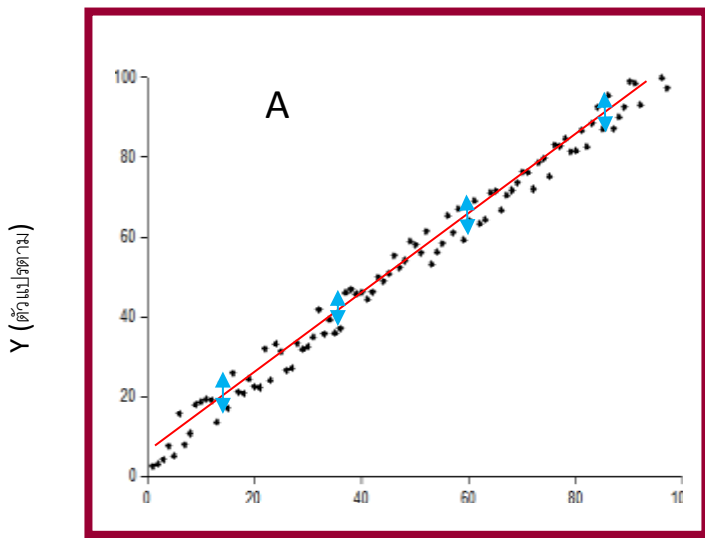
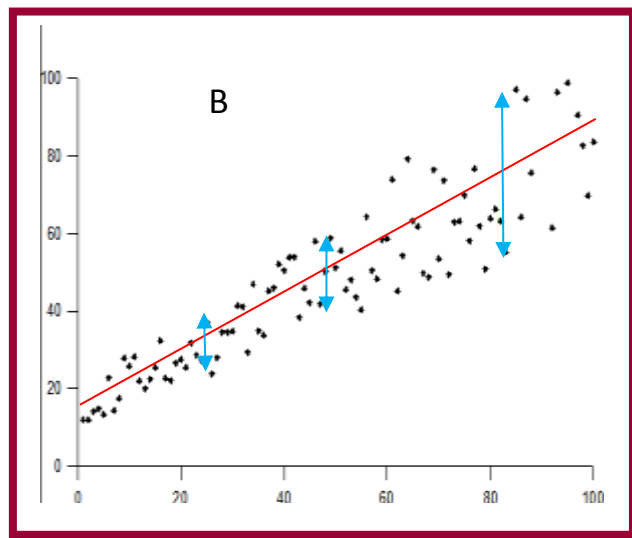


Heteroscedasticity

เมื่อใดก็ตามที่นักวิจัยใช้การวิเคราะห์การถดถอยอย่างง่าย (simple regression analysis) หรือการวิเคราะห์การถดถอยพหุคูณ (multiple regression analysis) ผลที่ได้ทางสถิติในรูปสมการ (equation) ขึ้นอยู่กับข้อสมมติฐานว่า ความผันผวนในค่าของ error term จะต้องคงที่ ที่เรียกเป็นศัพท์ทางสถิติว่า homoscedasticity และต้องไม่มีลักษณะที่เรียกว่า heteroscedasticity อธิบายได้ด้วยรูปสองมิติดังต่อไปนี้



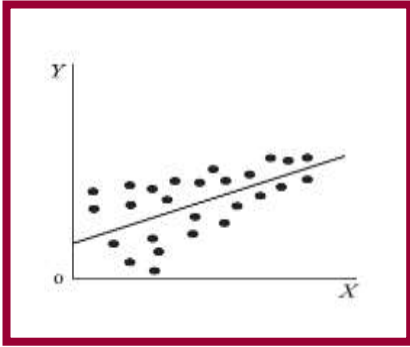
X(ตัวแปรอิสระ)



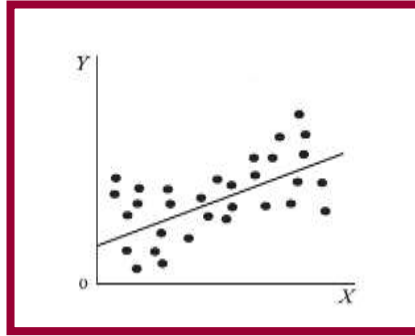
X(ตัวแปรอิสระ)

