مشاكل تقدير نماذج الإنحدار

الاستاذ المساعد

الدكتور احمد الحسيني

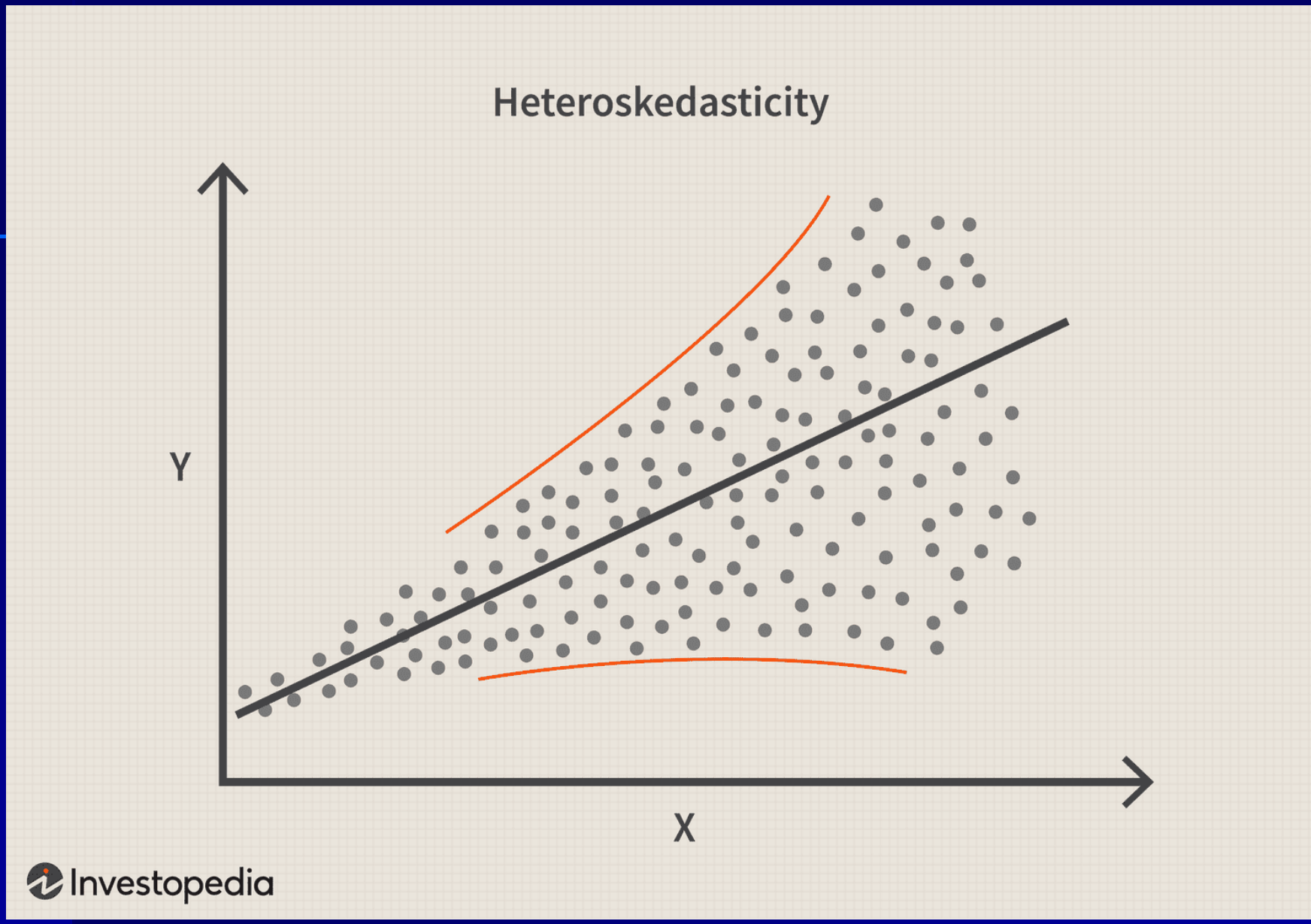
