68

The Multi Collinearity Problem — ثانياً - مشكلة التعدد الخطي (الازدواج الخطي)

The meaning of Multi Collinearity — مفهوم مشكلة التعدد

أول من أشار إلى هذه المشكلة هو الاقتصادي النرويجي (راكنر فريش). إن أحد الافتراضات الأساسية لنموذج الانحدار الخطي المتعدد هو عدم وجود ارتباط تام بين قيم مشاهدات المتغيرات المستقلة الداخلة في نموذج الانحدار المراد تقدير معلماته.

$r_{x_i x_j} \neq 1$

وفي حالة عدم صحة هذا المتغيرات المستقلة لا قيمة واحد أو أكثر من المتغيرات المستقلة في النموذج تحت الدراسة ينتفي الافتراض الخاص بعدم الارتباط فتظهر مشكلة تدعى بالتعدد الخطي أو الازدواج الخطي.

فإذا كان معامل الارتباط الخطي مساوياً للواحد أي أن الارتباط بين المتغيرات التوضيحية تكون ارتباطاً تاماً فيصبح من المستحيل الحصول على تقديرات لكل معلمة (معامل) بشكل منفصل فتكون (OLS) غير مجدية.

$r_{x_i x_j} = 1$

وتحصل مشكلة التعدد الخطي عندما يرتبط اثنان أو أكثر من المتغيرات التوضيحية المستقلة بعلاقة خطية قوية جداً تكون من الصعب فصل أثر كل متغير من المتغيرات المستقلة. وتجدر الإشارة هنا إلى أن التعدد الخطي في العلاقات الاقتصادية لا يشكل مشكلة إلا إذا تجاوزت شدته درجة معينة، حيث تكون هنالك علاقة بين المتغيرات التوضيحية وذلك نتيجة لتأثير المتغيرات التوضيحية ببعضها البعض.

إن الارتباط الخطي المتعدد ليس حالة أحادية أو شاذة في المجال الاقتصادي ولكنه نوعاً ما ظاهرة ملازمة في أغلب العلاقات نظراً لطبيعة العلاقات الاقتصادية، وأحياناً يوجد ارتباط الانحدار الخطي المتعدد ولكنه ليس مرتفعاً فلا يكون مؤثراً.

وتواجه مشكلة التعدد الخطي حينما تكون فيه احد المتغيرات المستقلة مساوية لكافة لكل هذه، او عندما تعتمد فيه الي المتغيرات المستقلة له قيمه واحدا والترمز المتغيرات المستقلة من النموذج المدروس، علماً ان هذه المشكلة تواجه الباحث سواء في ظل فرضية ثبات التباين او عدم ثبات التباين وسواء كانت البيانات بشكل السلاسل الزمنية او المقطعية.

ويتضح مما تقدم اعلاه بأن فرضية الكامل التعدد الخطي يتلخص بما يلي: لا توجد علاقة خطية تامة او شبه تامة بين اي من المتغيرات المستقلة، اضافة الا ذلك يجب ان تكون عدد المعالم المطلوب تقديرها اقل من حجم العينة كما يلي

$$\operatorname{rank}(X) = f(X) = K+1 < n$$

ويعتقد الدكتور klein ان مشكلة التعدد الخطي تحدث عندما

$$r_{x_i x_j} > R_y$$

$$r_{x_i x_j} > \sqrt{R^2}$$

ثالثاً أسباب حدوث ظاهرة التعدد الخطي

مصادر

تنشأ التعدد الخطي من عدة اسباب هي الاتي:

١ – اتجاه المتغيرات الاقتصادية للتغير مع مرور الوقت عن الظروف لتأثرها لنفس العوامل، مثال ذلك في فترة الازدهار والنمو حيث يتحول من الدخل والاستهلاك والادخار والاستثمار والاسعار والاستيراد، والعكس في حالة الركود او الكساد، ان حركة هذه العوامل لاتجاه الزمني في السلاسل الزمنية هي السبب في هذا الارتباط.

٢ – استخدام قيم متباطئة (متخلفة زمنيا) لبعض المتغيرات التوضيحية كمتغيرات توضيحية في تشخيص النماذج الذي يعطي نتائج جيدة

فإن ذلك سيتيح الشكل التالي والشكل السابق كمتغيرات توضيحية سوية مادامت لها الاستقلال الكلي ومن الطبيعي ان القيم المتطابقة لاي متغير احصائي تظهر توجهاً للارتباط ، ولهذا السبب تظهر مشكلة الارتباط الخطي المتعدد .

