المحاضرة 8

الارتباط الخطي المتعدد

1.8 مقدمه

2.8 مشكلة الارتباط الخطي

3.8 آثار وجود الارتباط الخطي المتعدد المرتفع

4.8 النماذج التي تحتوي اكثر من متغيرين مستقلين

5.8 طريقة كشف الارتباط الخطي المتعدد

6.8 طريقة معالجة الارتباط الخطي المتعدد

14 مقدمه

مشـــــــــكلة الارتباط المتعدد Multicollinearity خاصة بنماذج الانحدار المتعدد لأتحدث في نموذج الانحدار البسـيط. وهي مشكلة خاصة بالمتغيرات المستقلة. على سبيل المثال إذا كـان هناك نموذج ارتباط خطي متعدد يحتوي على متغيرين مستقلين X1, X2

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_t$$

إذا حدث إن هناك ارتباط X1,X2 بين كما يلي :

$$X_2 = AX_3$$

نقول إن فيه مشكلة ارتباط خطي متعدد. و معامل الارتباط البسيط بين المتغيرين المستقلين يساوي الواحد الصحيح:

$$r_{X2X3} = \frac{\sum x_2 x_3}{\sqrt{\sum x_2^2}\sqrt{\sum x_3^2}..}$$

باستخدام معامل الارتباط r_{X2X3} أو مر بعها معامل التحديد r^2_{X2X3} نحدد إذا كـان هناك ارتباط خطي متعدد. إذا كان الارتباط $r_{X2X3} = 0$ يساوي الصفر فانه ليس هناك مشكلة ارتباط خطي متعدد آما إذا كانت هناك ارتباط كامل آي أن $r_{X2X3} = 1$ فنول أن هناك ارتباط خطي متعدد تام. وعند حدوث الارتباط الخطي التام لا نستطيع أجراء المقدرات باستخدام م ص ع . ولكن هذه المشكلة لاتحدث كثيرا في الدراسات العملية إلا في ظروف استثنائية ويمكن معالجتها بحذف أحد المتغيرات لان الآخر يقوم مقامه.

<u>2.4 مشكلة الارتباط الخطي:</u> يمكن ان تقسم مشكلة الارتباط الخطي الى نوعين:

<u>1.2.4 مشكلة درجة الارتباط:</u>

عندما تكون $0 \leq r_{X1X2} < 1$ إذا كانت r بسيطة لا يمثل هذا مشكله معقده ولكن إذا كانت r مرتفعه تكون هناك مشكلة ارتباط خطي متعدد يجب معالجتها. أي أن المشكلة تتزايد حدتها بزيادة قيمة r أي بتزايد قوة الارتباط بين المتغيرين . أي إنها مشكلة درجة الارتباط.

<u>2.2.4 مشكلة العينة:</u>

في العينة التي استخدمناها لقياس الارتباط بين X_1, X_2 اذا وجد ارتباط بين هذين المتغيرين أن هذا لا يعني أن هناك ارتباط بينهم <u>في كل الأجزاء</u> ولكن أي انه يمكن استخدام عينه مختلفة في

سنوات قد يكون فيها اقل ارتباط أو جزء من عينه أخري يقل فيها الارتباط بين المتغيرين المستقلين.

3.4 آثار وجود الارتباط الخطي المتعدد المرتفع: (انظري ص 134-148)

تحت وجود الارتباط الخطي المرتفع يمكن تطبيق م ص ع بالطريقة المعتادة وتحتفظ بخواصها الكفاءة وغير متحيزة ولكن تباينها يكون مرتفع على الرغم إنها مازالت تحتفظ بخاصية آدني تباين. مما ينعكس على دقة القياسات.

1.3.4 التأثير على الانحراف المعياري:

يوجد علاقة بين تباين والارتباط بين وهذه العلاقة هي التي يعتمد عليها لتوضيح مشكلة الارتباط المتعدد على تباين المقدرات وبالتالي دقة القياسات. يختلف هذا حسب حالة الارتباط المتعدد:

في حالة النموذج الذي يوجد به متغيرين

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_t$$

يمكن تقسيم التأثير على حسب درجة الارتباط

أ -لا يوجد أي ارتباط بين X_1, X_2 $r^2_{X2X3} = 0 \Leftarrow$

$$V(\beta_2) = \frac{\sigma^2}{\sum x_2^2 (1 - r^2)}$$

و

$$V(\beta_3) = \frac{\sigma^2}{\sum x_3^2 (1 - r^2)}$$

وهذا يكون الوضع المثالي في نموذج الانحدار المتعدد والتباين يكون منخفض والدقة عالية. الوضع المثالي يسمى الانحدار المتعامد ويعني أن تأثير X_2على المتغير التابع منفصل تماما عن تأثير X_3 على المتغير التابع. والتباين منخفض والدقة عالية.