مشاكل تقدير نماذج الإنحدار

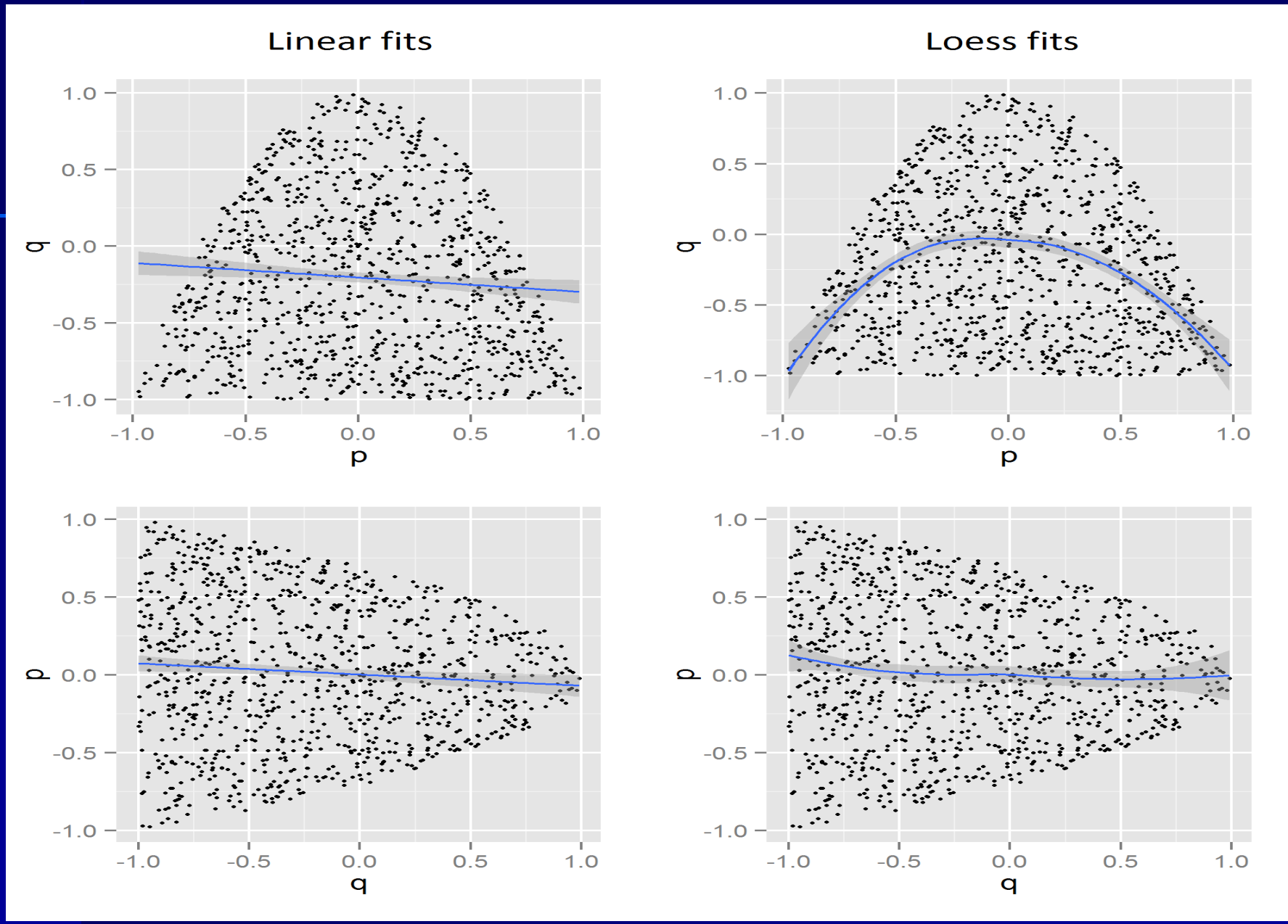
- هناك ثلاثة مشاكل قياسية تواجه الباحث :