الفترة الحالية   الفترة السابقة

$$C_t = b_0 + b_1 \cancel{X_t} + b_2 \cancel{X_{t-1}} + U_t$$

وإذا كانت

$$X_1 = 2X_2 \quad \text{or} \quad X_1 = X_2^2$$

فان ذلك يؤدي الى وجود مشكلة التعدد الخطي .

نتائج او الآثار المترتبة على الارتباط الخطي المتعدد Consequences of Multicollinearity

1 - وجود ارتباط عالي بين متغيرين مستقلين يجعل من الصعب معرفة تأثير كل متغير منفرداً على المتغير التابع، كما انه قد تظهر المشكلة على الرغم من تكون R^2 مرتفعة وبالرغم من معلمات النموذج المقدرة معنوية احصائياً وهنا نحتاج الى معالجة هذه الظاهرة .

2 - ان وجود ارتباط شديد بين المتغيرات المستقلة في النموذج يجعل قيمة محدد المصفوفة $|X'X|$ قريبة من الصفر ولذلك فان قيمة التباين والتباين المشترك تصبح كبيرة جداً

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

$$var-cov(\hat{b}) = \sigma^2 \frac{1}{|X'X|} adj(X'X)$$

3 - ان زيادة حجم التباين تؤدي الى ان تقديرات المربعات الصغرى (OLS) غير دقيقة وان معنوية معالم النموذج (t) تكون غير دقيقة فقد تؤشر لعدم معنوية معالم المتغيرات المستقلة في النموذج

$$t = \frac{b}{S(\hat{b})}$$

4 - وجود علاقة طردية بين معامل الارتباط بين المتغيرات المستقلة وبين قيمة التباين حيث تقترب قيمته من ∞ عندما تقترب قيمة الارتباط من الواحد الصحيح

ثالثاً: انواع التعدد الخطي
Taypes of multicollinearity

1- غياب التعدد الخطي Absence of multicollinearity

تحدث هذه الحالة عندما لا ترتبط المتغيرات المستقلة (X_i) في نموذج الانحدار الخطي المتعدد ارتباطاً خطياً وبالتالي فان المصفوفة $(X'X)$ تصبح مصفوفة قطرية تأخذ الشكل الاتي

$$(X'X) = \begin{bmatrix} \Sigma X_1^2 & 0 & 0 & \cdots & 0 \\ 0 & \Sigma X_2^2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \Sigma X_k^2 \end{bmatrix}$$

وان مقلوب المصفوفة سيكون قطرياً ايضاً وللعناصر المناظرة في المصفوفة $(X'X)$ وستصبح المعلمات $(\hat{B})$ بموجب الصيغة الاتية

$$b_i = \frac{\Sigma x_i y}{\Sigma x_i^2} \qquad (i = 1, 2, \ldots, k)$$

وبالتالي لا تتعقد حالة نموذج الانحدار المتعدد، حيث تقدر الثابت $\hat{a}$ وفقاً للصيغة

$$\hat{a} = \bar{Y} - \hat{b}_i \bar{X}_i \qquad (i = 1, 2, \ldots, k)$$

ويتضح ان هذه الصيغة مطابقة لانحدار Y البسيط لكل متغير من المتغيرات المستقلة X_i بشكل منفصل.

2- التعدد الخطي التام Exact multicollinearity

ان حالة التعدد الخطي التام هي حالة مثالية غير ممكنة التحقق في الواقع العملي، ذلك لانها تتحقق فقط في حالة وجود علاقة تامة بين اثنين او اكثر من المتغيرات المستقلة.

