حسن تخصيص النماذج/ مشاكل القياس

مشكلة الارتباط المتعدد -الازدواج الخطي-

(Multicollinearity)

تعريف: تظهر مشكلة الارتباط المتعدد أو الازدواج الخطي عندما يكون متغيران (أو أكثر) من المتغيرات المستقلة لهما ارتباط خطي عالي فيما بينها، للارتباط المتعدد نوعان:

أ/ ارتباط متعدد (ازدواج خطي) تام: ويحدث حين يظهر ارتباط خطي تام بين متغيرين مستقلين أو أكثر، ومعامل الارتباط بينهما يساوي ± 1، ومنه المصفوفة $(X'X)$ هي مصفوفة فردية ($|X'X| = 0$) وبالتالي لا تقبل مقلوب (أي لا يمكننا تقدير المعلمات β). حيث:

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

مثال

$$R = f(X_1, X_2, X_3)$$

R: عائد السهم.

X_1: عدد عمليات البيع والشراء أسبوعيا.

X_2: عدد عمليات البيع والشراء يوميا.

X_3: توزيعات الأرباح.

في هذه الحالة يمكن أن نلاحظ العلاقة الخطية الموجودة بين عدد عمليات البيع والشراء أسبوعيا (X_1) وعدد عمليات البيع والشراء يوميا (X_2)، وهنا المصفوفة $(X'X)$ لا تقبل مقلوب (أي لا يمكننا تقدير معلمات النموذج).

ب/ ارتباط متعدد (ازدواج خطي) غير تام: ويحدث حين يظهر ارتباط خطي قوي غير تام بين متغيرين مستقلين أو أكثر، ومعامل الارتباط بينهما يقترب من ± 1، إذن المصفوفة $(X'X)$ هي مصفوفة غير فردية (محددها يختلف عن 0) وبالتالي يمكن تقدير المعلمات B، بالرغم من عدم مصداقية نتائج التقدير.

1 – أعراض الارتباط المتعدد غير التام

- ✓ انحراف معياري عالي (قيم t منخفضة) لمقدرات المربعات الصغرى رغم ارتفاع قيمة معامل التحديد R^2.
- ✓ الانحدار حساس جدا تجاه زيادة أو حذف عدد قليل من المشاهدات.
- ✓ عوامل القطر الرئيسي للمصفوفة (X'X) أقل نسبيا من العوامل خارج القطر الرئيسي (مما يعني أن الجذور المميزة (eigenvalues) للمصفوفة (X'X) ذات قيم صغيرة).
- ✓ ارتباط خطي ثنائي كبير بين المتغيرات المستقلة.

2- تشخيص مشكل الارتباط المتعدد

يعطى الانحدار التالي $Y = f(X_1, X_2, X_3, \ldots, X_j) \ldots (*)$

أ- معامل تضخم التباين (VIF: Variance Inflation Factor)

$$VIF_j = \frac{1}{1 - r_j^2}$$

r_j^2: معامل التحديد لانحدار المتغير المستقل X_j على بقية المتغيرات المستقلة في الانحدار الأصلي $X_j = f(X_1, X_2, X_3, \ldots) (**)(*)$

معامل تضخم التباين ينقسم إلى نوعين الممركز وغير الممركز:

- **VIF** الممركز: وهو نسبة تباين المعلمة المقدرة من النموذج الأصلي (*) إلى تباين المعلمة المقدرة من الانحدار الأصلي بهذا المتغير المستقل X_j والثابت فقط.
- **VIF** غير الممركز: وهو نسبة تباين المعلمة المقدرة من النموذج الأصلي (*) إلى تباين المعلمة المقدرة له من الانحدار الأصلي بهذا المتغير المستقل X_j بدون ثابت.

مسلمة: إذا كان معامل تضخم التباين لمتغير **10** $<$، إذن هذا المتغير (X_j) لديه ارتباط خطي عالي مع المتغير المستقل الآخر (أو المتغيرات المستقلة الأخرى).

إذا كان معامل تضخم التباين لمتغيرين يساوي **1** ، فإنه لا وجود لارتباط خطي بين المتغير (X_j) والمتغير المستقل الآخر (أو المتغيرات المستقلة الأخرى).

ب- معامل التحمل (Tolerance)

$$Tol_j = \frac{1}{VIF_j}$$

إذا كان معامل التحمل للمتغير Xj $Tol_j = 0$، فإن المتغير (X_j) مرتبط ارتباطا خطيا تاما **perfect collinearity** مع المتغير المستقل الآخر (أو المتغيرات المستقلة الأخرى).

إذا كان معامل التحمل للمتغير Xj $Tol_j = 1$، فإنه لا وجود لارتباط خطي بين للمتغير (X_j) المتغير المستقل الآخر (أو المتغيرات المستقلة الأخرى).

ج- الرقم الشرطي أو المؤشر الشرطي (Condition Index)

$$CI = \frac{\sqrt{\lambda\ max}}{\sqrt{\lambda\ min}}$$

λ_i:الجذور المميزة (eigenvalues) للمصفوفة (X'X).

قاعدة:

1- عندما تأخذ قيم المؤشر الشرطي بين 30 و 50، فإن هناك احتمالية عالية لوجود مشكلة الارتباط المتعدد.

2- إذا تجاوزت قيم المؤشر الشرطي 50، فإن المشكلة تكون أعمق.

3- إذا قلّت قيم المؤشر الشرطي عن 30، فإنه لا وجود لمشكلة الارتباط المتعدد.