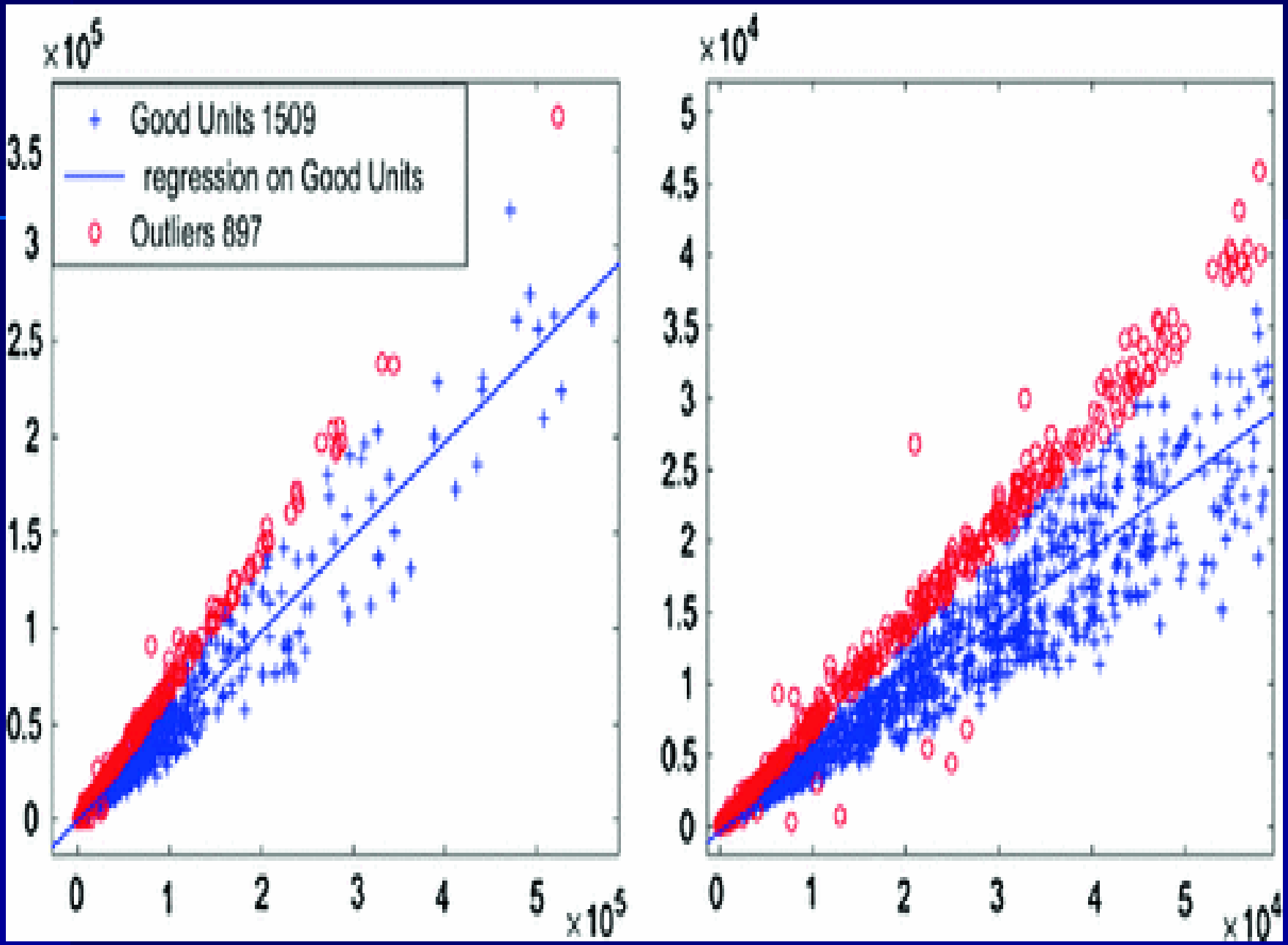
1. الإرتباط الذاتي من الدرجة الأولى Autocorrelation
2. عدم ثبات تباين حد الخطأ Heteroscedasticity
3. الإزدواج الخطيMulticollinearity