ولتوضيح هذه الحالة نفترض وجود نموذج انحدار خطي متعدد يحتوي على متغيرين مستقلين كالاتي

$$Y_i = a + b_1 X_1 + b_2 X_2 + U_i$$

والذي تمثلت تقديره وفقاً للصيغة الآتية

$$\beta = (X'X)^{-1}(X'Y)$$ باستخدام (OLS)

حيث محدد المصفوفة $(X'X)$ هو

$$|X'X| = \begin{vmatrix} \sum X_1^2 & \sum X_1X_2 \\ \sum X_2X_1 & \sum X_2^2 \end{vmatrix} \neq 0$$

$$\sum X_1^2 \sum X_2^2 - (\sum X_1X_2)^2 \neq 0$$

فلو وجدت علاقة خطية تامة بين قيم كل من هذين المتغيرين المستقلين X_1, X_2 يمكن التعبير عنها كالآتي

$$X_2 = KX_1$$

حيث أن K عبارة عن قيمة ثابتة لا تساوي صفر. فالمحدد $|X'X|$ سيكون مساوياً للصفر أي أن

$$|X'X| = \begin{vmatrix} \sum X_1^2 & K\sum X_1^2 \\ K\sum X_1^2 & K^2\sum X_1^2 \end{vmatrix} = K^2(\sum X_1^2)^2 - K^2(\sum X_1^2)^2 = 0$$

وبالتالي لا يمكن حساب معكوس المصفوفة أي $(X'X)^{-1}$ وتسمى هذه المصفوفة بالمصفوفة الشاذة وعليه لا يمكن الحصول على المقدرات (OLS) وفقاً للصيغة $\beta = (X'X)^{-1}(X'Y)$ لذلك تتعطل تقديرات (OLS) تحت وطأة الارتباط الخطي المتعدد التام.

② - التعدد الخطي غير التام

No Exact Multicollinearity

مما تقدم يتضح أن حالة التعدد الخطي التام وغياب التعدد الخطي هما حالتان لا تحدثان في الحياة العملية، وما يحدث هنالك دائماً ارتباط بدرجات متفاوتة بين المتغيرات الاقتصادية المستقلة

فعندما تكون معامل الارتباط ($r_{x_ix_j}$) قريباً من الواحد فإن معنوية التباين والتباين المشترك لمعلمات β ستزداد بالإضافة وأهم ما في هذا الموضوع هو اختبار فيما إذا كانت ($\hat{\beta}$) معنوية أم لا معنوية، ومن مقارنة (t) المحسوبة مع قيمة t الجدولية تتحدد المعنوية حيث

$$t = \frac{\hat{\beta}}{S(\hat{\beta})}$$

(78)

فان مصفوفة التباين والتباين المشترك

$$var-cov(\hat{\beta}) = \hat{\sigma}^2 (X'X)^{-1}$$

للمعالم وعليه

ان ارتفاع الارتباط الخطي المتعدد يجعل محدد مصفوفة $(X'X)$ صغيرة قريبة من الصفر مما يجعل $(X'X)^{-1}$ كبيرة وبالتالي كبر قيم الانحرافات المعيارية لقيم المعلمات (βs) لا يجعلها ذات معنوية.

اذا كان استخدام النموذج لاغراض التنبؤ فلا خوف من وجود التعدد الخطي في النموذج

اختبار وجود مشكلة التعدد الخطي

هناك كثير من الاختبارات للكشف عن وجود مشكلة التعدد الخطي ومن اهمها:

1- اختبار فارار وكلوبر The Farrar-Glauber test

2- اختبار فريش frisch test

3- اختبار كلاين Klien test

1- اختبار فارار كلوبر The farrar-Glauber test

ترجع اهمية هذا الاختبار الى انه يساعد في الاختبار لوجود الارتباط الخطي وتحديد درجة حدته ثم تحديد المتغيرات التي يوجد بينها الارتباط واخيرا معرفة من المتغيرات المتسببة في ذلك الارتباط وذلك باستخدام اختبارات X^2 و F و t على الترتيب

1- اختبار مربع كاي X^2 ويحسب وفقا للصيغة ادناه

$$X^2 = -\left[(n-1) - \frac{1}{6}(2K+5)\right] \ln|R|$$

H0: (X_j) is orthogonal — حيث تمثل فرضية العدم X_j متعامدة (المتغيرات المستقلة غير مرتبطة)

H1: (X_j) is not orthogonal — الفرض البديل X_j غير متعامدة، المتغيرات غير مستقلة (يوجد ارتباط خطي)

حيث أن (n) تمثل حجم العينة أو عدد المشاهدات

(k) تمثل عدد المتغيرات المستقلة

$Ln|R|$: تمثل اللوغاريتم الطبيعي لمحدد مصفوفة معاملات الارتباط الجزئية

$$R = \begin{vmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1k} \\ r_{21} & 1 & r_{23} & \cdots & r_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ r_{k1} & r_{k2} & r_{k3} & \cdots & 1 \end{vmatrix}$$