ب- الارتباط التام: $r^2_{X2X3} = 1$

$$\Leftarrow \quad v(\beta_2) = \frac{\sigma^2}{\sum x^2} = \frac{\sigma^2}{0} \quad = \quad \infty$$

حيث تتضح العلاقة الخطية بين أعمدة وصفوف هذه المصفوفة وينعكس ذلك على المحدد اذ يتخذ قيمة الصفروعلية لايمكن ايجاد قيمة β حسب القانون ينفجر التباين وينهار الانحدار لا يوجد تباين ولا يوجد معادلة للقياس.

ج- r^2_{X2X3}- مرتفعه قريبه من الواحد

مثلا 0.90 معيار المقام منخفض، التباين مرتفع الدقة قليلة.

4.4 النماذج التي تحتوى اكثر من متغيرين مستقلين:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u_t$$

لا نستطيع الاعتماد على الارتباط البسيط لأن معامل الارتباط البسيط يختص بمتغيرين فقط، يمكن تعميم قانون التباين

$$V(\beta_k) = \frac{\sigma^2}{\sum x_k^2 (1 - R_k^2)} \qquad k=2,3,4$$

معامل التحديد = R_k^2 الناتج من انحدار م ص ع بمتغير تابع X_k والبقية متغيرات مستقلة فمثلا لا استخراج R_3^2 يجري انحدار للمتغير X_3 على المتغيرات الأخرى X_4, X_2 ويعمل انحدار للمتغيرات المستقلة فقط.

$$X_3 = \gamma_0 + \gamma_1 X_2 + \gamma_2 X_4 + \eta_t \qquad R_3^2$$

ويقيس R_3^2 لهذا الانحدار مدى الارتباط بين X_3 و X_2, X_4 . و تباين $V(\beta_3)$ يساوي

$$V(\beta_3) = \frac{\sigma^2}{\sum x_3^2 (1 - R_3^2)}$$

ويستخدم $\frac{1}{1 - R_k^2}$ لقياس درجة التباين ، حيث يسمى بمعامل تضخم التباين و يستخدم كمؤشر لزيادة التباين آو لانخفاض الدقة وذلك إذا زاد الارتباط بين مجموعة المتغيرات المستقلة فان R_k^2 يرتفع وبالتالي ينخفض المقام ويرتفع معامل تضخم التباين ويرتفع التباين وبالتالي تقل دقة القياس.

الدرجة العالية من الارتباط الخطي المتعدد تسبب مشكلة إن مقدرات معالم الانحدار تصبح غير دقيقه وينشأ عدم الدقة من التباين المتزايد للمقدرات.

5.4 طريقة كشف الارتباط الخطي المتعدد:

1- معامل الارتباط:

إذا كان نموذج لمتغيرين مستقلين يمكن الاكتفاء بقياس r^2_{X2X3} و r_{X2X3} إذا كانت كبيره نقول إن هناك مشكلة الارتباط الخطي وإذا كانت منخفضة نقول انه لا يوجد مشكله. إما في حالة النماذج

التي تحتوي عدد كبير من المتغيرات المستقلة حيث يكون الارتباط بين عدد اكبر من المتغيرات المستقلة لا يستطيع معامل الارتباط قياس الارتباط لأنه يقيس لمتغيرين فقط. ولكن الارتباط قد يحدث بين المجموعة ككل وليس بين عنصرين آو متغيرين فقط. الارتباط قد يمتد ليشمل مجموعه المتغيرات المستقلة وليس بالضرورة إن يقتصر على اثنين فقط قد يكون بين أربعه آو خمسه .... آو بين المجموعة ككل. إذا كان الآمر كذلك علينا إن نعتمد على طرق قياس أخرى.

2-استخدام محدد مصفوفة المتغيرات المستقلة ومصفوفة الارتباط.

