เอกสารประกอบการสอน E-Views

The Violation on the Assumptions of the Classical Model and Remedial Measures

นิติพงษ์ ส่งศรีโรจน์

22 **กุมภาพันธ์** 2550

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ปัญหา MULTICOLLIEARITY

การประมาณสมการถดถอยพหุคูณหรือกรณีที่มีตัวแปรอิสระมากกว่า 1 ตัวแปรนั้น ในทางทฤษฎีนั้นเราได้สมมติว่าตัวแปรอิสระเหล่านั้นต้องไม่มีความสัมพันธ์กัน การเกิด Multicollinearity นั้นอาจจะเกิดได้หลายสาเหตุ ดังนี้

(1) วิธีในการเก็บรวบรวมข้อมูลซึ่งการเก็บข้อมูลตัวแปรอิสระนั้นค่าของมันอาจจะถูกจำกัดช่วง

(2) ข้อจำกัดของโมเดลหรือประชากรที่ถูกสุ่ม เช่น ต้องการประมาณสมการการบริโภคไฟฟ้ากับรายได้และ ขนาดของที่พักอาศัย ซึ่งตัวแปรสองตัวนั้นอาจจะมีความสัมพันธ์กัน เนื่องจากครอบครัวที่มีรายได้สูงมักจะมีที่พักอาศัยขนาดใหญ่ด้วย

(3) การกำหนดรูปแบบฟังก์ชันหรือโมเดล (Model Specification) เช่นการเพิ่มตัวแปรที่เป็นโพลิโนเมียลเข้าไปในสมการ เช่น X กับ X^2 ซึ่งเป็นกรณีที่ค่า X นั้นมีช่วงที่แคบ

(4) Overdetermined model เป็นกรณีที่โมเดลนั้นมีจำนวนตัวแปรอิสระมากกว่าจำนวนของข้อมูล เช่น มีตัวแปรอิสระ 20 ตัว แต่ข้อมูลที่จะนำมาประมาณสมการถดถอยนั้นมีเพียง 10 ตัวอย่างหรือ 10 ปีเป็นต้น ซึ่งจะเกิดขึ้นได้กับกรณีที่การเก็บข้อมูลของผู้ป่วยซึ่งอาจจะมีผู้ป่วยเพียงจำนวนน้อยแต่มีการเก็บข้อมูลของผู้ป่วยจำนวนที่มากกว่า

เหตุผลอื่นๆ กรณีที่เกิด Multicollinearity เช่นกรณีที่เกิดกับข้อมูลแบบ Time Series ซึ่งตัวแปรอิสระในตัวแบบนั้นอาจจะมีลักษณะของแนวโน้ม หรือ Common Trend ที่คล้ายคลึงกัน เช่น การประมาณสมการการบริโภคกับ รายได้ ความมั่งคั่ง จำนวนประชากร ซึ่งตัวแปรอิสระเหล่านี้อาจจะมีอัตราการเพิ่มขึ้นหรือลดลงที่เหมือนกันไม่มากก็น้อยซึ่งทำให้ตัวแปรอิสระมีความสัมพันธ์กัน

ผลจากการเกิด MULTICOLLIEARITY

การเกิด Multicollinearity ทำให้คุณสมบัติของสมการถดถอยซึ่งเรียกว่า BLUE (Best Linear Unbiased Estimator) นั้นยังคงมีอยู่ เมื่อเกิด Multicollinearity แล้วโดยเฉพาะอย่างยิ่งมีความสัมพันธ์ของตัวแปรอิสระที่สูงหรือเป็นแบบ Perfect Multicollinearity แล้วเรามักจะเผชิญกับผลที่เกิดขึ้นดังนี้

(1) ตัวประมาณที่ได้จาก OLS (Ordinary Least Square) จะมีค่า Variance และ Covariance สูง

(2) ผลจากข้อ (1) ส่งผลให้ช่วงความเชื่อมั่นกว้างขึ้น ทำให้มีแนวโน้มยอมรับสมมติฐานหลัก H_0

(3) ผลจากข้อ (1) ทำให้ค่า t-ratio ของสัมประสิทธิ์ 1 ตัวหรือมากกว่า มีแนวโน้มไม่มีนัยสำคัญทางสถิติ

(4) t-ratio ของสัมประสิทธิ์ 1 ตัวหรือมากกว่าไม่มีนัยสำคัญ โดยที่มีค่า R^2 ที่สูงมากๆ หรือมีค่า R^2 ที่สูงมากๆ แต่ t-ratio ของสัมประสิทธิ์มีนัยสำคัญทางสถิติเพียงจำนวนน้อย (ตัวแปรอิสระมีนัยสำคัญทางสถิติเพียงไม่กี่ตัวแปร)

(5) ตัวประมาณด้วย OLS และค่า standard error มีความอ่อนไหวต่อข้อมูลที่มีการเปลี่ยนแปลงเพียงเล็กน้อย นั่นคือถ้าตัวแปรหนึ่งซึ่งมีข้อมูลค่าหนึ่งที่นำมาประมาณเมื่อเราเปลี่ยนค่ามันเพียงเล็กน้อยจะทำให้ค่าสัมประสิทธิ์เดิมกับค่าสัมประสิทธิ์อันหลังมีความแตกต่างกันมาก รวมทั้งทำให้ค่า Covariance แตกต่างจากเดิมมาก นอกจากนี้ตัวแปรดังกล่าวที่เคยมีนัยสำคัญก็กลับไม่มีนัยสำคัญ

(6) ผลของ Multicolliearity อาจจะทำให้ขนาดและเครื่องหมายของสัมประสิทธิ์เปลี่ยนไป

การตรวจสอบการเกิด MULTICOLLIEARITY

เราจะทราบได้อย่างไรว่าจริงๆแล้วการประมาณสมการของเรานั้นมีปัญหา Multicolliearity หรือไม่ โดยเฉพาะอย่างยิ่งหากโมเดลของเรามีตัวแปรอิสระมากกว่าสองตัวแปร เนื่องจากปัญหา Multicolliearity เป็นปรากฏการณ์ของหนึ่งที่เกิดขึ้นกับข้อมูลกลุ่มตัวอย่างที่นำมาประมาณสมการถดถอย เราไม่มีเครื่องมือที่เป็นมาตรฐานที่ใช้ค้นหาปัญหา Multicolliearity สิ่งที่ทำได้คือการใช้หลักการทั่วๆไปทั้งที่เป็นทางการและไม่เป็นทางการ ซึ่งเราพอที่จะตรวจสอบถึงปัญหา Multicolliearity ได้ โดยมีประเด็นที่ใช้ในการพิจารณาดังต่อไปนี้

(1) ค่า R^2 สูงแต่ตัวแปรอิสระมีนัยสำคัญทางสถิติเพียงไม่กี่ตัวแปร

(2) ค่าความสัมพันธ์แบบจับคู่ของตัวแปรอิสระมีค่าสูง เช่น เกิน 0.80 ถือว่ามีปัญหา Multicolliearity ที่ไม่สามารถยอมรับได้

(3) ตรวจสอบค่า Partial Correlations

(4) ใช้ Auxiliary Regression

(5) พิจารณา Tolerance and Variance Inflation Factor (TOL and VIF)

การแก้ปัญหา MULTICOLLIEARITY

เมื่อเกิดปัญหา Multicolliearity ที่รุนแรง เราจะทำอย่างไรกับปัญหาดังกล่าว มีวิธีที่จะลดปัญหาดังกล่าวหรือขจัดให้หมดไปได้หรือไม่ เรามีทางเลือกอยู่ 2 ประการ คือ ไม่ต้องทำอะไรเลย กับ ใช้หลักเกณฑ์ที่ใช้กันทั่วไป

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

กรณีที่ไม่ต้องทำอะไรเลย เนื่องจากปัญหา Multicolliearity นั้นเป็นความบกพร่องของข้อมูลที่เราใช้ในการประมาณและเราก็ไม่มีทางเลือกซึ่งต้องยอมรับในข้อมูลที่มีอยู่นั้นเพื่อใช้ในการวิเคราะห์เชิงประจักษ์ ถึงแม้ว่าตัวแปรเหล่านั้นอาจจะไม่มีนัยสำคัญทางสถิติ กรณีที่ใช้หลักเกณฑ์ทั่วไปเพื่อบรรเทาหรือขจัดปัญหา Multicolliearity นั้นสามารถพิจารณาในประเด็นดังต่อไปนี้

(1) A priori information หรือการกำหนดข้อจำกัดให้กับโมเดลก่อน เช่น เราต้องประมาณสมการการบริโภค (Y_i) กับรายได้ (X_1) และความมั่งคั่ง (X_2) นั่นคือ $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + U_i$ และพบว่า รายได้กับความมั่งคั่งนั้นมีความสัมพันธ์กันสูง เพื่อแก้ไขปัญหาดังกล่าว และเราเชื่อว่าสัมประสิทธิ์ของ β_2 น่าจะมีค่าเป็น 0.10 เท่าของ β_1 หรือ $\beta_2 = 0.10\beta_1$ ดังนั้นเราจะพบว่า $Y_i = \beta_0 + \beta_1(X_1 + 0.10X_2) + U_i$ แล้วเราจะทราบได้อย่างไรว่าข้อมูลจำเป็นอะไรบ้างที่เราต้องทราบมาก่อน เราสามารถทราบได้จากงานวิจัยเชิงประจักษ์ในอดีตและทฤษฎีในสาขานั้นๆ

(2) การใช้ข้อมูลทั้งแบบภาคตัดขวางและแบบ Time series ร่วมกัน

(3) การละทิ้งตัวแปรอิสระบางตัวแต่จะทำให้เกิด Specification error หรือ Specification bias แต่การละทิ้งตัวแปรอิสระสำคัญในทางทฤษฎีปัญหาจะมีความรุนแรงกว่า เพราะว่าถึงแม้จะมีปัญหา Multicollinearity แต่การประมาณสมการก็ยังเป็น BLUE