عدم ثبات تباين حد الخطأ
Heteroscedasticity

- تعني أن حدود الأخطاء للبيانات المراد إجراء الإنحدار الخطي عليها ليس لها نفس التباين .
- يرتبط عدم ثبات تباين حد الخطأ ببيانات المقطع المستعرض Cross-section Data أكثر من بيانات السلسلة الزمنية Time-series Data
- بيانات المقطع المستعرض عبارة عن بيانات يتم تجميعها عن متغير ما في لحظة زمنية معينة .
- هناك فرض من فروض حد الخطأ العشوائي ينص على ((توزيع حد الخطأ العشوائي يبقى نفسه لكل مشاهدات (Xi) اي بمعنى ان تباين حد الخطأ يبقى نفسه لكل قيم المتغيرات المستقلة اي ان تباين حد الخطأ سيكون ثابت او متجانس (Homo) لكل قيم مشاهدات المتغيرات المستقلة))


Heteroskedasticity
Y
X
Investopedia


Linear fits
Loess fits
q
p
1.0
0.5
0.0
-0.5
-1.0
p
q


Good Units 1509
regression on Good Units
Outliers 897
x 10^5
x 10^4

أسباب عدم ثبات تباين حد الخطأ

- تتعدد أسباب عدم ثبات تباين حد الخطأ لإرتباطها بصورة مباشرة بالسلوك الشخصي لأفراد المجتمع والذي تؤثر عليه على سبيل المثال لا الحصر مستويات التعليم ، زيادة دخول الأفراد ، طبيعة العمل ... إلخ .
- الاخطاء في القياس Measurement Errors
- اخطاء الحذف Omission Errors

آثار عدم ثبات تباين حد الخطأ

- إذا طبقت طريقة المربعات الصغرى العادية في تقدير نموذج ما مع وجود عدم ثبات تباين حد الخطأ فإن :

1. القيم المقدرة لمعاملات الإنحدار سوف تكون غير متحيزة .
2. تباين القيم المقدرة لمعاملات الإنحدار سوف لايكون أقل ما يمكن .
3. اختبار الفروض وتكوين حدود الثقة سوف يكون صعباً .
4. التنبؤ بإستخدام نتائج تقدير هذا النموذج سوف يكون غير ممكناً .

اكتشاف عدم ثبات تباين حد الخطأ

- يتم اكتشاف عدم ثبات تباين حد الخطأ بواسطة عدة اختبارات منها ما يلي :

1. اختبار Park .
2. اختبار Goldfeld-Quandt .
3. اختبار معامل ارتباط Spearman .

خطوات اختبار Park

1. تطبيق طريقة المربعات الصغرى العادية على المعادلة المطلوب تقديرها .
2. الحصول على البواقي أو القيم المقدرة لحد الخطأ من المعادلة .
3. استخدام طريقة المربعات الصغرى العادية في إجراء إنحدار $\ln e_i^2$ على $\ln X_i$ فينتج مايلي :

$$\ln \hat{e}_i^2 = \hat{\alpha} + \hat{\beta} \ln X_i$$

خطوات اختبار Park

4. ايجاد القيمة المحسوبة لاختبار t بالنسبة لـ β :

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

5. ايجاد القيمة الجدولية لإختبار t عند درجات حرية معينة (N-K+1) .
6. مقارنة القيمة المحسوبة لإختبار t بالقيمة الجدولية له :

خطوات اختبار Park

I. فإذا كانت قيمة t المحسوبة أكبر من قيمة t الجدولية يتم قبول الفرض البديل القائل بأن قيمة β المقدرة لا تساوي الصفر ويقال في هذه الحالة أن β معنوية إحصائياً ويدل هذا على وجود مشكلة عدم ثبات تباين حد الخطأ

II. أما إذا كانت قيمة t المحسوبة أقل من قيمة t الجدولية يتم قبول فرض العدم القائل بأن قيمة β المقدرة تساوي الصفر ويقال في هذه الحالة أن β غير معنوية إحصائياً ويدل هذا على وجود افتراض ثبات تباين حد الخطأ

مثال تطبيقي

Y_i	X_i	Y_i	X_i
8.0	10	20.6	25
8.2	10	18.6	25
8.3	10	22.4	25
8.1	10	23.1	25
8.7	10	15.1	25
12.3	15	24.2	30
9.4	15	16.7	30
11.6	15	27.0	30
12.0	15	26.0	30
8.9	15	22.1	30
15.0	20		
16.0	20		
12.0	20		
11.3	20		
19.1	20		

