المحور05: التعدد الخطي (أو التداخل الخطي أو الاشتراك الخطي) Multicollinearity

المحاضرة 06:

- يحدث التعدد الخطي في نماذج الانحدار المتعدد.
- التعدد الخطي يحدث بين المتغيرات المستقلة و يعني وجود إرتباط قوي بين المتغيرات المستقلة.
- هناك نوعين من الارتباط الخطي بين المتغيرات المستقلة الذي ينتج لنا التعدد الخطي:
  - **الإرتباط الخطي التام PREFECT MULTICOLLINEARITY**: مثلا $[X_{1t} = 2X_{2t}]$ أو $[X_{1t} = 3X_{3t}]$ أو $[X_{2t} = X_{3t}]$ ..... وفي حال حدوث ارتباط خطي تام بين المتغيرات المستقلة في النموذج، فإنه يستحيل أصلا تقدير النموذج، لأن محدد المصفوفة $(X'X)$ يكون مساويا للصفر $[\det(X'X) = 0]$ وبالتالي لا يمكن تقدير النموذج لأن معكوس المصفوفة $(X'X)$ لا يمكن حسابه.
  - **الارتباط الخطي غير التام IMPERFECT MULTICOLLINEARITY** : مثلا $[X_{1t} = X_{2t} + 1]$ في حالة الارتباط الخطي غير التام يمكن تقدير معالم النموذج، لأنه يمكننا ايجاد معكوس المصفوفة ، طالما أن محددها غير معدوم. ولكن لا يمكننا التأكد من وجود مشكلة التعدد الخطي إلا باختبارات الكشف عن هذه المشكلة.
- يحدث التعدد الخطي بسبب البيانات المأخوذة ونوعيتها، فاستخدام البيانات الخاطئة والمرتبة وفقا لترتيب معين أو سلوك بعض من المتغيرات سلوكا باتجاه واحد، يجعل هذه المتغيرات تترابط فيما بينها.
- التعدد الخطي يظهر بشكل كبير في بيانات السلاسل الزمنية أكثر مما يظهر في الأنواع الأخرى من البيانات.
- في وجود التعدد الخطي:
  - قيمة التباين للمعلمات المقدرة تكون كبيرة جدا.
  - قد تكون قيمة معامل التحديد عالية جدا، وقد تصل إلى 100%.
  - إحصائية FISHER المحسوبة تكون عالية جدا.
  - عدم معنوية معلمات النموذج نظرا لصغر قيمة احصائية STUDENT المحسوبة وهو ما ينتج عنه انحدارا مزيفا لا يعكس حقيقة الظاهرة المدروسة.
  - مجلات الثقة للمعلمات تكون واسعة، مما ينتج عنه قبول الفرضية الصفرية بكل سهولة.
- في وجود التعدد الخطي، مقدرات OLS تعتبر غير دقيقة ومغلوطة. إذ يمكن الحصول على إشارات للمعلمات المقدرات مخالفة للنظرية الاقتصادية والمنطق الاقتصادي.
- التعدد الخطي يعتبر مشكلة.

1- أسباب مشكلة التعدد الخطي:

من أهم الأسباب التي تؤدي إلى ظهور مشكلة التعدد الخطي نذكر مايلي:

- اتجاه المتغيرات الاقتصادية للتغير معا مع مرور الزمن، فإذا أخذنا المتغيرات الاقتصادية التالية: الدخل، الاستهلاك، الاستثمار والعمالة، فنجد أنها بمرور الزمن ستتزايد، وبما أنه هناك ارتباط بين هذه المتغيرات، فالتعدد الخطي سوف يظهر.
- إدراج متغيرات ذات ابطاء زمني كمتغيرات مستقلة في النموذج الخطي المتعدد، فدخل الفترة الحالية مثلا يتحدد جزئيا بدخل الفترة السابقة، وهذا ما يعني أن هناك ارتباط بين القيم المتتالية لمتغير معين، وبالتالي نشوء مشكلة التعدد الخطي.
- طريقة جمع البيانات.

2- الكشف عن مشكلة التعدد الخطي:

توجد عدة اختبارات للكشف عن وجود التعدد الخطي من عدمه نذكر أهمها:

- معامل تضخيم التباين VIF.
- شرط العدد CONDITION NUMBER.
- اختبار كلاين KLEIN TEST.
- اختبار FARRAR-GLAUBER.

ما يهمنا نحن باستخدام البرنامج الاحصائي EVIEWS، هو استخدام معامل تضخيم التباين (VARIANCE INFLATION FACTOR) الذي يحدد درجة التعدد الخطي. حيث يوفر برنامج EVIEWS هذا الاختبار.

يعطى معامل تضخيم التباين بالعلاقة التالية:

$$V.I.F\left(\hat{\beta}_i\right) = \frac{1}{1 - R_i^2}$$

ملاحظات:

- القيمة $[1 - R_i^2]$: تسمى بمقياس التحمل TOLERANCE. يتم حسابه من خلال اجراء معادلة انحدارية لكل متغير مستقل بدلالة المتغيرات المستقلة الأخرى وحساب مربع معامل الارتباط المتعدد.
- إذا كان معامل تضخيم التباين يساوي 10 أو أكثر، فذلك يدل على وجود مشكلة التعدد الخطي MULTICOLLINEARITY.
- معامل تضخيم التباين يساوي 10 يعني أن التباين تضاعف 10 مرات عن القيمة التي يمكن أن يحصل عليها إذا كانت المتغيرات المستقلة غير مرتبطة أي مستقلة عن بعضها البعض.