(4) การแปลงค่าตัวแปร (Transformation of variavles) กรณีข้อมูลแบบ Time series เราจะใช้วิธีการที่เรียกว่า first difference form กล่าวคือ เราแปลงจากสมการ $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t$ ให้อยู่ในรูป $Y_t - Y_{t-1} = \beta_0 + \beta_1(X_{1t} - X_{1,t-1}) + \beta_2(X_{2t} - X_{2,t-1}) + U_t - U_{t-1}$ (หมายเหตุ: อาจทำให้ Y_t ไม่เป็น stationary นั่นคือ ค่าเฉลี่ยและค่าความแปรปรวนมีการเปลี่ยนแปลงไปตามเวลา) การแปลงข้อมูลรูปแบบอื่นๆ ที่ใช้ซึ่งเรียกว่า ratio transformation เช่น การประมาณสมการค่าใช้จ่ายในการบริโภค (Y_t) กับ real GDP (X_{2t}) และ ประชาการ (X_{3t}) เนื่องจากประชากรและ real GDP นั้นมีอัตราเพิ่มขึ้นตลอดเวลา ดังนั้นเป็นไปได้ว่าตัวแปรสองตัวจะมีความสัมพันธ์กัน วิธีแก้อาจทำได้โดย การนำประชากรมาหารทุกตัวแปรในสมการ หรือทำให้อยู่ในรูปแบบ per capita ทุกตัวแปร อย่างไรก็ตามวิธีการแปลงข้อมูลทั้งสองวิธีข้างต้นต้องทำอย่างระมัดระวัง เพราะอาจทำให้เกิดปัญหาตามมาอีก เช่น disturbance term อาจเกิด Serial Correlation และอาจจะทำให้ Degree of Freedom ลดลงไปโดยเฉพาะในกรณีที่มีกลุ่มตัวอย่างจำนวนน้อย นอกจากนี้กรณี $\frac{U_t}{X_{3t}}$ อาจจะเกิดปัญหา Heteroscedasticity

(5) การเพิ่มข้อมูลกลุ่มตัวอย่าง

(6) กรณีตัวแปรในสมการที่มีรูปโพลิโนเมียล เช่น การประมาณสมการต้นทุนทั้งหมด ตัวแปรที่อยู่ในรูปยกกำลังสอง ยกกำลังสามอาจจะมีความสัมพันธ์กันซึ่งอาจจะแก้ไขด้วยการทำให้อยู่ในรูป Devation Form หรือถ้าพบว่ายังมีปัญหาอาจจะใช้ Orthogonal polynomials

(7) อาจจะใช้ Factor Analysis, Ridge regression

ตัวอย่าง ปัญหา Multicollinearity โดยใช้ข้อมูลที่ปรากฏอยู่ใน Gujarati (2003) หน้า 371 ดังนี้ (นิสิตสามารถดาวน์โหลดข้อมูลได้ที่ http://www.kaset51.com (คลิ๊กหัวข้อเอกสารประกอบการสอน)

ตารางที่ 1 ข้อมูลสำหรับวิเคราะห์ปัญหา Multicollinearity

OBS	Y	X1	X2	X3	X4	X5	TIME
1947	60,323	830	234,289	2,356	1,590	107,608	1
1948	61,122	885	259,426	2,325	1,456	108,632	2
1949	60,171	882	258,054	3,682	1,616	109,773	3
1950	61,187	895	284,599	3,351	1,650	110,929	4
1951	63,221	962	328,975	2,099	3,099	112,075	5
1952	63,639	981	346,999	1,932	3,594	113,270	6
1953	64,989	990	365,385	1,870	3,547	115,094	7
1954	63,761	1,000	363,112	3,578	3,350	116,219	8
1955	66,019	1,012	397,469	2,904	3,048	117,388	9
1956	67,857	1,046	419,180	2,822	2,857	118,734	10
1957	68,169	1,084	442,769	2,936	2,798	120,445	11
1958	66,513	1,108	444,546	4,681	2,637	121,950	12
1959	68,655	1,126	482,704	3,813	2,552	123,366	13
1960	69,564	1,142	502,601	3,931	2,514	125,368	14
1961	69,331	1,157	518,173	4,806	2,572	127,852	15
1962	70,551	1,169	554,894	4,007	2,827	130,081	16

โดยที่ Y = จำนวนคนที่มีงานทำ (หน่วยเป็น 1,000 คน)

X1 = GNP implicit price deflator

X2 = GNP (หน่วยเป็นล้านดอลลาร์)

X3 = จำนวนคนที่ไม่มีงานทำ (หน่วยเป็น 1,000 คน)

X4 = จำนวนคนที่เป็นทหาร

X5 = จำนวนประชากรที่อายุมากกว่า 14 ปี ที่ไม่ได้สังกัดสถาบัน

TIME = ปี

สมมติว่าเราต้องการประมาณ Y กับตัวแปรอิสระข้างต้น ผลปรากฏในโปรแกรม E-Views ดังนี้

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ตารางที่ 2 ผลการประมาณสมการด้วย E-Views ปี 1947-1962

Dependent Variable: Y
Method: Least Squares
Date: 02/18/07 Time: 14:42
Sample: 1947 1962
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	77270.12	22506.71	3.433204	0.0075
X1	1.506187	8.491493	0.177376	0.8631
X2	-0.035819	0.033491	-1.069516	0.3127
X3	-2.020230	0.488400	-4.136427	0.0025
X4	-1.033227	0.214274	-4.821985	0.0009
X5	-0.051104	0.226073	-0.226051	0.8262
TIME	1829.151	455.4785	4.015890	0.0030

R-squared	0.995479	Mean dependent var	65317.00
Adjusted R-squared	0.992465	S.D. dependent var	3511.968
S.E. of regression	304.8541	Akaike info criterion	14.57718
Sum squared resid	836424.1	Schwarz criterion	14.91519
Log likelihood	-109.6174	F-statistic	330.2853
Durbin-Watson stat	2.559488	Prob(F-statistic)	0.000000

จากผลการประมาณ พบว่า R^2 มีค่าสูงมากแต่ มีตัวแปรสองสามตัวที่ไม่มีนัยสำคัญทางสถิติ (X1, X2 และ X5) เพื่อให้เห็นชัดเจนเราจะพิจารณา Correlation Matrix ของตัวแปรอิสระด้วย โดยใช้คำสั่ง Quick Group Statistics Correlations ซึ่งปรากฏตามตารางด้านล่าง ซึ่งพบว่าตัวแปรอิสระมีความสัมพันธ์กันสูง

ตารางที่ 3 Correlation Matrix

	X1	X2	X3	X4	X5	TIME
X1	1.000000	0.991589	0.620633	0.464744	0.979163	0.991149
X2	0.991589	1.000000	0.604261	0.446437	0.991090	0.995273
X3	0.620633	0.604261	1.000000	-0.177421	0.686552	0.668257
X4	0.464744	0.446437	-0.177421	1.000000	0.364416	0.417245
X5	0.979163	0.991090	0.686552	0.364416	1.000000	0.993953
TIME	0.991149	0.995273	0.668257	0.417245	0.993953	1.000000

เพื่อให้มั่นใจมากกว่าเดิมเราจะลองใช้วิธีที่เรียกว่า Auxiliary Regressions คือการประมาณสมการตัวแปรอิสระ X หนึ่งๆ กับตัวแปร X ที่เหลือ ซึ่งผลปรากฏดังนี้

ตารางที่ 4 Auxiliary Regressions

Dependent Variable: X1
Method: Least Squares
Date: 02/18/07 Time: 15:08
Sample: 1947 1962
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2044.583	533.3698	3.833331	0.0033
X2	0.002561	0.000948	2.700628	0.0223
X3	0.031922	0.015130	2.109831	0.0611
X4	0.008802	0.007478	1.176973	0.2665
X5	-0.017550	0.006331	-2.771998	0.0197
TIME	-9.992189	16.66535	-0.599579	0.5621

R-squared	0.992622	Mean dependent var	1016.812
Adjusted R-squared	0.988933	S.D. dependent var	107.9155
S.E. of regression	11.35293	Akaike info criterion	7.976825
Sum squared resid	1288.890	Schwarz criterion	8.266546
Log likelihood	-57.81460	F-statistic	269.0649
Durbin-Watson stat	1.870344	Prob(F-statistic)	0.000000

ซึ่งจะพบว่า กรณีประมาณสมการโดยให้ X1 เป็นตัวแปรตาม $R^2 = 0.9926$ ส่วนกรณีอื่นๆ ปรากฏดังนี้

ตารางที่ 5 R^2 ของ Auxiliary Regressions

ตัวแปรตาม	R^2	Tolerence (TOL)=1 - R^2
X1	0.9926	0.0074
X2	0.9994	0.0006
X3	0.9702	0.0298
X4	0.7213	0.2787
X5	0.9970	0.0030
Time	0.9986	0.0014

จากตารางข้างต้นจะพบว่ามีค่า R^2 ของ Auxiliary Regressions อยู่ 3 ใน 6 ค่าที่มีค่ามากกว่า R^2 กรณีตัวแปรตามเป็น Y ($R^2 = 0.99547$) ซึ่งสะท้อนให้เห็นว่าข้อมูลนั้นมีปัญหา Multicollinearity (หมายเหตุ: R^2 ของ Auxiliary Regressions ที่ได้มานั้น F-statistic จะต้องมีนัยสำคัญทางสถิติ

กรณีประเด็นเรื่องความอ่อนไหวของการเปลี่ยนแปลงข้อมูลกลุ่มตัวอย่าง เช่นตัดข้อมูลในปี 1962 ออกไป แล้วประมาณสมการใหม่อีกครั้งผลปรากฏดังนี้

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ตารางที่ 6 ผลการประมาณสมการด้วย E-Views ปี 1947-1961