يتم فحص المصفوفة التي تضم معاملات الارتباط البسيط بين كل المتغيرات المستقلة كعناصر. ففي النموذج العام يتم بناء المصفوفة المتماثلة بحذف الصف الاول والعمود الاول الخاصين بالمتغير التابع Y من المصفوفة:

$$R = \begin{vmatrix} 1 & r_{23} & r_{24}\ldots & r_{2k} \\ & 1 & r_{34}\ldots & r_{2k} \\ & & 1 \quad \ldots & r_{2k} \end{vmatrix}$$

اذا جرى حساب محددة المصفوفة فاذا كان المحدد يساوي للصفر كان ذلك دلالة على وجود ارتباط تام يجمع بين المتغيرات المستقله.

3-استخدام اختبارF وt كمؤشر لوجود الارتباط:

إذا كان R_k^2 مرتفع بينما F آو t منخفضة جدا فانه يكون هناك ارتباط خطي متعدد. على سبيل المثال إذا كانت لديك الانحدار التالي:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_t$$
$$Y_i = 10 + 2X_1 + 4X_2 + u_t$$
$$(t=0.5) \quad (t=1) \qquad\qquad R^2 = 0.95$$

هذا الوضع يشير إلى وجود مشكله

$$SST = SSR + SSE$$
$$\frac{SST}{SST} = \frac{SSR}{SST} + \frac{SSE}{SST}$$
$$1 = R^2 + \frac{SSE}{SST}$$

$$F = \frac{R^2(SST)/(k-1)}{(1-R^2)/SST/n-k)}$$

$$F = \frac{R^2/(k-1)}{(1-R^2)/n-k)}$$

إذا كانت R_3^2 عالية و t الخاصة بالمقدرات هي المهمة فإذا كانت منخفضة بينما R_3^2 مرتفعه فان هذا مؤشر لوجود الارتباط الخطي المتعدد. حيث إن R_3^2 في نفس الوقت المرتفعة سوف تؤدي إلى رفض فرضية العدم وقبول الفرض البديل بان النموذج صالح. لكن من الناحية الأخرى نقول إن X_1 و X_2غير معنوية بسبب انخفاض قيمة t آي انه إذا اختبرنا المتغيرات على انفراد نصل إلى نتيجه مناقضة إذا آخذنا النموذج ككل.

<u>التباين مرتفع عند وجود الارتباط</u>: معناه أن الأخطاء المعيارية ستكون مرتفعه و tستكون منخفضة . التضارب الناتج من وجود الارتباط الخطي المتعدد.

لدراسة الدالة التالية Y والعوامل المؤثرةX_1 و X_2 تحصلنا على النتائج التاليه:

$$\boldsymbol{Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u_t}$$

$$\boldsymbol{Y = 5.47 + 0.086 X_1 + 0.414 X_2 + u_t}$$

$$\boldsymbol{SE \qquad (0.211) \qquad (0.231)}$$

$$\boldsymbol{R^2 = 0.90 \qquad n = 20\backslash}$$

$$\boldsymbol{SSE = 129.025 \qquad \sum y^2 = 18102}$$

$$\boldsymbol{\sum x_1^2 = 5362 \quad \sum x_2^2 = 4482 \quad \sum x_1^2 x_2^2 = 4824}$$

$$\boldsymbol{\sigma^2{}_{xy} = 2.754}$$

من المعلومات أعلاة حددي :

1) جدول تحليل التباين للنموذج.
2) معامل الارتباط بين المتغيرات المستقلة.
3) حددي معنوية اختبار العلاقة بين المتغيرات Y والعوامل المؤثرةX_1 و X_2 .
4) اذا كانت قيمة معامل التحديد مرتفعه واختبارF منخفضة و يتم رفض انه يوجد علاقه بين المتغيرات مالمعني التحليل لهذا.

تم حذف المتغير X_1 وتحصلنا على النتائج التالية.هل تحسن النموذج وهل استفدنا من حذف المتغير , X2

$$Y = \beta_0 + \beta_1 X_1 + u_t$$
$$Y = 5.31 + 0.458 X_1 + u_t$$
$$SE \qquad (0.039)$$
$$R^2 = 0.88 \qquad n = 20 \backslash$$
$$SSE = 153.39 \qquad F = 132.43[0.000]$$
$$\sigma_{xy} = 2.912$$

6.4 طريقة معالجة الارتباط الخطي المتعدد:

1 -معلومات اضافيه للعينة آو استعمال معلومات من خارج العينة.

