية المحدد للاختبار ؟

يلاحظ انه اذا اخترنا مستوى معنوية ادق من 0.05 وليكن 0.01 واستخدمناه في اختبار معنوية معاملات انحدار المتغيرات التفسيرية في كل طريقة من الطرائق السابقة فان ذلك لا يؤثر في اختيار متغيرات الانحدار التي تشكل افضل نموذج انحدار خطي بينما اذا لم نتشدد في اختيار مستوى المعنوية بحيث اخترنا قيمته 0.10 أي رضينا باحتمال الوقوع في الخطأ من النوع الاول وهو احتمال

* قيم معايير المفاضلة الثلاثة الاولى (C_p, R_y^2, MSE) وردت في كتاب Drapper & Smith Applied regression Analysis ص298 و301 وكذلك في كتاب (المدخل الى تحليل الانحدار) للاستاذ الدكتور خاشع الراوي ص266، اما قيم المعيارين الاخيرين فقد تم حسابها من قبل الباحث.

رفض فرضية العدم وهي صحيحة بقيمة 0.10 فان نتائج بعض الطرائق ستتغير كما في الجدول الاتي.

جدول رقم (5)

متغيرات النموذج الجزئي الافضل حسب الطريقة المطبقة

عندما يتغير مستوى معنوية الاختبار

المتغيرات في النموذج عندما $\alpha = 0.10$	المتغيرات في النموذج عندما α=0.05 او 0.01	الطريقة المطبقة
x_1, x_2 & x_4	x_1 & x_4	Forward Selection
x_1 & x_2	x_1 & x_2	Back ward Elimination
x_1 & x_2	x_1 & x_4	Step Wise
x_3 & x_4	x_3 & x_4	$R^2_{(i)}(VIF)$ - Elimination
x_1 & x_2	x_1 & x_2	RUF - Procedure

ويلاحظ من الجدول رقم (5) ان طريقة **Stepwise** قد تحسنت وفقا لمعايير المفاضلة في الوصول الى النموذج الجزئي الافضل والذي هو بدلالة x_1 & x_2 فقط وذلك عندما ارتفع احتمال الوقوع في الخطأ من النوع الاول، الى 0.10، وكذلك تحسنت طريقة **Forward selection** حيث انخفضت قيمة MSE الى 5.33 وازداد R^2_a, R^2_y الى 0.98 وصغرت قيمة C_p بحيث اصبحت 3.02 ومقدار التحيز الى 0.98، ولكن هذه الطريقة وباحتمال 0.10 قد ابتعدت فيها تقديرات معلمات الانحدار عن الدقة بنسبة 13 مرة تقريبا وذلك بسبب تضخم متوسط تبايناتها(*) اما طريقتنا المقترحة **RUF-procedure** وكذلك طريقة الحذف التراجعي **Backward Elimination** فقد حافظت على مستواها الافضل وفي كل الاحتمالات.

5. الاستنتاجات والتوصيات:

أ. الاستنتاجات:

(*) لاحظ هذه النتائج في الدورة الثانية من استخدام طريقتنا المقترحة RUF-Procedure حيث معادلة الانحدار بدلالة x_4, x_2, x_1 او في جدول رقم (4) عندما عدد المعلمات المقدرة يساوي 4.

بعد هذا الاستعراض والتطبيق للطرائق الشائعة والطريقة التي اقترحها الباحث للكشف عن مشكلة العلاقات الخطية المتعددة في بيانات المتغيرات التفسيرية ومعالجتها، واخضاع هذه الطرائق جميعها لنفس معايير المفاضلة: نتوصل الى الاستنتاجات التالية:

1. ان معيار او مؤشر الكشف عن وجود مشكلة العلاقات الخطية المتعددة باستخدام عامل الاستقرار النسبي RUF واختبار $|t|$ هو اسهل وادق المعايير او المؤشرات فضلا عن كونه اداة بيد الباحث لاختيار متغيرات الانحدار التي تبقى في النموذج، وعلى اساس علمي رصين.
2. ان الطريقة التي اقترحها الباحث باسم طريقة عامل عدم الاستقرار النسبي RUF procedure هي الطريقة الفضلى – في حالة وجود مشكلة التعدد الخطي – التي تضمن اختيار اهم المتغيرات معا في معادلة الانحدار الخطي المتعدد مع اقل تحيز وادق تقدير لمعاملات الانحدار.
3. ان طريقة الحذف التراجعي Backward Elimination تتفوق على طريقة الاختبار المتقدم Forward selection والطريقة المعدلة لها Step-wise procedure في حالة وجود مشكلة العلاقات الخطية المتعددة.
4. ان التخلص من مشكلة العلاقات الخطية المتعددة بطريقة حذف متغير تفسيري واحد او اكثر على اساس قوة ارتباطه الخطي مع المتغيرات التفسيرية الاخرى هي طريقة غير صحيحة ليس لانها لا تحقق الهدف المطلوب وانما لانها تقود الى مشكلة جديدة وكبرى تتمثل في الاختيار السيء للمتغيرات الباقية في معادلة الانحدار الخطي المتعدد، مما يجعل المعادلة الناتجة متحيزة جدا وبعيدة عن مطابقة البيانات وذلك لكبر متوسط مربعات الخطأ فيها وكبر إحصاءة Mallow قياسا بعدد المعلمات المقدرة بها.
5. ان جميع الطرائق سواء التي تستخدم المربعات الصغرى الاعتيادية OLS في تقدير معاملات الانحدار او التي تحولت عنها الى طرائق اخرى (المكونات الرئيسية P.C، انحدار الـ Ridge) انما تضحي بجانب معين من خصائص المقدرات الجيدة ذات المقدرة على تفسير علاقة الانحدار، وذلك وصولا الى تجاوز مشكلة العلاقات الخطية المتعددة.

ب- التوصيات:

1. يوصي الباحث باستخدام طريقته المقترحة RUF-procedure في حالة وجود مشكلة العلاقات الخطية المتعددة ما بين المتغيرات التفسيرية (المهمة) وذلك لسهولتها وقدرتها على الوفاء بكل معايير المفاضلة تقريبا.
2. يوصي الباحث بعدم استخدام طريقة Step-wise (ولا طريقة Forward selection طبعا) في حالة وجود مشكلة العلاقات الخطية المتعددة، وفي حالة استخدام 0.05 فاقل، كمقدار لاحتمال الوقوع في الخطأ من النوع الاول، لان ذلك يؤدي الى اختيار غير دقيق لمتغيرات الانحدار المهمة معا في النموذج مما يجعل تقديرات معاملات الانحدار متحيزة وغير دقيقة.
3. يوصي الباحث بعدم استخدام طريقة حذف متغيرات الانحدار على اساس قوة ارتباطها الخطي مع بعضها، لانها تقود الى مشكلة جديدة وكبرى تتمثل في الاختيار السيء للمتغيرات الباقية في معادلة الانحدار الخطي المتعدد مما يجعل المعادلة الناتجة متحيزة جدا وبعيدة جدا عن مطابقة البيانات.

المصادر References:

1. سامبريت جاترجي وبيرترام برايس، "تحليل الانحدار بالامثلة" ترجمة محمد مناجد الدليمي، مطابع التعليم العالي في الموصل، 1990، ص182.
2. BRICE. BOWER MAN & RICHARD T: D'CONNELL, "Applied statistics: Improving Business processes'' Richard D.Irwin, a times mirror higher education group, Inc, company, 1997, P.837.
3. جورج او – ويسولوسكي، "الانحدار المتعدد وتحليل التباين" ترجمة د.شلال حبيب الجبوري، مطابع التعليم العالي في الموصل، 1990، ص73.
4. دومينيك سالفاتور، "نظريات ومسائل في الاحصاء والاقتصاد القياسي" ترجمة د. سعيدة حافظ منتصر، دار ماكجروهيل للنشر، 1990، ص210.
5. N.R. Draper & H. Smith, "Applied Regression Analysis'' 2nd Edition, John wiley & Sons, 1981, p. 258.
6. سامبريت جاترجي ويرترام برايس، مصدر سابق، ص199.
7. Harold Hotelling "Analysis of a complex of statistical variables in to principal components'' Journal of Educational Psychology, 24, 1994, PP. 417-441 & 489-520.
8. A.E Hoerl & R.W. Rennard, ''Ridge regression: biased estimation for nonorthogonal problems'' Tech nometrics, 12, 1970, pp. 55-67.
9. N.R. Draper & H. Smoth, ''Ibid" p.314.

10. A. Hald, "statistical theory with engineering applications", wiley, New York, 1952, P. 647.

11. الراوي – خاشع محمود، "المدخل الى تحليل الانحدار" دار الكتب للطباعة والنشر، جامعة الموصل، 1987، ص275–294.