Dependent Variable: Y
Method: Least Squares
Date: 02/18/07 Time: 15:37
Sample: 1947 1961
Included observations: 15

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	67271.28	23237.42	2.894954	0.0200
X1	-2.051082	8.709740	-0.235493	0.8197
X2	-0.027334	0.033175	-0.823945	0.4338
X3	-1.952293	0.476701	-4.095429	0.0035
X4	-0.958239	0.216227	-4.431634	0.0022
X5	0.051340	0.233968	0.219430	0.8318
TIME	1585.156	482.6832	3.284049	0.0111

R-squared	0.995512	Mean dependent var	64968.07
Adjusted R-squared	0.992146	S.D. dependent var	3335.820
S.E. of regression	295.6219	Akaike info criterion	14.52076
Sum squared resid	699138.2	Schwarz criterion	14.85119
Log likelihood	-101.9057	F-statistic	295.7710
Durbin-Watson stat	2.492491	Prob(F-statistic)	0.000000

จากตารางที่ 6 เมื่อเปรียบเทียบกับ ตารางที่ 2 พบว่า ขนาดของสัมประสิทธิ์ ขนาดของ Standard Error เปลี่ยนไปมาก รวมทั้งเครื่องหมายก็เปลี่ยนไปด้วย แล้วเราจะแก้ปัญหาเหล่านี้อย่างไร เนื่องจาก GNP อยู่ในรูป Nominal term เราสามารถเปลี่ยนในรูป real term ได้นั่น คือ x2/x1 = RGNP ส่วน X5 กับ X6 นั้นมีความสัมพันธ์กันเราจะตัด X6 ออกไปซึ่งเป็นค่าที่แสดงแนวโน้มเท่านั้น ส่วน X3 ก็ตัดออกเนื่องจากไม่มีเหตุผลสนับสนุนในทางทฤษฎี ดังนั้นผลการประมาณสมการใหม่ที่ได้ ดังแสดงในตารางที่ 7

ตารางที่ 7 ผลการประมาณสมการภายหลังการปรับโมเดลเพื่อลดปัญหา Multicollinearity

Dependent Variable: Y
Method: Least Squares
Date: 02/18/07 Time: 15:54
Sample: 1947 1962
Included observations: 16

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	65720.37	10624.81	6.185558	0.0000
RGNP	97.36496	17.91552	5.434671	0.0002
X4	-0.687966	0.322238	-2.134965	0.0541
X5	-0.299537	0.141761	-2.112965	0.0562

R-squared	0.981404	Mean dependent var	65317.00
Adjusted R-squared	0.976755	S.D. dependent var	3511.968
S.E. of regression	535.4492	Akaike info criterion	15.61641
Sum squared resid	3440470.	Schwarz criterion	15.80955
Log likelihood	-120.9313	F-statistic	211.0972
Durbin-Watson stat	1.654069	Prob(F-statistic)	0.000000

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

จากตารางที่ 7 พบว่าค่า R^2 ลดลงบ้างเล็กน้อยแต่ก็ยังคงมีค่าสูงและเครื่องหมายสัมประสิทธิ์ก็อธิบายเหตุผลทางเศรษฐศาสตร์ได้

ปัญหา AUTOCORRELATION

กรณีข้อมูลที่เป็น Time series นั้นเป็นไปได้ว่าข้อมูลในแต่ละช่วงเวลานั้นอาจจะมีความสัมพันธ์ซึ่งกันและกันได้ เมื่อเกิดสถานการณีเช่นนี้ก็จะทำให้ error term ไม่เป็นไปตามข้อสมมติเกี่ยวกับการประมาณสมการถดถอย นั่นคือ $E(u_i u_j) \neq 0$ โดยที่ $i \neq j$ จึงมีคำถามว่าแล้ว Autocorrelation มันเกิดจากอะไร ซึ่งมีหลายเหตุผลหลายประการ ดังนี้

(1) ความเฉื่อย (Inertia) หรือที่เรียกว่า Sluggishness ซึ่งหมายถึง ความเคลื่อนไหวอย่างช้าๆ ลักษณะเด่นของข้อมูล Time Series ในทางเศรษฐศาสตร์นั้นมักจะมี Sluggishness เช่น GNP ดัชนีราคา การผลิต การจ้างงาน หรือวัฏจักรธุรกิจ กล่าวคือ เมื่อเริ่มอยู่ที่จุดต่ำสุดของภาวะกาถดถอย เมื่อมีการฟื้นตัวทางเศรษฐกิจก็จะทำให้ข้อมูล Time Series ดังกล่าวเริ่มเคลื่อนไปทิศทางที่สูงขึ้น ดังนั้น โดยตัวของมันเองมันจึงมีแรงขับในตัวมันเองและดำเนินไปจนกว่าจะมีปัจจัยอย่างอื่นมากระทบที่ทำให้มันช้าลง เช่น การเพิ่มขึ้นของอัตราดอกเบี้ยหรือการเพิ่มภาษี เป็นต้น ดังนั้นข้อมูลของช่วงเวลาก่อนหน้าอาจจะส่งผลในช่วงเวลาถัดไป จึงทำให้ของมูลสองช่วงเวลาที่ต่อเนื่องกันมีความสัมพันธ์กันได้

(2) Specific Bias นั่นคือ กรณีที่ตัวแปรสำคัญไม่ได้อยู่ในโมเดล หรืออาจจะเป็นกรณีการกำหนดรูปแบบฟังก์ชันหรือโมเดลผิดพลาด เช่นฟังก์ชันต้นทุนส่วนเพิ่มปกติต้องเป็นรูปแบบที่มีผลผลิตยกกำลังสอง แต่กลับเป็นแบบฟังก์ชันเส้นตรง เป็นต้น

(3) ปรากฏการณ์ Cobweb ซึ่งมักจะเกิดกับอุปทานของสินค้าเกษตร กล่าวคือ การตัดสินใจทางด้านอุปทานจะพิจารณาจากราคาที่เกิดขึ้นในช่วงเวลาที่แล้ว (Lag 1 ช่วงเวลา)

(4) Lag กรณี การใช้จ่ายในการบริโภคกับรายได้ กับการใช้จ่ายการบริโภคที่เกิดขึ้นในช่วงเวลาที่แล้ว ซึ่งสามารถเขียนเป็นสมการ คือ

$Consumption_t = \beta_1 + \beta_2 income + \beta_3 consumption_{t-1} + u_i$ สมการนี้จะเรียกว่า “Autoregression” เนื่องจากมีตัวแปรอิสระที่เกิดจากการ Lag ตัวแปรตามในสมการ ถ้าหากไม่มีเทอม $consumption_{t-1}$ จะทำให้ error term แสดงถึงแบบแผนความเคลื่อนไหวที่เป็นระบบอันเนื่องมาจากอิทธิพลของการบริโภคในช่วงเวลาที่แล้วต่อการบริโภคในปัจจุบัน

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

(5) การจัดการข้อมูล (Manupulation of Data) เช่นการรวมข้อมูลจากรายไตรมาส เป็นข้อมูลรายเดือน หรือรายเดือนเป็นรายไตรมาส หรือกันเป็นรายปี หรือทำรายปีเป็นรายไตรมาส เป็นต้น นอกจากนี้การจัดการข้อมูลที่เรียกว่า Interpolation (วิธีคำนวณหาข้อมูลหรือ Series ที่ขาดหายไปเนื่องจากข้อมูลไม่ครบในระหว่างปี) และ Extrapolation (วิธีการคำนวณหาข้อมูลในอนาคต เช่นมีข้อมูลปี 1990-2000 แต่ต้องการข้อมูลเพิ่มจึงทำการประมาณข้อมูลขึ้นมาในอนาคตอีก 5 ปี เป็นต้น

(6) การแปลงข้อมูล (Data Transformation) ให้อยู่ในรูป Difference Form ถ้าสมการเดิมไม่มี autocorrelation เมื่อแปลงข้อมูลแล้ว error term ตัวใหม่ อาจจะทำให้เกิด autocorrelation

ผลของการเกิด AUTOCORRELATION

ถึงแม้ว่าจะเกิดปัญหา Autocorrelation แต่ตัวประมาณด้วยวิธี OLS ก็ยังคงเป็น Linear Unbiased รวมทั้งมีความคงเส้นคงวา (Consistency) และเป็น Asymptotically Normally distrinuted เพียงแต่ไม่เป็น Minimum Variance หรือกล่าวได้ว่าไม่เป็น BLUE และส่งผลให้ตัวแปรไม่มีนัยสำคัญทางสถิติ (ควรใช้ GLS จะดีกว่าในการทดสอบสมมติฐาน) นอกจากนี้ยังมีผลอื่นๆ ที่เกิดขึ้นเมื่อเรายังคงใช้ Variance ตัวประมาณที่เกิดปัญหา Autocorrelation นั่นคือ

(1) Residual Variance $\hat{\sigma}^2 = \sum \hat{u}_t^2 /(n-2)$ อาจจะมีค่าต่ำกว่าค่าจริง (σ^2)

(2) จากข้อ (1) ส่งผลให้ค่า R^2 มีค่าที่สูงเกินจริง

(3) ถึงแม้ σ^2 จะไม่มีค่าต่ำกว่าความเป็นจริง แต่ค่า Variance ภายใต้ AR(1) ก็ยังไม่มีประสิทธิภาพเท่ากับ Variance ของตัวประมาณกรณีที่ใช้ GLS

(4) ดังนั้นค่านัยสำคัญทางสถิติ t และ F อาจจะไม่ถูกต้อง เนื่องจากค่า Standard Error กรณี OLS จะต่ำกว่าความเป็นจริง ทำให้ค่า t มากผิดปกติซึ่งอาจจะส่งผลให้ปฏิเสธสมมติฐานหลักได้ง่ายหรือตัวประมาณมีนัยสำคัญนั่นเอง แต่ถ้า σ^2 มีค่าต่ำกว่าความเป็นจริงและค่า R^2 มาก ค่า F ก็จะยิ่งมากซึ่งจะทำให้ค่าสถิติมีนัยสำคัญได้ และหากนำไปใช้ก็จะทำให้ได้ข้อสรุปเกี่ยวกับนัยสำคัญทางสถิติของตัวประมาณผิดพลาดไปอย่างมาก