مثال تطبيقي:

باستخدام البيانات التالية على البرنامج الاحصائي EVIEWS، اختبر وجود مشكلة التعدد الخطي.

Y	X_1	X_2	X_3
49	1300	7.5	0.012
50.2	1300	9	0.012
50.5	1300	11	0.0115
48.5	1300	13.5	0.013
47.5	1300	17	0.0135
44.5	1300	23	0.012
28	1200	5.3	0.04
31.5	1200	7.5	0.038
34.5	1200	11	0.032
35	1200	13.5	0.026
38	1200	17	0.034
38.5	1200	23	0.041
15	1100	5.3	0.084
17	1100	7.5	0.098
20.5	1100	11	0.092
29.5	1100	17	0.086

الحل:

Variance Inflation Factors
Date: 03/18/24 Time: 06:26
Sample: 2007 2022
Included observations: 16

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	3073.118	3464.903	NA
X1	0.001779	2961.596	12.22505
X2	0.031335	6.532563	1.061838
X3	11648.50	33.66826	12.32496

3- معالجة مشكلة التعدد الخطي: توجد عدة إقتراحات لمعالجة مشكلة التعدد الخطي، أهمها:

- زيادة حجم العينة.
- حذف متغيرات من النموذج، عندما يكون هناك تداخل خطي بين متغيرين مستقلين، يلجأ الباحث أحيانا إلى حذف أحد المتغيرين للتخلص من هذا التداخل. ويجب أن يكون حذف المتغيرات عملية تكرارية تبدأ بالمتغير الذي له أكبر قيمة VIF.
- إجراء بعض التحويلات على المتغيرات، مثل مزج المتغيرات في متغير واحد، أو تحويل متغيرات النموذج إلى الفروق الأولى: $\Delta X_t = X_t - X_{t-1}$

مثال: نأخذ المثال السابق.

نتائج التقدير:

Dependent Variable: Y
Method: Least Squares
Date: 04/15/24 Time: 05:17
Sample: 2007 2022
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-121.2696	55.43571	-2.187572	0.0492
X1	0.126854	0.042182	3.007300	0.0109
X2	0.348158	0.177017	1.966802	0.0728
X3	-19.02170	107.9282	-0.176244	0.8630

R-squared	0.919815	Mean dependent var	36.10625
Adjusted R-squared	0.899769	S.D. dependent var	11.89877
S.E. of regression	3.767073	Akaike info criterion	5.702792
Sum squared resid	170.2901	Schwarz criterion	5.895939
Log likelihood	-41.62233	Hannan-Quinn criter.	5.712682
F-statistic	45.88452	Durbin-Watson stat	0.775471
Prob(F-statistic)	0.000001		

المتغيرين X_2 و X_3 غير معنويين عند 5% ومعامل التحديد قيمته عالية وتقترب من 1 ، وهذا مؤشر مبدئي عن وجود مشكلة التعدد الخطي. وكما رأينا فإن نتائج معامل تضخيم التباين دلت على وجود تعدد خطي:

Variance Inflation Factors
Date: 04/15/24 Time: 05:22
Sample: 2007 2022
Included observations: 16

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	3073.118	3464.903	NA
X1	0.001779	2961.596	12.22505
X2	0.031335	6.532563	1.061838
X3	11648.50	33.66826	12.32496

كما قلنا، من بين الطرق لمعالجة التعدد الخطي نقوم بحذف المتغير الذي له أكبر قيمة في VIF. في حالتنا هذه نقوم بحذف المتغير X_3:

تصبح نتائج التقدير هي:

Dependent Variable: Y
Method: Least Squares
Date: 04/15/24 Time: 05:26
Sample: 2007 2022
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-130.6899	14.14571	-9.238835	0.0000
X1	0.133961	0.011908	11.25008	0.0000
X2	0.351061	0.169553	2.070504	0.0487

R-squared	0.919607	Mean dependent var	36.10625
Adjusted R-squared	0.907239	S.D. dependent var	11.89877
S.E. of regression	3.623968	Akaike info criterion	5.580377
Sum squared resid	170.7308	Schwarz criterion	5.725237
Log likelihood	-41.64301	Hannan-Quinn criter.	5.587795
F-statistic	74.35306	Durbin-Watson stat	0.792878
Prob(F-statistic)	0.000000		

أما نتائج اختبار الكشف عن التعدد الخطي باستخدام VIF فهي:

Variance Inflation Factors
Date: 04/15/24 Time: 05:29
Sample: 2007 2022
Included observations: 16

Variable	Coefficient Variance	Uncentered VIF	Centered VIF
C	200.1010	243.7815	NA
X1	0.000142	255.0093	1.052642
X2	0.028748	6.475985	1.052642