Observation	*Predicted Y_i*	e_i	e_i^2	$\ln e_i^2$	$\ln X_i$
1	7.588	0.412	0.169744	-1.77346	2.302585
2	7.588	0.612	0.374544	-0.98205	2.302585
3	7.588	0.712	0.506944	-0.67935	2.302585
4	7.588	0.512	0.262144	-1.33886	2.302585
5	7.588	1.112	1.236544	0.21232	2.302585
6	11.458	0.802	0.643204	-0.44129	2.70805
7	11.458	-2.098	4.401604	1.481969	2.70805
8	11.458	0.102	0.010404	-4.56556	2.70805
9	11.458	0.502	0.252004	-1.37831	2.70805
10	11.458	-2.598	6.749604	1.909484	2.70805
11	15.328	-0.408	0.166464	-1.79298	2.995732
12	15.328	0.592	0.350464	-1.0485	2.995732
13	15.328	-3.408	11.61446	2.452251	2.995732
14	15.328	-4.108	16.87566	2.825873	2.995732
15	15.328	3.692	13.63086	2.612337	2.995732
16	19.198	-0.218	0.047524	-3.04652	3.218876
17	19.198	1.282	1.643524	0.496843	3.218876
18	19.198	3.082	9.498724	2.251157	3.218876
19	19.198	3.782	14.30352	2.660506	3.218876
20	19.198	-4.218	17.79152	2.878722	3.218876
21	23.068	0.972	0.944784	-0.0568	3.401197
22	23.068	-6.528	42.61478	3.752201	3.401197
23	23.068	3.772	14.22798	2.655211	3.401197
24	23.068	2.772	7.683984	2.039138	3.401197
25	23.068	-1.128	1.272384	0.240892	3.401197

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-6.925	2.950	-2.347	0.028
$\ln X_i$	2.523	1.000	2.523	0.019

$$\ln \hat{e}_i^2 = -6.925 + 2.523 \ln X_i$$

- مقارنة قيمة T المحسوبة بالقيمة الجدولية عند درجات حرية 22 (2.074) نجد أن β معنوية إحصائياً ويدل هذا على وجود مشكلة عدم ثبات تباين حد الخطأ .

خطوات اختبار Goldfeld-Quandt

1. ترتيب مشاهدات X ترتيباً تصاعدياً
2. استبعاد المشاهدات الوسطى لكل من X و Y ، ثم تكوين مجموعتين من المشاهدات بحيث يكون لكل مجموعة على حدة معادلة خاصة بها .
3. تطبيق طريقة المربعات الصغرى العادية على المعادلتين كلاً على حدة .
4. الحصول على القيم المقدرة لحد الخطأ للمعادلتين .

خطوات اختبار Goldfeld-Quandt

5. إيجاد القيمة المحسوبة لإحصائية F كما يلي :

$$\hat{F} = \frac{\sum e_{2i}^2}{\sum e_{1i}^2}$$

6. ايجاد درجات الحرية لمعادلة F المحسوبة كما يلي :

$$DF = \frac{N - M - 2(K+1)}{2}$$

7. ايجاد القيمة الجدولية لإحصائية F عند درجات الحرية لكل من البسط والمقام ومستوى معنوية معين .

خطوات اختبار Goldfeld-Quandt

8. مقارنة بين القيم المحسوبة لإحصائية F والقيمة الجدولية لها :
   1. فإذا كانت قيمة F المحسوبة أكبر من الجدولية ، يتم قبول الفرض البديل القائل بوجود مشكلة عدم ثبات تباين حد الخطأ .
   2. أما إذا كانت قيمة F المحسوبة أقل من الجدولية يتم قبول فرض العدم المتمثل في وجود افتراض ثبات تباين حد الخطأ .

- يمكن تطبيق هذا الإختبار على أي متغير مستقل في نموذج الإنحدار المتعدد يُحتمل أن يكون المسبب في وجود مشكلة عدم ثبات تباين حد الخطأ .