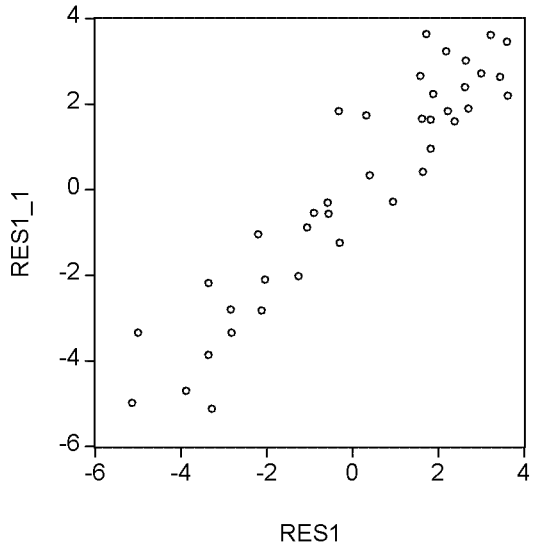
นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

การตรวจสอบการเกิด AUTOCORRELATION

การค้นหาปัญหา Autocorrelation นั้นมีวิธีการให้พิจารณาหลายแบบ ดังนี้

(1) วิธีกราฟ ซึ่งอาจจะใช้ Time sequence plot หรือใช้ Standardized residual plot นอกจากนี้อาจจะแสดงกราฟระหว่าง Residual ในเวลา t กับ t-1 ก็ได้ ดังในภาพที่ 1 และ 2

ภาพที่ 1 กราฟแสดง Residual กับ เวลา



ภาพที่ 2 กราฟแสดงความสัมพันธ์ Residual เวลา t กับ t-1

(2) Durbin-Watson d Test ค่าเข้าใกล้ 2 จะไม่มีปัญหา Autocorrelation ในการใช้วิธีการนี้มีข้อสมมติอยู่ว่า ประการที่หนึ่ง ตัวแปรอิสระของสมการถดถอยต้องเป็น nonstochastic หรือไม่เป็นตัวแปรเชิงสุ่ม ประการที่สอง error term จะต้องมีการกระจายแบบปกติ และประการที่สาม ในสมการถดถอยจะต้องไม่มีค่า lag ของตัวแปรตามที่เป็นตัวแปรอิสระ (เพราะถ้ามีค่า Durbin-Watson จะเข้าใกล้สองอยู่แล้ว นั่นคือไม่มี first-order autocorrelation) แต่ถ้ามี lag ของตัวแปรตาม Durbin ให้ใช้ h test แต่วิธีการนี้ก็ยังไม่มีอำนาจที่แข่งแกร่งเท่ากับวิธีของ Breusch-Godfrey Test

(3) ใช้ General Test of Autocorrelation เพื่อหลีกเหลี่ยงข้อบกพร่องที่เกิดจากการใช้ Durbin-Watson d Test ซึ่งมีชื่อว่า Breusch-Godfrey Test (BG Test) หรือ LM Test ซึ่งสามารถใช้ได้กับกรณี lag ของตัวแปรตามที่เป็นตัวแปรอิสระ ใช้กับ higher-order autoregressive เช่น AR(1), AR(2) เป็นต้น และใช้ได้กับกรณี simple หรือ higher-order moving averagevb ของ White noise error term วิธีการของ BG Test พอสรุปได้ดังนี้

3.1) ประมาณสมการถดถอยตามปกติ โดนวิธี OLS เพื่อให้ได้ค่า Residual : $\hat{u}_t$

3.2) ประมาณสมการ $\hat{u}_t$ กับตัวแปรอิสระทุกตัว รวมทั้ง $\hat{u}_{t-1}$, $\hat{u}_{t-2}$, ..., $\hat{u}_{t-\rho}$

3.3) จากข้อ 3.2) จะได้ R^2 โดยที่ถ้า n มีค่ามาก แล้ว $(n-\rho)R^2 \sim \chi^2_\rho$ ถ้าค่านี้เกิน Critical Chi-square value ณ ระดับนัยสำคัญที่เลือก ดังนั้นเราจะปฏิเสธสมมติฐานหลัก นั่นคือค่า ρ ไม่เท่ากับศูนย์ แสดงว่ามีปัญหา Autocorrelation อีกประการหนึ่งในการเลือกว่าจะใช้ ρ เท่ากับเท่าไรเพื่อมาใช้คำนวณหาค่าสถิตินั้นไม่มีหลักเกณฑ์ตายตัว แต่อาจจะใช้ Akaike กับ Schwarz Information Criteria (AIC กับ SIC) ศึกษาเพิ่มเติมได้ในตำราเศรษฐมิติ เช่น Gujarati (2003)

การแก้ปัญหา AUTOCORRELATION

เมื่อทราบได้แน่ชัดว่าเกิดปัญหา Autocorrelation แล้ว เราจะแก้ไขปัญหานี้อย่างไร มีทางเลือกอยู่ 4 ประการ ดังนี้

(1) ดูว่าปัญหา Autocorrelation เป็นแบบ Pure Autocorrelation หรือไม่ และโมเดลต้องมีการกำหนดที่ถูกต้องทั้งในแง่รูปแบบโมเดลและการไม่ละทิ้งตัวแปรที่สำคัญ

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

(2) ถ้าเป็น Pure Autocorrelation อาจจะใช้การแปลงฟังก์ชันโมเดลต้นแบบ หรือใช้วิธี Generalized Least-Square (GLS)

(3) กรณีมีตัวอย่างขนาดใหญ่ สามารถใช้ Newey-West Method เพื่อทำให้ค่า Standard Error มีความถูกต้อง

สำหรับกรณทางเลือกที่ได้กล่าวข้างต้นจะได้กล่าวถึงในรายละเอียด ดังต่อไปนี้

Model miss-Specification กับ Pure Autocorrelation

การทดสอบว่าเป็น Pure Autocorrelation หรือไม่ก็คือการนำเอา Time Trend เข้าไปในโมเดลด้วย ตัวอย่างเช่น จากสมการข้างล่าง โดยที่ Y_t คือ the index of real wage, X_t คือ the productivity index และ t คือ time trend การตีความจากสมการดังกล่าว กรณีตัวแปร t ก็คือ the index of real wage จะลดลงในอัตราประมาณ 0.90 หน่วยต่อปี และถ้า the productivity index เพิ่มขึ้น 1 หน่วย โดยเฉลี่ยแล้ว the index of real wage จะเพิ่มขึ้นประมาณ 1.30 หน่วย

$$\begin{aligned}\hat{Y}_t &= 1.4752 + 1.3057X_t - 0.9032t \\ se &= (13.18)\quad (0.2765)\quad (0.4203) \\ t &= (0.1119)\,(4.7230)\quad (-2.1490) \\ \mathrm{R}^2 &= 0.9631 \qquad d = 0.2046\end{aligned} \tag{1}$$

จากสมการที่ (1) ถึงแม้เราจะนำตัวแปร time trend เข้าไปในสมการแล้วแต่ก็ยังปรากฏว่าค่า d นั้นยังคงมีค่าที่ต่ำมาก ซึ่งแสดงว่าสมการดังกล่าวเกิด Pure Autocorrelation โดยที่ไม่ได้เป็นผลมาจากการกำหนดโมเดลผิด ซึ่งทราบได้จาก การกำหนดตัวแปร X^2 เพิ่มเติมเข้าไปในโมเดล (ไม่ต้องมี time trend) จะพบว่า ค่าสัมประสิทธิ์นั้นจะมีนัยสำคัญทางสถิติที่สูงทุกตัว แต่ค่า d นั้นยังคงต่ำหรือมีปัญหา Autocorrelation (คำเตือน: ค่า d นั้นใช้กับกรณีที่ residual เป็นการแจกแจงปกติเท่านั้น ซึ่งทดสอบโดยใช้ Jarque-Bera test of normality ในกรณีข้างต้นพบว่า residual มีการแจกแจงปกติ) ดังนั้นเราจึงมั่นใจได้ว่าโมเดลเริ่มต้นนั้นมีปัญหา Pure Autocorrelation จริง แล้วเราจะจัดการกับปัญหานี้อย่างไร วิธีการจัดการมีสองวิธีที่ได้กล่าวมาแล้ว คือ ใช้ GLS หรือไม่ก็ The Newey-West Method เพื่อทำให้ Standard Error ของ OLS นั้นมีค่าที่ถูกต้อง (รายละเอียด GLS และวิธีการทำให้ Standard Error ถูกต้อง ศึกษาเพิ่มเติมได้ใน Gujarati (2003: 477-487))

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ตัวอย่าง ปัญหา Autocorrelation โดยใช้ข้อมูลที่ปรากฏอยู่ใน Gujarati (2003) หน้า 460 ดังนี้ (นิสิตสามารถดาวน์โหลดข้อมูลได้ที่ http://www.kaset51.com (คลิ๊กหัวข้อเอกสารประกอบการสอน)) จากตารางที่ 8 ตัวแปร Y_t คือ the index of real wage, X_t คือ the productivity index