وتحسب معاملات الارتباط الجزئية $r_{x_i x_j}$ حسب الصيغة

$$r_{x_i x_j} = r_{ij} = \frac{\sum x_i x_j}{\sqrt{\sum x_i^2 \cdot \sum x_j^2}}$$

وبعد احتساب قيمة χ^{*2} نجري عملية مقارنتها مع قيمة χ^2 الجدولية بدرجة حرية مساوية الى $\left(k(k-1)/2\right)$ عند مستوى معنوية معين فاذا كانت

$$\chi^{*2} < \chi^2$$

فلا توجد مشكلة التعدد الخطي العلاقة

واذا كانت

$$\chi^{*2} > \chi^2$$

فذلك يعني وجود مشكلة التعدد الخطي

ملاحظة: ان قيمة محدد $(X'X)$ وهو $|R|$ تقع بين الصفر والواحد لذا فان $Ln|R|$ دائماً ستكون سالبة.

المتغير المستقل هو الاول

(ب) يتم تحديد المتغيرات المستقلة المرتبطة خطياً في النموذج بواسطة اختبار F حيث يستخرج قيم F^* المحسوبة وفقاً للصيغة التالية

$$F_j = \frac{R^2_{j \cdot j=1,2,3} \quad k/k-1}{(1 - R^2_{j \cdot j=1,2,\ldots n)\, k)\, n-k}}$$

حيث أن $R^2_{j \cdot 23 \ldots k}$ تمثل معامل الارتباط المتعدد بين كل متغير مستقل X_j في النموذج وبين المتغيرات المستقلة الأخرى

$j = 1, 2, 3 \ldots k$

فعلى سبيل المثال فإن معامل التحديد بين X_1 مع معاملات مستقلة يحسب وفقاً للصيغة الآتية

$$R^2_{1 \cdot 234} = 1 - (1 - r^2_{12 \cdot 34}) \cdot (1 - r^2_{13 \cdot 24})(1 - r^2_{14 \cdot 23})$$

حيث أن $r_{13 \cdot 24}$ و $r_{14 \cdot 23}$ تمثل معاملات الارتباط الجزئية

لاختبار فرضية العدم التالية:

$H_0: R^2_{j \cdot 23 \ldots k} = 0$

$H_1: R^2_{j \cdot 23 \ldots k} \neq 0$ مقابل الفرضية البديلة

فإذا كانت

$$F_j^{\text{المحسوبة}} > F$$

F الجدولية عند مستوى معنوية معين ودرجات حرية مساوية الى $[(k-1), (n-k)]$ للبسط والمقام فإن ذلك يعني وجود ارتباط خطي متعدد بين المتغير المستقل X_j وبين المتغيرات المستقلة الأخرى k

وإذا كانت

$$F_j < F$$

فهذا يعني عدم وجود ارتباط خطي بين المتغير المستقل وبين المتغيرات المستقلة الأخرى، وبالتالي فإن هذا المتغير ليس له علاقة بوجود مشكلة التعدد الخطي ~~التام~~ الذي يعاني منه النموذج

٢- لغرض تحديد المتغيرات المستقلة المسؤولة عن حصول مشكلة التعدد الخطي، فإن ذلك يستوجب إجراء اختبار (t) والذي يحتاج بدوره الى حساب معاملات الارتباط الجزئية بين كل اثنين من المتغيرات المستقلة ويتم احتسابه وفقاً للصيغة الآتية:

$$t^*_{ij} = \frac{r_{ij\cdot 12\ldots k}\sqrt{n-k}}{\sqrt{1-r^2_{ij\cdot 12\ldots k}}}$$

حيث أن

$r_{ij\cdot 12\ldots k}$ تمثل معامل الارتباط الجزئي بين X_i و X_j بافتراض بقية المتغيرات المستقلة ثابتة

$H_0: r_{ij\cdot 12\ldots k} = 0$ فرض العدم

$H_1: r_{ij\cdot 12\ldots k} \neq 0$ فرض البديل

(n) حجم العينة

(k) عدد المتغيرات المستقلة

ثم يتم مقارنة قيمة (t^*_{ij}) مع نظيرتها الجدولية بدرجة حرية ($n-k$) ومستوى معنوية معين فإذا كانت

$$t^*_{ij} > t$$

فذلك يعني أن الارتباط الجزئي بين X_i و X_j معنوي

وبإجراء الاختبار لكافة المتغيرات المستقلة والتي تبين من الاختبارات السابقة (اختبار-F) أنها تشكل مصدر لمشكلة التعدد الخطي. وبذلك تشخص بشكل نهائي المتغيرات المستقلة المسببة لحصول مشكلة التعدد الخطي.