مثال تطبيقي

Y_i	X_i	Y_i	X_i
8.0	10	20.6	25
8.2	10	18.6	25
8.3	10	22.4	25
8.1	10	23.1	25
8.7	10	15.1	25
12.3	15	24.2	30
9.4	15	16.7	30
11.6	15	27.0	30
12.0	15	26.0	30
8.9	15	22.1	30
15.0	20		
16.0	20		
12.0	20		
11.3	20		
19.1	20		

Y_{1i}	X_{1i}		Y_{2i}	X_{2i}
8.0	10		19.1	25
8.2	10		20.6	25
8.3	10		22.4	25
8.1	10		23.1	25
8.7	10		15.1	25
12.3	15		24.2	30
9.4	15		16.7	30
11.6	15		27.0	30
12.0	15		26.0	30
8.9	15		22.1	30

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	3.100	1.819	1.704	0.127
Xi	0.516	0.143	3.616	0.007

$$\hat{Y}_{1i} = 3.10 + 0.516 X_i$$

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	4.360	12.790	0.341	0.742
Xi	0.628	0.463	1.356	0.212

$$\hat{Y}_{2i} = 4.36 + 0.628 X_i$$

Predicted Y_{1i}	e_{1i}	e_{1i}^2
8.26	-0.26	0.0676
8.26	-0.06	0.0036
8.26	0.04	0.0016
8.26	-0.16	0.0256
8.26	0.44	0.1936
10.84	1.46	2.1316
10.84	-1.44	2.0736
10.84	0.76	0.5776
10.84	1.16	1.3456
10.84	-1.94	3.7636
		10.184

Predicted Y_{2i}	e_{2i}	e_{2i}^2
20.06	-0.96	0.9216
20.06	0.54	0.2916
20.06	2.34	5.4756
20.06	3.04	9.2416
20.06	-4.96	24.6016
23.2	1.00	1.00
23.2	-6.50	42.25
23.2	3.80	14.44
23.2	2.80	7.84
23.2	-1.10	1.21
		107.272

- ايجاد القيمة المحسوبة لإحصائية F كما يلي :

$$\hat{F} = \frac{\sum e_{2i}^2}{\sum e_{1i}^2} = \frac{107.272}{10.184} = 10.53$$

- ايجاد درجات الحرية كما يلي :

$$DF = \frac{N - M - 2(K+1)}{2} = \frac{25 - 5 - 2(3)}{2} = 7$$

- بالبحث في جدول F عند درجات حرية عددها 7 لكل من البسط والمقام ومستوى معنوية 5% تبين أن قيمة F الجدولية تساوي 3.79 . وعليه يقبل الفرض البديل القائل بوجود مشكلة عدم ثبات تباين حد الخطأ .

اختبار معامل ارتباط الرتب Spearman

- يتم تقدير الدالة بطريقة المربعات الصغرى الاعتيادية والحصول على قيم ($\hat{Y}$) و ($\hat{X}$)
- ايجاد حد الخطأ العشوائي من خلال :ei=Yi- ŷi
- ايجاد القيم المطلقة لحد الخطأ العشوائي / ei /
- ترتب / ei / و X المقدرة ترتيبا تصاعديا او تنازليا
- يتم اعطاء رتب (ارقام) تسلسلية مع اخذ متوسط الرتب للقيم المتكررة ومن ثم يحسب منها فروق الرتب

- يتم احتسابSpearman معامل ارتباط الرتب

$$rs = 1 - \frac{6 \sum_{Di}{}^{2}}{n(n^2-1)}$$

Di = R/ei/ - Rxi

$$trs^* = \frac{rs\sqrt{n-k}}{\sqrt{1-rs^2}}$$

- If (trs*) > t
  - نقبل فرض عدم تجانس التباين اي نقبل افتراض (Hetero) والمشكلة موجودة
- If (trs*) < t
  - نقبل فرض ثبات التباين اي نقبل افتراض (Homo) والمشكلة غير موجودة

Homework

From this table assume that (Y) is the Consumption Income for a person , and (X) is the Disposable Income for same person .