ตารางที่ 8 ข้อมูลสำหรับวิเคราะห์ปัญหา Autocorrelation

OBS	Y	X	OBS	Y	X
1959	58.5	47.2	1979	90	79.7
1960	59.9	48	1980	89.7	79.8
1961	61.7	49.8	1981	89.8	81.4
1962	63.9	52.1	1982	91.1	81.2
1963	65.3	54.1	1983	91.2	84
1964	67.8	56.6	1984	91.5	86.4
1965	69.3	58.6	1985	92.8	88.1
1966	71.8	61	1986	95.9	90.7
1967	73.7	62.3	1987	96.3	91.3
1968	76.5	64.5	1988	97.3	92.4
1969	77.6	64.8	1989	95.8	93.3
1970	79	66.2	1990	96.4	94.5
1971	80.5	68.8	1991	97.4	95.9
1972	82.9	71	1992	100	100
1973	84.7	73.1	1993	99.9	100.1
1974	83.7	72.2	1994	99.7	101.4
1975	84.5	74.8	1995	99.1	102.2
1976	87	77.2	1996	99.6	105.2
1977	88.1	78.4	1997	101.1	107.5
1978	89.7	79.5	1998	105.1	110.5

ประมาณสมการ Y_t คือ the index of real wage กับ X_t คือ the productivity index ผลปรากฏใน E-Views ดังนี้

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

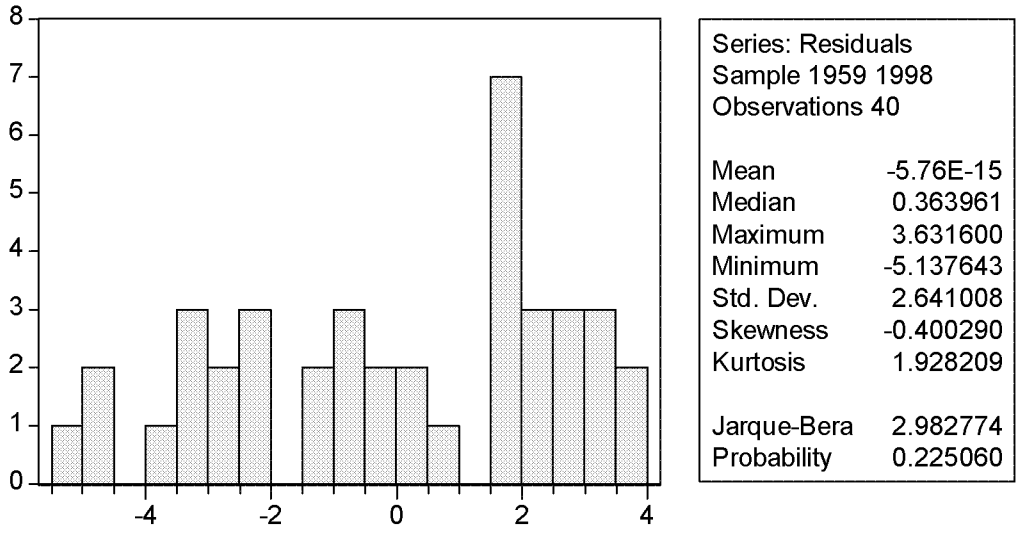
ตารางที่ 8 ผลการประมาณสมการด้วย E-Views ปี 1959-1998

Dependent Variable: Y
Method: Least Squares
Date: 02/20/07 Time: 13:53
Sample: 1959 1998
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29.51925	1.942347	15.19773	0.0000
X	0.713659	0.024105	29.60658	0.0000

R-squared	0.958449	Mean dependent var	85.64500
Adjusted R-squared	0.957356	S.D. dependent var	12.95632
S.E. of regression	2.675533	Akaike info criterion	4.854881
Sum squared resid	272.0220	Schwarz criterion	4.939325
Log likelihood	-95.09761	F-statistic	876.5495
Durbin-Watson stat	0.122904	Prob(F-statistic)	0.000000

วิธีการตรวจสอบเราจะดูที่ residual ด้วยวิธีกราฟก่อนซึ่งผลจะปรากฏดังในภาพที่ 1 และภาพที่ 2 ซึ่งแสดงว่ามีปัญหา Autocorrelation และเมื่อดูที่ว่า d พบว่ามีค่าต่ำมาก ไม่เข้าใกล้ 2 เลย และใช้ค่า d ได้ เนื่องจากทดสอบ Normality โดย Jarque-Bera test of normality แล้วพบว่า residual มีการแจกแจงแบบปกติ โดยการตรวจสอบให้เข้าไปหน้าจอ แสดงผลการประมาณสมการแล้ว เลือกที่ View Rsidual Test Histogram-Normality Test ปรากฏผลดังในภาพที่ 3 ถ้า Residual เป็นการแจกแจงปกติแล้ว Histogram ควรจะมีลักษณะเป็นระฆังคว่ำและค่า Jarque-Bera ต้องไม่มีนัยสำคัญทางสถิติ หาก rsidual ไม่เป็นไปตามข้อสมมติให้ใช้ Breusch-Godfrey Test (BG Test)



ภาพที่ 3 Jarque-Bera test of normality

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

แต่อย่างไรก็ตามเราก็ต้องสงสัยว่า ที่เกิดปัญหา Autocorrelation นั้นมันเป็น Pure Autocorrelation หรือไม่ หรือเป็นเพราะการกำหนดโมเดลผิด ซึ่งตรวจสอบโดยการนำ Time Trend เข้าไปในโมเดล ปรากฏผลดังในตารางที่ 9

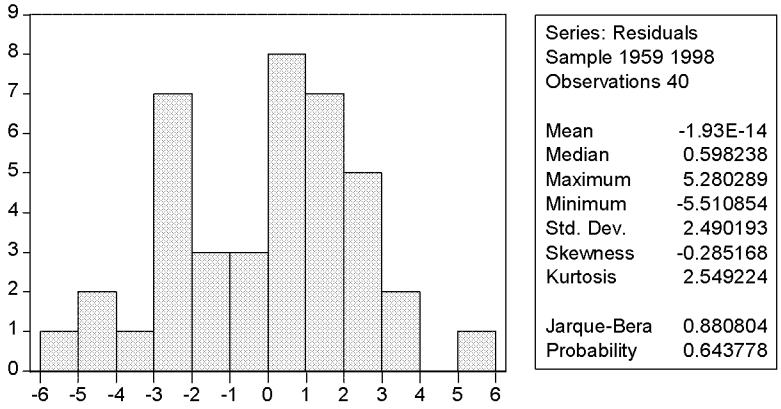
ตารางที่ 9 ผลการประมาณสมการด้วย E-Views ปี 1959-1998

Dependent Variable: Y
Method: Least Squares
Date: 02/20/07 Time: 14:22
Sample: 1959 1998
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.475191	13.18221	0.111908	0.9115
X	1.305693	0.276476	4.722620	0.0000
T	-0.903238	0.420341	-2.148822	0.0383

R-squared	0.963059	Mean dependent var	85.64500
Adjusted R-squared	0.961063	S.D. dependent var	12.95632
S.E. of regression	2.556609	Akaike info criterion	4.787279
Sum squared resid	241.8413	Schwarz criterion	4.913945
Log likelihood	-92.74559	F-statistic	482.3053
Durbin-Watson stat	0.204600	Prob(F-statistic)	0.000000

จะพบว่า residual ก็ยังคงมีการแจกแจงปกติ ดังแผนภาพที่ 4 และจากตารางที่ 9 ค่า d ก็ยังคงต่ำอยู่ และต่อไปจะดูว่ากำหนดโมเดลผิดพลาดหรือไม่โดยเพิ่ม X^2ในการประมาณสมการ ผลปรากฏดังแสดงในตารางที่ 10



ภาพที่ 4 Jarque-Bera test of normality

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ตารางที่ 10 ผลการประมาณสมการด้วย E-Views ปี 1959-1998 โดยเพิ่มตัวแปร X^2

Dependent Variable: Y
Method: Least Squares
Date: 02/20/07 Time: 14:32
Sample: 1959 1998
Included observations: 40

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-16.21817	2.954592	-5.489141	0.0000
X	1.948830	0.077994	24.98681	0.0000
X_SQUARED	-0.007917	0.000497	-15.93638	0.0000

R-squared	0.994716	Mean dependent var	85.64500
Adjusted R-squared	0.994431	S.D. dependent var	12.95632
S.E. of regression	0.966895	Akaike info criterion	2.842584
Sum squared resid	34.59077	Schwarz criterion	2.969250
Log likelihood	-53.85169	F-statistic	3482.880
Durbin-Watson stat	1.029978	Prob(F-statistic)	0.000000

จากตารางที่ 9 และ 10 พบว่าเมื่อกำหนดโมเดลใหม่แล้ว ค่า d ก็ยังต่ำอยู่ ไม่เข้าใกล้ 2 และเมื่อใส่ Time Trend เข้าไปแล้วค่า d ก็ยังคงต่ำมากและไม่เข้าใกล้ 2 จึงสรุปได้ว่าโมเดลเริ่มต้นนั้นมีปัญหา Pure Autocorrelation

การแก้ปัญหาข้างต้นจะเริ่มต้นด้วยวิธี GLS กรณีทราบค่า ρ โดยการทำให้อยู่ในรูปแบบที่เรียกว่า difference form นั่นคือทำให้อยู่ในรูป

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \quad (2)$$

$$Y_t^* = \beta_1 + \beta_2^* X_2^* + \varepsilon_t \quad (3)$$

โดยที่ $\varepsilon_t = (u_t - \rho u_{t-1})$ และเราก็จะประมาณสมการที่ (3) ด้วย OLS ซึ่งจะเรียว่าการใช้ GLS การแปลงโมเดลในรูป difference form ทำให้ observation หายไปหนึ่งตัว เพื่อหลีกเลี่ยงการสูญเสีย observation อาจจะใช้รูปแบบการแปลงโมเดลในรูปแบบ $Y_1\sqrt{1-\rho^2}$ และ $X_1\sqrt{1-\rho^2}$ ซึ่งเรียกว่า Prais-Winsten transformation