ในที่นี้แกนนอนแสดงค่าของตัวแปรอิสระ (independent variable) ส่วนแกนตั้งแสดงค่าของตัวแปรตาม (dependent variable) หากเราสมมติว่าเส้นสีแดงที่ลากผ่านจุดต่าง ๆ เป็นเส้นที่คำนวณได้จากการวิเคราะห์การถดถอย ระยะห่างในแนวตั้งจากจุดต่าง ๆ จากเส้นสีแดง (มีทั้งเป็นบวกและลบ) เป็นค่าของความคลาดเคลื่อนหรือ error terms ในภาพ A error terms จะมีความผันผวนคงที่รอบเส้นแดง แต่ในภาพ B error term จะมีความผันผวนไม่แน่นอน โดยจะอยู่ในระดับต่ำเมื่อค่าของ x มีค่าต่ำ และจะเพิ่มสูงขึ้นเมื่อค่าของ x เพิ่มขึ้นในกรณีเช่นนี้ error terms ตามภาพ A มีลักษณะที่เรียกว่า homoscedastic ส่วน error terms ตามภาพ B มีลักษณะที่เรียกว่า heteroscedastic

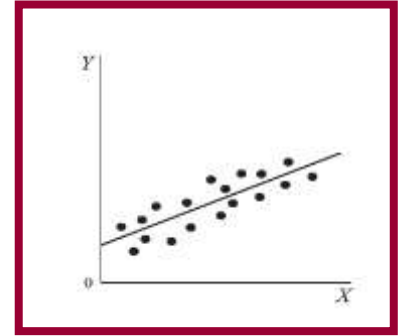
Error term ที่มีลักษณะ heteroscedastic อาจไม่ได้มีกรณีเดียว ลองพิจารณาภาพ 3 ภาพต่อไปนี้



ภาพที่ 1



ภาพที่ 2



ภาพที่ 3

ภาพที่ 1 แสดงถึง variance ของ error terms มีลักษณะ homoscedastic แต่แตกต่างจากภาพ B ด้านบนตรงที่ว่า ยิ่งค่าของตัวแปรอิสระยิ่งสูง variance ของ error terms จะต่ำ

ภาพที่ 2 แสดงถึง variance ของ error terms มีลักษณะ heteroscedastic เช่นเดียวกัน แต่ variance จะมีค่าต่ำเมื่อตัวแปรอิสระมีค่าอยู่ใกล้ๆ ค่ากลางๆ และ variance จะมีค่าสูงเมื่อตัวแปรอิสระมีค่าต่ำหรือสูง

ภาพที่ 3 แสดงถึง variance ของ error terms มีลักษณะ homoscedastic

การที่ error terms มีความผันผวนที่มีลักษณะเป็น heteroscedastic ผลข้างเคียงที่เกิดขึ้นก็คือ การทดสอบสมมติฐานทางสถิติต่าง ๆ อาจไม่มีความน่าเชื่อถือ เนื่องจากค่าของ standard error ที่คำนวณได้ อาจต่ำกว่าความเป็นจริง

จะตรวจสอบการมีheteroscedasticity ได้อย่างไร?

ในกรณีที่เป็นการวิเคราะห์การถดถอยอย่างง่าย (simple regression) และ error terms มีลักษณะเช่นภาพ B Goldfeld-Quandt ได้เสนอวิธีการดังนี้

1. ให้เรียงลำดับตัวแปรอิสระและตัวแปรตามขึ้นใหม่ โดยยึดถือค่าของตัวแปรอิสระเป็นหลักจากน้อยไปหามาก
2. แบ่งข้อมูลออกเป็นสองช่วงจำนวนเท่า ๆ กัน โดยเว้นข้อมูลช่วงกลางไว้เล็กน้อย
3. คำนวณหา regression ของตัวแปรตามกับตัวแปรอิสระจากข้อมูลทั้งสองช่วง เราจะได้ข้อมูลเกี่ยวกับ error sum of squares (ESS)
4. นำ ESS ช่วงที่ตัวแปรอิสระมีค่าสูงหารด้วย ESS ช่วงที่ตัวแปรอิสระมีค่าต่ำ
5. นำผลที่ได้ไปเปรียบเทียบกับค่า F ที่ degree of freedom ทั้งเศษและส่วนเท่ากับ $(n-d-2k)/2$ หากมีค่าสูงกว่า แสดงว่า variance ของ error terms มีลักษณะที่เป็น heteroscedastic ในที่นี้ n = ขนาดของตัวอย่างทั้งหมด d = ขนาดของข้อมูลที่เว้นไว้ k = ค่าของ parameters ที่ต้องการหาค่า (ในกรณีของสมการถดถอยอย่างง่าย ต้องหาทั้งค่าคงที่ และค่าสัมประสิทธิ์ของตัวแปร X ดังนั้น $k=2$)