Find :

Is this model consider a Heteroscedasticity Problem or not ? Using a spearman method to detect it .

معالجة عدم ثبات تباين حد الخطأ

- تتم معالجة مشكلة عدم ثبات تباين حد الخطأ من خلال إجراء تحويل للنموذج الأصلي . ويتوقف شكل النموذج الأصلي المحول على نمط عدم ثبات تباين حد الخطأ المكتشف في النموذج الأصلي المقدر .

1. التحويل اللوغاريتمي المزدوج سوف يؤدي غالباً إلى تقليل درجة عدم ثبات تباين حد الخطأ . ومن ثم طبقاً لهذا الإفتراض تكون المعادلة المحولة المناسبة للنموذج الأصلي كما يلي :

$$\ln Y_i = \alpha + \beta \ln X_i + \in_i$$

Year	Yi	Xi
2002	26.1	38.3
2003	29.3	43.5
2004	35.6	53.5
2005	39.4	60.8
2006	42.7	66.4
2007	46.43	71.2
2008	50.1	77.2
2009	54.5	86.1
2010	60.1	94.6
2011	64.9	102.4
2012	69.2	109.9
2013	73.1	115.6

معالجة عدم ثبات تباين حد الخطأ

2. من أبرز الطرق المستخدمة لتصحيح هذه المشكلة هي طريقة المربعات الصغرى المرجحة GLS وذلك بتحويل قيم المتغيرات عن ضربها في مقلوب الإنحراف المعياري للبواقي e_i ، فيصبح النموذج المعدل كما يلي :

$$\frac{Y_i}{\delta} = \alpha \frac{1}{\delta} + \beta_1 \frac{X_{1i}}{\delta} + \beta_2 \frac{X_{2i}}{\delta} + \frac{\in_i}{\delta}$$

الارتباط الخطي المتعدد
Multicollinearity

- وهو يعني وجود ارتباط بين المتغيرات المستقلة في نموذج الإنحدار المراد تقديره .
- يعني وجود ارتباط خطي تام بين بعض أو كل المتغيرات المستقلة في نموذج الإنحدار .
- فإذا كان معامل الإرتباط بين متغيرين مستقلين يساوي الواحد الصحيح فإن معاملات الإنحدار تكون غير محددة .
- أما إذا كان معامل الارتباط بينهما يساوي الصفر ، فلا يوجد هناك أي مشاكل تتعلق بتقديرات معاملات الإنحدار .
- ومن ثم يمكن القول أن مشكلة الإزدواج الخطي تؤثر على دقة تقديرات معاملات الإنحدار .

أسباب الازدواج الخطي

1. إتجاه المتغيرات الإقتصادية معاً للتغير مع مرور الزمن
2. استخدام متغيرات مستقلة ذات فترات إبطاء في المعادلة المراد تقديرها .
3. استخدام متغيرات مستقلة ذات علاقة صريحة في المعادلة المراد تقديرها

آثار الازدواج الخطي

1. القيم المقدرة لمعاملات الانحدار سوف تكون غير محددة ، ومن ثم سوف تكون غير دقيقة .
2. الأخطاء المعيارية للقيم المقدرة لمعاملات الانحدار سوف تكون كبيرة جداً

اكتشاف الإزدواج الخطي

- هناك عدة اختبارات لاكتشاف الإزدواج الخطي ، من أبرزها ايجاد مصفوفة معامل الارتباط بين المتغيرات المستقلة في المعادلة المراد تقديرها .
- معامل الارتباط القوي يشير لوجود ارتباط خطي بين المتغيرين المستقلين المعنيين .

معالجة الإزدواج الخطي

1. زيادة حجم المشاهدات
2. إحلال متغيرات ذات فترات إبطاء محل المتغيرات المستقلة الأخرى في نماذج فترات الإبطاء الموزعة .
3. حذف متغير أو أكثر من المتغيرات المستقلة .
4. إضافة معادلات جديدة للنموذج .

- إن وجود الإزدواج الخطي لا يعتبر مشكلة وإنما المشكلة تتمثل في درجة الإزدواج الخطي