กรณีที่เราไม่ทราบค่า ρ ซึ่งส่วนมากจะเป็นเช่นนั้น วิธีการที่สามารถทำได้ก็คือ การใช้ First-Difference Method โดยการสมมติให้ $\rho = 1$ ดังนั้นในสมการ (2) ก็จะลดรูปลงเหลือเป็นสมการที่ (4)

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

$$Y_t - Y_{t-1} = +\beta_2(X_t - X_{t-1}) + \varepsilon_t \text{ หรือ } \Delta Y_t = +\beta_2 \Delta X_t + \varepsilon_t \quad (4)$$

ซึ่งวิธีการ First-Difference นั้นจะเหมาะสมในกรณีโมเดลเริ่มต้นนั้นมีค่า $d<R^2$(Maddala, 2001) จากตารางที่ 8 พบว่า $d<R^2$ เมื่อประมาณสมการที่ (4) จะปรากฏผลดังนี้

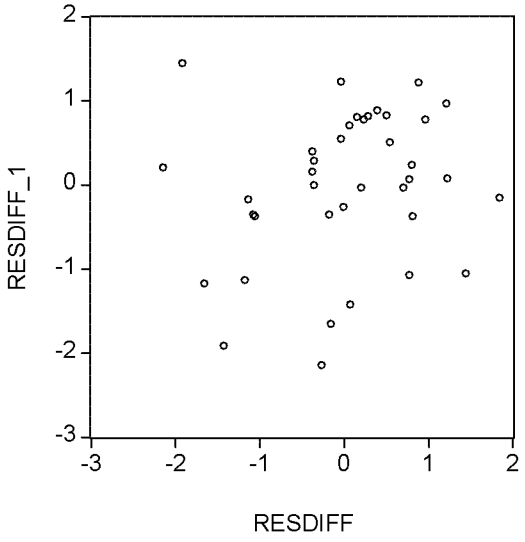
ตารางที่ 11 ผลการประมาณสมการ First-Difference

Dependent Variable: DIFFY
Method: Least Squares
Date: 02/20/07 Time: 15:34
Sample(adjusted): 1960 1998
Included observations: 39 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DIFFX	0.719956	0.078194	9.207333	0.0000

R-squared	0.361092	Mean dependent var	1.194872
Adjusted R-squared	0.361092	S.D. dependent var	1.173378
S.E. of regression	0.937901	Akaike info criterion	2.734962
Sum squared resid	33.42701	Schwarz criterion	2.777617
Log likelihood	-52.33175	Durbin-Watson stat	1.509651

จากตารางที่ 11 พบว่า ค่า d เท่ากับ 1.509 แสดงว่าเกิด Autocorrelation เล็กน้อยในสมการ First-Difference แต่ถ้าแสดงกราฟของ residual ในเวลา t กับ t-1 ของสมการ First-Difference ก็ไม่เห็นว่าจะเกิดปัญหา Autocorrelation มากนัก ดังภาพที่ 5



ภาพที่ 5 Residual ในเวลา t กับ t-1 ของสมการ First-Difference

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

หากจะทดสอบด้วย BG Test ก็ทำได้สมมติให้ lag เท่ากับ 4 โดยนำค่า residual ของสมการ First-Difference มาประมาณสมการกับ ค่า ΔX รวมทั้ง $\hat{u}_{t-1}$, $\hat{u}_{t-2}, \hat{u}_{t-3}, \hat{u}_{t-4}$ ดังตารางที่ 12 แล้วนำ R^2 มาใช้พบว่า $(n-4)R^2 \sim \chi^2_4$ เท่ากับ $(35-4)(0.090343)=2.80 \sim \chi^2_4$ เมื่อเปิดตารางไควสแควร์ที่ df=4 ณ ระดับนัยสำคัญ ร้อยละ 1 มีค่าเท่ากับ 13.28 ซึ่งค่า 2.80 ไม่ตกอยู่ในขอบเขตวิกฤตินั้นคือยอมรับสมมติฐานหลัก นั่นคือม ρ ทุกตัวมีค่าเท่ากับศูนย์ และจากตารางที่ 12 ก็จะเห็นได้อีกว่าค่าสัมประสิทธิ์ตัวแปร Rsidual นั้นไม่มีนัยสำคัญทางสถิติ

ตารางที่ 12 BG Test

Dependent Variable: RESDIFF
Method: Least Squares
Date: 02/20/07 Time: 16:03
Sample(adjusted): 1964 1998
Included observations: 35 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.149394	0.325769	0.458589	0.6499
DIFFX	-0.108856	0.176510	-0.616714	0.5422
RESDIFF_1	0.208565	0.192361	1.084236	0.2872
RESDIFF_2	-0.078282	0.201146	-0.389180	0.7000
RESDIFF_3	-0.173395	0.207692	-0.834868	0.4106
RESDIFF_4	0.248263	0.210671	1.178440	0.2482
R-squared	0.090343	Mean dependent var		-0.023014
Adjusted R-squared	-0.066494	S.D. dependent var		0.972832
S.E. of regression	1.004656	Akaike info criterion		3.001972
Sum squared resid	29.27066	Schwarz criterion		3.268603
Log likelihood	-46.53451	F-statistic		0.576031
Durbin-Watson stat	1.770732	Prob(F-statistic)		0.717842

ที่ผ่านมาพอสรุปได้ว่าสมการที่เราประมาณจากตารางที่ 11 นั้นมีความเหมาะสม อย่างไรก็ตามพึงระลึกว่า First-Difference Transformation นั้นมีความเหมาะสมกับกรณี ρ มีค่าสูงและ d มีค่าต่ำหรือใช้ได้ดีและถูกต้องในกรณี $\rho = 1$ เราจะทดสอบต่อไปว่า จริงๆ แล้ว ที่เรากำหนด $\rho = 1$ นั้นถูกต้องหรือไม่ วิธีการทดสอบเรียกว่า g statistic หรือ Berenblutt-Webb test โดยใช้สูตร $g = \dfrac{\sum_{2}^{n} \hat{e}_t^2}{\sum_{1}^{n} \hat{u}_t^2}$ โดยที่ $\hat{u}_t$ มาจาก OLS ของสมการดั้งเดิม และ $\hat{e}_t$ มาจาก OLS ในสมการ First-Difference (โดยไม่มี Intercept term) เมื่อได้ค่า g แล้วก็นำมาพิจารณากับตารางของ Durbin-Watson โดยใช้ n=39 และ k=1 โดยการทดสอบมีสมมติฐานหลักว่า $\rho = 1$ และสมมติฐานทางเลือก คือ $\rho = 0$ ผลการประมาณหาค่า residual

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ทั้งสองสมการ จากตารางที่ 8 และ 11 พบว่า ค่า $g = \frac{33.4270}{272.0220} = 0.0012$ จากตาราง Durbin-Watson นั้นกำหนด n=39 และ k=1 ที่ระดับนัยสำคัญร้อยละ 5 จะได้ $d_L = 1.435$ $d_U = 1.540$ แล้วพิจารณาจากตารางที่ 13 ซึ่งพบว่าค่า 0.0012 นั้นมีค่าน้อยกว่า d_L นั่นคือ $\rho = 1$

ตารางที่ 13 ขอบเขตการยอมรับและการปฏิเสธสมมติฐานหลัก กรณีการใช้ Durbin-Watson Test

Regions of Acceptance and Rejection of the Null Hypothesis				
Zero to d_l	d_l to d_u	d_u to (4 - d_u)	(4-d_u) to (4-d_l)	(4-d_l) to 4
Reject Null H_o: **Positive** Autocorrelation	Neither accept or reject	Accept the Null Hypothesis	Neither accept or reject	Reject Null H_o: **Negative** Autocorrelation

ที่มา: http://www.csus.edu/indiv/j/jensena/mgmt105/durbin.htm (20 ก.พ. 2550)

จากทั้งหมดนั้นเราก็พอจะสรุปได้ว่าสมการในรูป Difference Form นั้นเหมาะสมแล้ว นอกจากการสมมติให้ $\rho = 1$ แล้วยังมีวิธีการอื่นๆ อีกกรณีที่ ρ นั้นไม่เข้าใกล้หนึ่งมากพอ ซึ่งจะได้กล่าวดังนี้

การหา ρ จากค่าสถิติ Durbin-Watson แล้วนำค่านี้ไปใช้ในการแปลงโมเดลในรูป Difference Form โดยสามารถหาค่า $\hat{\rho} \approx 1 - \frac{d}{2}$ แต่ต้องใช้ในกรณีที่มีตัวอย่างขนาดใหญ่เท่านั้นถ้ามีน้อยเทอมดังกล่าวจะไม่จริงเสมอไป ซึ่งจากตารางที่ 8 เราจะได้ $\hat{\rho} \approx 0.9386$ เราก็นำค่านี้ไปใช้กับสมการที่ (2) และ (3) (ต้องมี Intercept term ด้วย) ผลปรากฏดังแสดงในตารางที่ 14

ตารางที่ 14 ผลการประมาณสมการ โดยใช้ ρ จากค่าสถิติ Durbin-Watson

Dependent Variable: Y_DIFF_DURBIN
Method: Least Squares
Date: 02/20/07 Time: 17:48
Sample(adjusted): 1960 1998
Included observations: 39 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.154433	0.629869	5.008080	0.0000
X_DIFF_DURBIN	0.510552	0.095938	5.321687	0.0000
R-squared	0.433561	Mean dependent var		6.422846
Adjusted R-squared	0.418252	S.D. dependent var		1.144289
S.E. of regression	0.872777	Akaike info criterion		2.615646
Sum squared resid	28.18434	Schwarz criterion		2.700957
Log likelihood	-49.00510	F-statistic		28.32035
Durbin-Watson stat	1.601499	Prob(F-statistic)		0.000005