เพื่อให้เห็นการทดสอบheteroscedasticity ตามแนวของ Goldfeld-Quandt เราจะใช้ข้อมูลของค่าจ้างโดยเฉลี่ย (Y) จำนวนพนักงานที่จ้าง (X) ของบริษัท 30 บริษัทเป็นตัวอย่าง

บริษัท	ค่าจ้างเฉลี่ย	แรงงาน	บริษัท	ค่าจ้างเฉลี่ย	แรงงาน	บริษัท	ค่าจ้างเฉลี่ย	แรงงาน
1	8.40	100	11	9.40	200	21	10.90	400
2	8.40	100	12	9.60	200	22	11.30	400
3	8.60	100	13	9.50	300	23	11.50	400
4	8.70	100	14	9.80	300	24	11.70	400
5	8.90	100	15	9.90	300	25	11.60	500
6	9.00	100	16	10.30	300	26	11.80	500
7	8.90	200	17	10.30	300	27	12.10	500
8	9.10	200	18	10.50	300	28	12.50	500
9	9.30	200	19	10.30	400	29	12.70	500
10	9.30	200	20	10.60	400	30	13.10	500

เราจะใช้โปรแกรมสำเร็จรูป EVIEWS ช่วยในการวิเคราะห์ของเรา

ภาพด้านล่างแสดงข้อมูลworkfile ของ EVIEWS

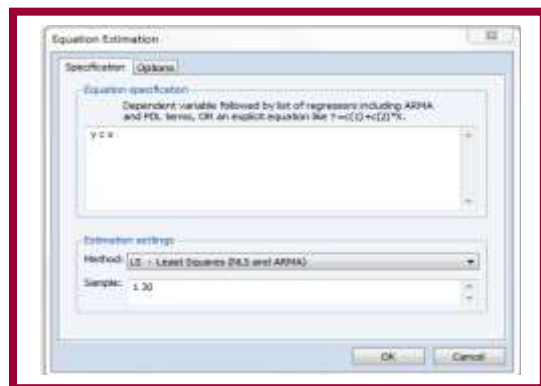
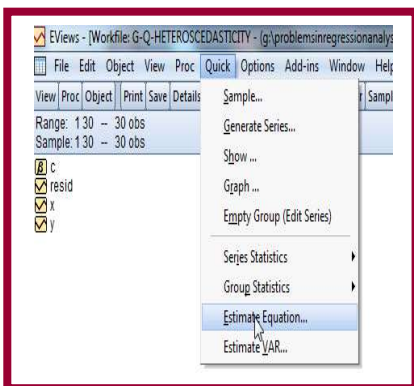
obs	X	Y
1	100	8.4
2	100	8.4
3	100	8.6
4	100	8.7
5	100	8.9
6	100	9.0
7	200	8.9
8	200	9.1
9	200	9.3
10	200	9.3
11	200	9.4
12	200	9.6
13	300	9.5
14	300	9.8
15	300	9.9
16	300	10.3
17	300	10.3
18	300	10.5
19	400	10.3
20	400	10.6
21	400	10.9
22	400	11.3
23	400	11.5
24	400	11.7
25	500	11.6
26	500	11.8
27	500	12.1
28	500	12.5
29	500	12.7
30	500	13.1

ในที่นี้ Y= ค่าจ้างโดยเฉลี่ย (ข้อมูลคอลัมน์ขวาสุดของworkfile)

X= จำนวนแรงงาน (ข้อมูลคอลัมน์กลางของ workfile)

•คลิกที่ปุ่ม Quick>Estimate Equation

----->พิมพ์ y c x แล้วระบุ Sample 1 30(ลำดับที่ 1-30)