ملاحظة: لو سأل السؤال معامل الارتباط الجزئي بين متغيرين مستقلين بثبوت الثالث نحسبه وفقاً للصيغة الآتية

$$r_{12\cdot 3} = \frac{r_{12} - (r_{13})(r_{23})}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}}$$

(77) Klien test ⑧ اختبار كلاين

يستخدم هذا الاختبار للكشف عن وجود التعدد الخطي حيث يتم مقارنة معامل التحديد R^2 مع مربع معامل الارتباط بين المتغيرات المستقلة، فاذا كان معامل التحديد R^2 اكبر من مربع معامل الارتباط بين المتغيرات المستقلة فهذا يعني عدم وجود مشكلة التعدد الخطي وان كانت موجودة فهذا لا يؤثر او يكون غير مؤثر اي ان

$$R^2 > r^2_{x_i x_j}$$

طرق معالجة مشكلة التعدد الخطي

ونشرحها بالتالي

① - محاولة توسيع حجم العينة وذلك باضافة بيانات كافية عن المتغيرات الظاهرة المدروسة، اذ تزداد التقديرات دقة بزيادة عدد البيانات التي تعتمد عليها عملية التقدير نظراً لوجود علاقة عكسية بين حجم العينة وحجم التباين، فكلما كبر حجم العينة كما تم الحصول على معلومات اضافية تساعد على تخفيض حجم التباين.

② - حذف المتغير المستقل او المتغيرات المستقلة التي تسببت في ظهور المشكلة في النموذج، وللسهولة وان سبيل هذا الحل للمشكلة مشكلة اخرى، اذا ان حذف متغير مستقل يعمل له اهمية في النموذج يؤدي الى مشكلة التوصيف وعدم ادخال المتغيرات المهمة في النموذج مما يرفع من احتمال تحيز المقدرات في تلك الحالة.

③ - تحويل شكل الدالة باستعمال النسب او الفروقات عوضاً عن المتغيرات الاصلية، فعلى سبيل المثال النموذج التالي

$$Y_i = a + b_1X_1 + b_2X_2 + u_i$$

حيث ان احد المتغيرات المستقلة وليكن X_2 كعامل (نقسم طرفي النموذج عليه) فنحصل على متغيرات جديدة كما يلي

$$\frac{Y_i}{X_2} = a\frac{1}{X_2} + b_1\frac{X_1}{X_2} + b_2 + \frac{u_i}{X_2}$$

(24)

غير انه يلاحظ ان النموذج الجديد قد استوفى افتراض (كان الاعتيادية الا هو ثبات التباين اي انه لا يعاني من عدم ثبات تباين ثابت كما لاحظنا في $\frac{u_i}{X_2}$ حيث

$$E\left(\frac{u_i}{X_2}\right)^2 = \frac{E(u_i^2)}{X_2^2} = \frac{\sigma^2}{X_2^2} \neq \sigma^2$$

مما يعني استبدال مشكلة بمشكلة اخرى.

4) - استخدام اسلوب الدمج بين بيانات السلاسل الزمنية والبيانات المقطعية مثلاً يمكن استخدام المعلمة المستخرجة بواسطة البيانات المقطعية مع العلاقة المقدرة بواسطة السلاسل الزمنية، فاذا استخدم مقدر معلمة الميل الحدي للاستهلاك لفترة معينة ولقطاع معين من دراسة المقاطع العرضية مع العلاقة بين الدخل والاستهلاك لنفس الفترة والقطاع من دراسات السلاسل الزمنية وفضلاً عن ذلك، توجد طرق اخرى يمكن استخدامها لمعالجة مشكلة التعدد الخطي نذكر منها اسلوب او طريقة الـ (Ridge Regression) (طريقة او اسلوب انحدار الحرف) وطريقة المركبات الرئيسية incripe components