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

การหาค่า ρ จากค่า Residuals ถ้าเราเชื่อว่าเป็นกรณี AR(1) นั่นคือ $u_t = \rho u_{t-1} + \varepsilon_t$ เราก็สามารถหา ρ จาก $\hat{u}_t = \rho \hat{u}_{t-1} + v_t$ โดยใช้สมการดั้งเดิมในการหา $\hat{u}_t$ ผลที่ได้ดังแสดงในตารางที่ 15 ดังนั้น $\hat{\rho} = 0.9142$ จากนั้นนำไปใช้ในสมการที่ (2) และ (3)

ตารางที่ 15 ผลการวิเคราะห์หา ρ โดยเชื่อว่าเป็นกรณี AR(1)

Dependent Variable: RES1
Method: Least Squares
Date: 02/20/07 Time: 18:06
Sample(adjusted): 1960 1998
Included observations: 39 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RES1_1	0.914245	0.056337	16.22811	0.0000

R-squared	0.873615	Mean dependent var	0.120615
Adjusted R-squared	0.873615	S.D. dependent var	2.561492
S.E. of regression	0.910629	Akaike info criterion	2.675945
Sum squared resid	31.51134	Schwarz criterion	2.718600
Log likelihood	-51.18093	Durbin-Watson stat	1.472987

การหาค่า ρ โดย Cochran-Orcutt Iterative Method ซึ่งใช้ได้กรณีที่ไม่ใช่เพียง AR (1) แต่ใช้กับกรณี higher-order autoregressive ด้วย ซึ่งวิธีการนี้จะทำให้ได้ค่า ρ หลายค่าแต่เราจะเลือกค่าที่เป็นค่าสุดท้ายหลังจากการทำ Iterative กณีสมมติ AR (1) โดยวิธี Cochran-Orcutt Iterative จะได้ค่าต่างๆ ดังนี้ (ค่าเหล่านี้ได้มาจากการใช้ Cochran-Orcutt Iterative Method ในโปรแกรม STATA 9.2)

Iteration 0:	rho = 0.0000	Iteration 1:	rho = 0.9142
Iteration 2:	rho = 0.9053	Iteration 3:	rho = 0.8992
Iteration 4:	rho = 0.8956	Iteration 5:	rho = 0.8936
Iteration 6:	rho = 0.8925	Iteration 7:	rho = 0.8919
Iteration 8:	rho = 0.8916	Iteration 9:	rho = 0.8915
Iteration 10:	rho = 0.8914	Iteration 11:	rho = 0.8914
Iteration 12:	rho = 0.8914	Iteration 13:	rho = 0.8914
Iteration 14:	rho = 0.8913	Iteration 15:	rho = 0.8913
Iteration 16:	rho = 0.8913	Iteration 17:	rho = 0.8913

จากการ Iterative ข้างต้นจะได้ค่าที่จำไปใช้ในการ Transformation คือ 0.8913

อย่างไรก็ตามสำหรับใน E-View กรณี Cochran-Orcutt Iterative Method เราจะใส่ AR(1) เข้าไปในโมเดลซึ่งผลปรากฏดังตารางที่ 16

ตารางที่ 16 ผลการประมาณสมการโดยใช้ AR (1) ใน E-Views

Dependent Variable: Y
Method: Least Squares
Date: 02/20/07 Time: 20:46
Sample(adjusted): 1960 1998
Included observations: 39 after adjusting endpoints
Convergence achieved after 7 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	45.03925	8.081898	5.572855	0.0000
X	0.550889	0.079019	6.971646	0.0000
AR(1)	0.891345	0.042583	20.93178	0.0000
R-squared	0.995339	Mean dependent var		86.34103
Adjusted R-squared	0.995080	S.D. dependent var		12.34486
S.E. of regression	0.865881	Akaike info criterion		2.623665
Sum squared resid	26.99100	Schwarz criterion		2.751631
Log likelihood	-48.16147	F-statistic		3843.976
Durbin-Watson stat	1.603995	Prob(F-statistic)		0.000000

หากเปรียบเทียบผลกับการวิเคราะห์ใน STATA 9.2 ปรากฏผลดังข้างล่าง

```
Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.9142
Iteration 2:  rho = 0.9053
Iteration 3:  rho = 0.8992
Iteration 4:  rho = 0.8956
Iteration 5:  rho = 0.8936
Iteration 6:  rho = 0.8925
Iteration 7:  rho = 0.8919
Iteration 8:  rho = 0.8916
Iteration 9:  rho = 0.8915
Iteration 10:  rho = 0.8914
Iteration 11:  rho = 0.8914
Iteration 12:  rho = 0.8914
Iteration 13:  rho = 0.8914
Iteration 14:  rho = 0.8913
Iteration 15:  rho = 0.8913
Iteration 16:  rho = 0.8913
Iteration 17:  rho = 0.8913

Cochrane-Orcutt AR(1) regression -- iterated estimates

      Source |       SS       df       MS              Number of obs =      39
-------------+------------------------------           F(  1,    37) =   71.85
       Model |   52.414941     1   52.414941           Prob > F      =  0.0000
    Residual |  26.9910037    37  .729486586           R-squared     =  0.6601
-------------+------------------------------           Adj R-squared =  0.6509
       Total |  79.4059446    38  2.08963012           Root MSE      =   .8541

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .5508882   .0649897     8.48   0.000     .4192066    .6825699
       _cons |   45.03935   6.158845     7.31   0.000     32.56035    57.51836
-------------+----------------------------------------------------------------
         rho |   .8913456
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.122905
Durbin-Watson statistic (transformed) 1.603998
```

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

หากพิจารณากรณีการใช้วิธีของ Prais-Winsten ซึ่งเป็นกรณีที่การทำ first observation ไม่หายไป ซึ่งผลปรากฏใน STATA 9.2 ดังนี้

```
Iteration 0:  rho = 0.0000
Iteration 1:  rho = 0.9142
Iteration 2:  rho = 0.9463
Iteration 3:  rho = 0.9556
Iteration 4:  rho = 0.9591
Iteration 5:  rho = 0.9605
Iteration 6:  rho = 0.9611
Iteration 7:  rho = 0.9613
Iteration 8:  rho = 0.9614
Iteration 9:  rho = 0.9615
Iteration 10:  rho = 0.9615
Iteration 11:  rho = 0.9615
Iteration 12:  rho = 0.9615
Iteration 13:  rho = 0.9615
Iteration 14:  rho = 0.9615

Prais-Winsten AR(1) regression -- iterated estimates

      Source |       SS       df       MS              Number of obs =      40
-------------+------------------------------           F(  1,    38) =  164.73
       Model |  143.420425     1  143.420425            Prob > F      =  0.0000
    Residual |  33.0833322    38  .870614006           R-squared     =  0.8126
-------------+------------------------------           Adj R-squared =  0.8076
       Total |  176.503757    39  4.52573736           Root MSE      =  .93307

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |    .724553   .0614294    11.79   0.000     .6001957    .8489103
       _cons |   26.44449   5.471857     4.83   0.000      15.3673    37.52169
-------------+----------------------------------------------------------------
         rho |   .9615163
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    0.122905
Durbin-Watson statistic (transformed) 1.528538
```

จากที่กล่าวมาทั้งหมดเป็นเรื่องราวเกี่ยวกับการแปลงโมเดลเพื่อแก้ไขปัญหา Autocorrelation เสียส่วนใหญ่ ในอีกด้านหนึ่งก็มีวิธีการอื่นที่ใช้เมื่อเกิดปัญหา Autocorrelation ก็คือ การทำให้ Standard Error กรณีที่ใช้ OLS ให้มีความถูกต้อง ซึ่งวิธีหนึ่งที่ใช้ คือ Newey-West Method ซึ่งเหมาะกับกรณีมีตัวอย่างขนาดใหญ่เท่านั้น จากสมการดั้งเดิมผลที่ได้จากการใช้ Newey-West Method ใน E-Views ดังแสดงในตารางที่ 17 ซึ่งเทียบผลกับตารางที่ 8 พบว่าค่าสัมประสิทธิ์ต่างๆเหมือนกัน R^2 ก็เหมือนกัน ต่างกันที่ค่า Standard Error กรณีใช้ Newey-West Method จะที่มีค่าที่สูงกว่า ดังนั้นทำให้ค่า t น้อยกว่ากรณีใช้ OLS ส่วนผลของค่า d นั้นมีค่าเท่ากันทั้งสองกรณี ซึ่งไม่ต้องเป็นห่วงเนื่องจากการใช้ Newey-West Method นั้นได้ทำให้ค่า Standard Error ของวิธี OLS ถูกต้องแล้ว

(หมายเหตุ: .ใน Gujarati (2003) ได้ระบุไว้ว่า ถ้ามีข้อมูล 15-20 observations แสดงว่ากลุ่มตัวอย่างมีขนาดเล็ก แต่ถ้ามีข้อมูล 50 observations ก็พอจะถือได้ว่ามีตัวอย่างขนาดใหญ่)

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ตารางที่ 17 ผลการประมาณสมการโดยใช้ Newey-West Method

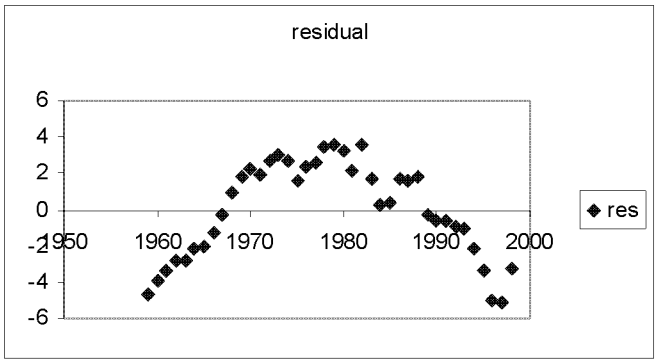
Dependent Variable: Y
Method: Least Squares
Date: 02/20/07 Time: 21:06
Sample: 1959 1998
Included observations: 40
Newey-West HAC Standard Errors & Covariance (lag truncation=3)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29.51925	4.118072	7.168223	0.0000
X	0.713659	0.051281	13.91661	0.0000
R-squared	0.958449	Mean dependent var		85.64500
Adjusted R-squared	0.957356	S.D. dependent var		12.95632
S.E. of regression	2.675533	Akaike info criterion		4.854881
Sum squared resid	272.0220	Schwarz criterion		4.939325
Log likelihood	-95.09761	F-statistic		876.5495
Durbin-Watson stat	0.122904	Prob(F-statistic)		0.000000