Computer output จะปรากฏในตารางด้านล่าง

Dependent Variable: Y
 Method: Least Squares
 Date: 06/01/15 Time: 18:41
 Sample: 1 30
 Included observations: 30

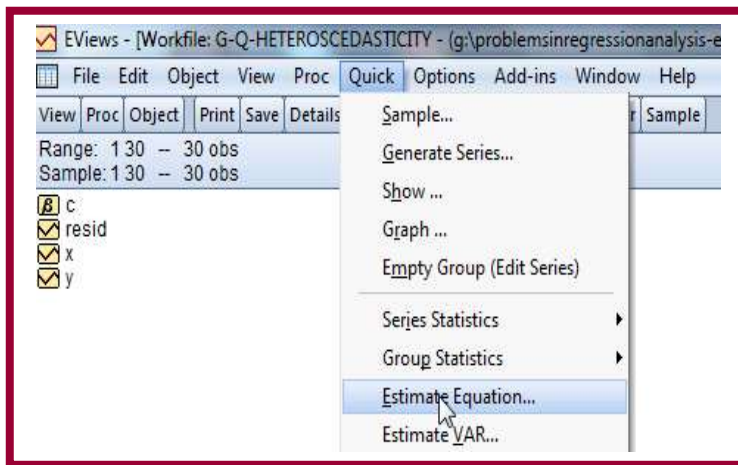
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.551667	0.188125	40.14177	0.0000
X	0.009050	0.000567	15.95507	0.0000

R-squared	0.900907	Mean dependent var	10.26667
Adjusted R-squared	0.897368	S.D. dependent var	1.371466
S.E. of regression	0.439365	Akaike info criterion	1.257368
Sum squared resid	5.405167	Schwarz criterion	1.350781
Log likelihood	-16.86052	Hannan-Quinn criter.	1.287252
F-statistic	254.5642	Durbin-Watson stat	1.052900
Prob(F-statistic)	0.000000		

สมการที่ได้คือ $Y = 7.552 + 0.009X$
 (40.14) (15.96)

$R^2 = 0.90$

เราจะแบ่งข้อมูลออกเป็นสองส่วนอย่างละเท่า ๆ กัน (ลำดับที่ 1-12 และ 19-20) โดยคงเหลือข้อมูลตรงกลางตั้งแต่ลำดับที่ 13-18 คลิกที่ปุ่ม Quick>Estimate Equation ให้กรอกคำสั่งเหมือนเดิม แต่ระบุข้อมูลลำดับที่ 1-12



Computer output ของสมการถดถอยจาก X 12 ตัวแรก แสดงได้ดังภาพต่อไปนี้

Dependent Variable: Y
 Method: Least Squares
 Date: 06/01/15 Time: 18:42
 Sample: 1 12
 Included observations: 12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.066667	0.224846	35.87647	0.0000
X	0.006000	0.001422	4.219265	0.0018

R-squared	0.640316	Mean dependent var	8.966667
Adjusted R-squared	0.604348	S.D. dependent var	0.391578
S.E. of regression	0.246306	Akaike info criterion	0.186528
Sum squared resid	0.606667	Schwarz criterion	0.267346
Log likelihood	0.880832	Hannan-Quinn criter.	0.156606
F-statistic	17.80220	Durbin-Watson stat	1.186813
Prob(F-statistic)	0.001774		

สมการที่ได้คือ $Y = 8.067 + .006X$
 (35.88) (4.22)

$R^2 = 0.64$

$ESS_1 = 0.607$

ในส่วน computer output ของสมการถดถอยจาก X 12 ตัวหลัง แสดงได้ดังภาพต่อไปนี้

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.050000	1.454275	4.160148	0.0019
X	0.012500	0.003212	3.891709	0.0030

R-squared	0.602313	Mean dependent var	11.67500
Adjusted R-squared	0.562544	S.D. dependent var	0.841130
S.E. of regression	0.556327	Akaike info criterion	1.816092
Sum squared resid	3.095000	Schwarz criterion	1.896909
Log likelihood	-8.896550	Hannan-Quinn criter.	1.786170
F-statistic	15.14540	Durbin-Watson stat	0.882876
Prob(F-statistic)	0.003001		

$$\text{สมการที่ได้คือ } Y = 6.05 + 0.0125X$$

$$(4.16) \quad (3.89)$$

$$R^2 = 0.60$$

$$ESS_2 = 3.095$$

เนื่องจาก $ESS_2/ESS_1 = 5.099$ ซึ่งสูงกว่าค่า $F_{10,10}$ ซึ่งมีค่าเท่ากับ 2.97 ดังนั้น heteroscedasticity มีอยู่จริง