2 -حذف متغيرات مستقلة من النموذج.

3 -وضع المتغيرات في شكل نسب يعني قسمة النموذج على أحد المتغيرين المستقلتين.

تمرين

	استهلاك C	الدخل المتاح (Y)	الأصول (A)	c	y	a	cy	ac	a2
1	20	25	250	-14.14	-15	-157.14	212.1	2221.96	24692.98
2	23	30	310	-11.14	-10	-97.14	111.4	1082.14	9436.18
3	28	35	330	-6.14	-5	-77.14	30.7	473.6396	5950.58
4	38	40	390	3.86	0	-17.14	0	-66.1604	293.7796
5	35	45	480	0.86	5	72.86	4.3	62.6596	5308.58
6	50	50	540	15.86	10	132.86	158.6	2107.16	17651.78
7	45	55	550	10.86	15	142.86	162.9	1551.46	20408.98
المجموع	**239**	**280**	**2850**				**680**	**7432.857**	**83742.86**
المتوسط	**34.14**	**40**	**407.14**						

y2	ya	u	u2
225	2357.1	-341.757	116798.1
100	971.4	-333.639	111314.9
25	385.7	-325.228	105773.5
0	0	-310.11	96168.14
25	364.3	-306.71	94071.26
100	1328.6	-286.592	82134.9
225	2142.9	-288.608	83294.8
700	**7550**		**689555.6**

$$\hat{\beta}_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\hat{\beta}_1 = \frac{((680)(83742.86) - (7432.86)(7550)}{(700)(83742.86) - (7550)} = 0.5113$$

$$\hat{\beta}_{21} = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\beta_2 = \frac{(83742.86)(700) - (680)(7550)}{(700)(83742.86) - (7550)} = 0.0427$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$$\beta_0 = 34.14 - 0.5113(40) - 0.0427)(407.14) = -367.83$$

$$V(\hat{\beta}_1) = \frac{\sum u^2}{n-k} \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$
$$V(\hat{\beta}_2) = \frac{\sum u^2}{n-k} \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

$$V(\hat{\beta}_1) = \frac{\sum u^2}{n-k} \quad \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$
$$V(\beta_1) = \frac{689555.6}{7-3} \quad \frac{837428.6}{(700)(7550) - 7432.857} = 1.0623$$
$$SE(\beta_1) = \sqrt{V(\beta 1)} = \sqrt{1.0623} = 1.037$$
$$V(\hat{\beta}_2) = \frac{\sum u^2}{n-k} \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$
$$V(\beta_2) = \frac{689555.6}{7-3} \quad \frac{(700)}{(700)(7550) - 7432.857} = 0.205209$$
$$SE(\beta_2) = \sqrt{V(\beta_1)} = \sqrt{0.05209} = 0.0942$$
..

$$\hat{C} = -367.83 + 0.5113Y + 0.0427a$$
$$= \quad (1.0307) \quad (0.0942)$$
$$t = \quad 0.496 \quad 0.453 \quad \bar{R}^2 = 0.835$$

اذا تم حذف المتغير a الاصول النقدية ستكون المعادلة كما يلي

$$\hat{C} = -471.43 + 0.9714yd$$
$$(0.157)$$
$$t = \quad 6.187 \quad \bar{R}^2 = 0.861$$

ماذا حدث من تغيرات ولماذا؟؟

1 - ارتفاع قيمة t لمتغير الدخل المتاح.

2 - الارتباط بين المتغيرين الدخل المتاح والاصول النقدية مرتفع $r_{ya} = 0.986$ مما يؤدي الى ارتفاع الخطأ المعياري ومن ثم انخفاض قيمة t حيث ارنخفض من قيمة 0.157 الى 1.03

$$r_{ya} = \frac{\sum x_1 x_2}{\sqrt{x_1^2}\sqrt{x_2^2}} = \frac{7550}{\sqrt{700}\sqrt{837428.6}} = 0.986$$

3 - قيمة معامل التحديد متساوي على الرغم من اختلاف معنوية المتغيرين المفسرين في المعادلتين. 4- هذه نتائج المعادلة عندما يكون هناك ارتباط خطي متعدد.

$$\hat{\beta}_1 = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\hat{\beta}_{21} = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \hat{\beta}_2 \overline{X}_2$$