ในกรณีของการวิเคราะห์การถดถอยพหุคูณ (Multiple regression analysis) ที่มีหลายมิติ สิ่งที่ทำทายนักวิจัยคือ ไม่สามารถแสดงออกมาในรูปสองมิติได้ ยิ่งไปกว่านั้นหากมีตัวแปรอิสระสองตัว การเรียงลำดับขนาดของตัวแปรอิสระอาจทำให้นักวิจัยเกิดสับสนและไม่ง่ายเหมือนอย่างในการวิเคราะห์การถดถอยอย่างง่าย (Simple regression analysis)

หากเกิดกรณี heteroscedasticity จะต้องทำอะไรเพื่อให้มีทางออกทางสถิติ?

อาจต้องอาศัยการแปลง (transformation) ตัวแปรเดิมไปเป็นตัวแปรตัวใหม่อย่างในกรณีตัวอย่างที่ใช้ในเอกสารนี้ หาก error terms มี variance ที่เป็นสัดส่วนกับ X^2 หากเราหารสมการทั้งสมการด้วยตัวแปร X เราก็จะได้ตัวแปรใหม่ จากนั้นก็สามารถดำเนินการวิเคราะห์สมการถดถอยไปตามปกติ ซึ่งเอกสารนี้ของดเว้นไม่กล่าวถึง

Contribution this issue: ดร. ดนัย ปัตตพงษ์

อยากเรียนรู้การนำสถิติข้างต้นนี้ไปใช้ในการวิจัยระดับสารนิพนธ์ (independent study) วิทยานิพนธ์ (thesis) ดุษฎีนิพนธ์ (dissertation) ปรึกษาได้ที่ dpattaphongse@gmail.com / dpattaphongse@yahoo.com

- * ผู้แต่ง MBA's Made Easy (160+ issues) เอกสารวิชาการด้านศาสตร์การบริหารธุรกิจที่ช่วยให้ธุรกิจสามารถยืนหยัดและอยู่รอดได้ในภาวะที่โลกเปลี่ยนแปลงอยู่ตลอดเวลา
- * ผู้พัฒนา FINALYSIS... a dedicated software สำหรับช่วยในการวิเคราะห์ความเป็นไปได้ทางการเงินของโครงการพัฒนาอสังหาริมทรัพย์ (บ้านจัดสรร/จัดสรรที่ดินเพื่อการอุตสาหกรรม/อาคารชุด/อาคารสำนักงานให้เช่า) โรงแรม โรงพยาบาลเอกชน ห้างสรรพสินค้า โรงงานน้ำตาล โรงงานกระดาษ โรงไฟฟ้าชีวมวล ฯลฯ ได้เห็นตัวเลขก่อนโครงการเกิด หลีกเลี่ยงความผิดพลาดเป็นร้อยเป็นพันล้านบาทเกิดการลงทุนจริง (กำหนด debut 1 เมษายน 2569)
- * ผู้แต่งหนังสือ”การวิเคราะห์ความเป็นไปได้ทางการเงินและการจัดวงเงินเครดิตของโครงการลงทุน”ประกอบด้วยตัวอย่างของธุรกิจจริงที่ไม่เปิดเผยชื่อนับ 100 บริษัท ครอบคลุมอุตสาหกรรม 24 อุตสาหกรรม
- * Co-developer ซอฟต์แวร์ enogex@cel[®] สำหรับใช้ทดสอบ/เรียนรู้ศัพท์ (ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 90 บทครอบคลุมศัพท์ระดับ SAT/IELTS/TOEFL กว่า 12,000 คำ) และไวยากรณ์อังกฤษ (ประกอบด้วยแบบฝึกหัดและเฉลยกว่า 160 บทหรือกว่า 10,000 ข้อครอบคลุมเนื้อหาในระดับอุดมศึกษาและTOEFL) มาพร้อมกับไฟล์เสียง/ไฟล์ข้อมูล/ฯลฯ อีกมาก (กำหนด debut 1 เมษายน 2569)