นอกจากนี้หากนำผลการใช้ Newey-West Method ใน STATA 9.2 มาเปรียบเทียบผลปรากฏดังนี้

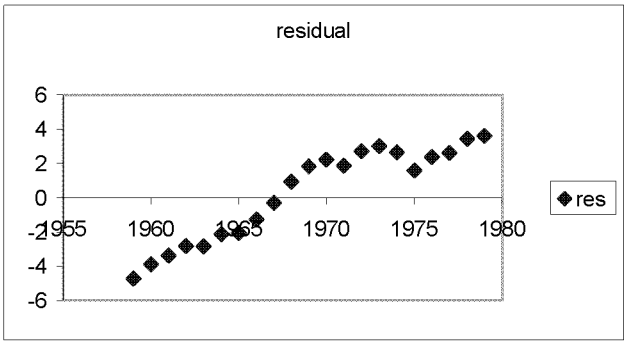
```
Regression with Newey-West standard errors                Number of obs  =        40
maximum lag: 0                                            F(  1,    38)  =    587.27
                                                          Prob > F       =    0.0000

------------------------------------------------------------------------------
             |             Newey-West
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |   .7136594   .0294492    24.23   0.000     .6540427    .7732761
       _cons |   29.51926   2.355621    12.53   0.000     24.75055    34.28796
------------------------------------------------------------------------------
```

ส่วนเพิ่มเติม เกี่ยวกับกราฟที่แสดงถึงปัญหาการมี Autocorrelation และ ไม่มี Autocorrelation

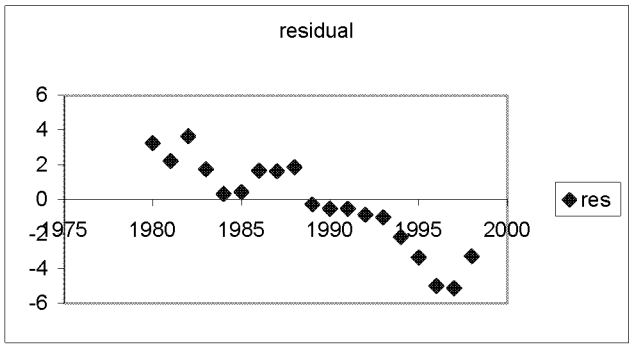


ภาพที่ 6 มี Autocorrelation

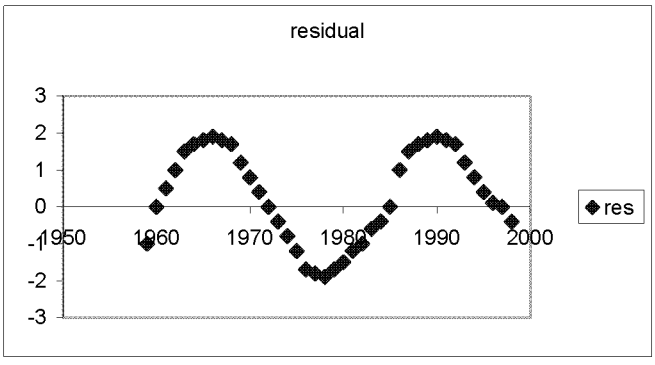
นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com



ภาพที่ 7 มี Autocorrelation

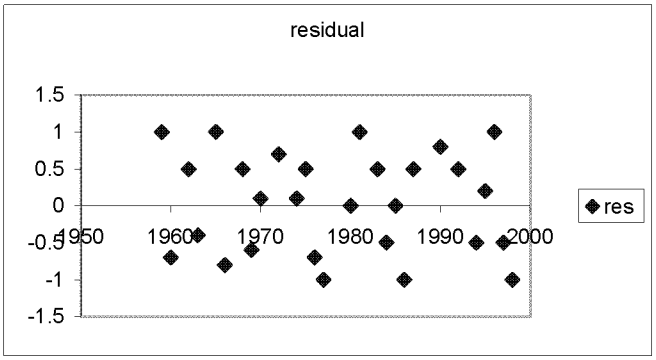


ภาพที่ 8 มี Autocorrelation



ภาพที่ 9 มี Autocorrelation

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com



ภาพที่ 10 ไม่มี Autocorrelation

ปัญหา HETEROSCEDASTICITY

ปัญหา Heteroscedasticity คือการที่ $E(u_i^2) \neq \sigma^2$ ลักษณะของการเกิดปัญหาดังกล่าวมีหลายเหตุผลที่ทำไมค่าความแปรปรวนของ u_i จึงไม่คงที่ ซึ่งจะได้กล่าวดังนี้

(1) ผลจากการเรียนรู้ข้อผิดพลาด ทำให้ค่าความแปรปรวนลดลง เช่น การประมาณสมการความผิดพลาดในการพิมพ์กับจำนวนชั่วโมงของการฝึกพิมพ์ดีด

(2) การเติบโตในรายได้ กลุ่มที่มีรายได้สูงทำให้มีทางเลือกในการออมมากขึ้น เช่น การประมาณสมการการออมกับรายได้ เมื่อรายได้เพิ่ม การออมก็เพิ่มแต่การออมนั้นมีการกระจายมากขึ้นหรือมีค่าความแปรปรวนลดลง

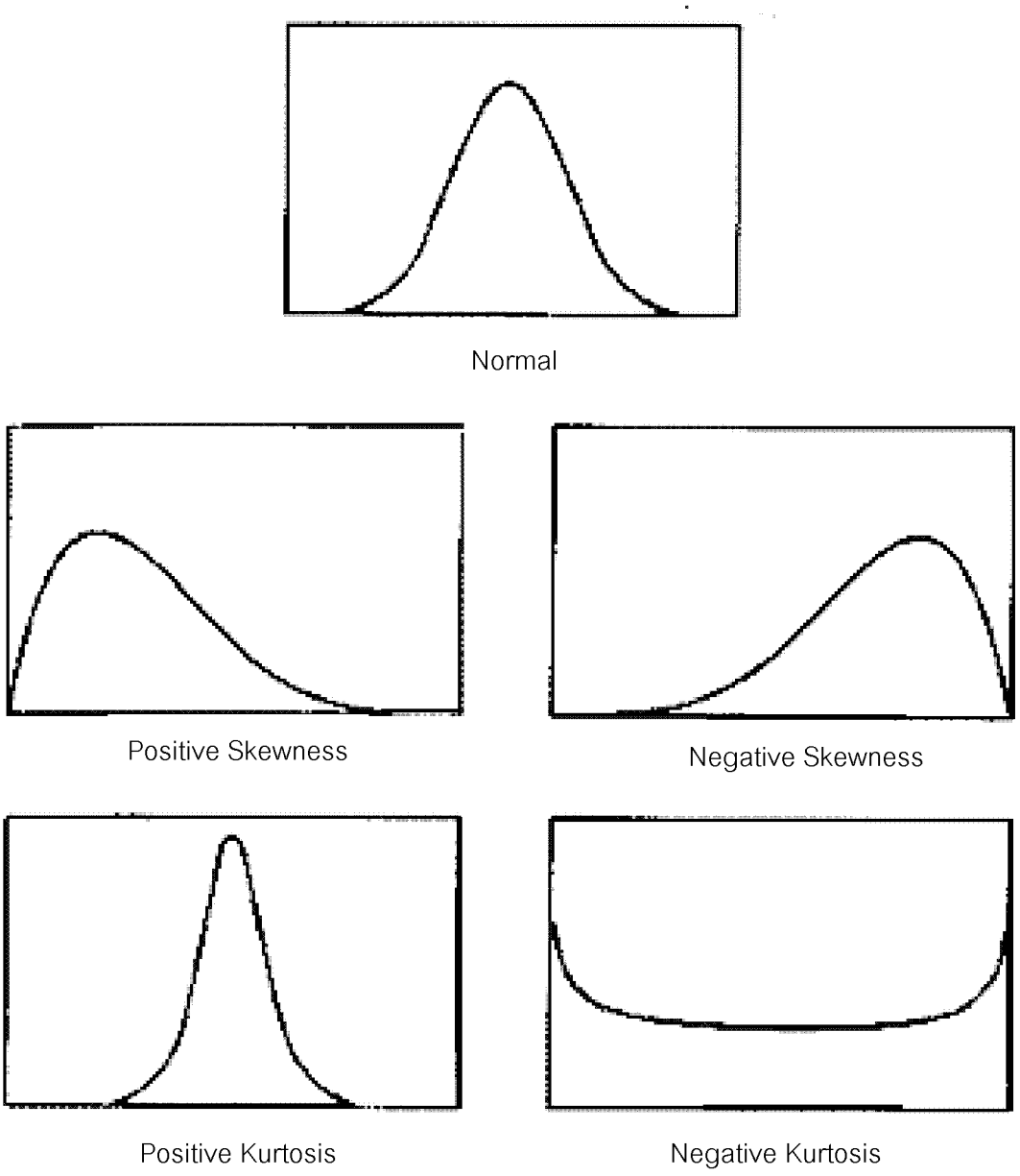
(3) เทคนิคการเก็บข้อมูล การมีเทคนิค วิธีการเก็บข้อมูลที่ดีมีคุณภาพช่วยให้ค่าความแปรปรวนลดลง

(4) ผลจาก Outlier กล่าวคือ ค่าที่ต่ำมากและค่าที่สูงมาก การนำค่าดังกล่าวเข้ามาหรือตัดออกไปมีผลต่อการประมาณสมการถดถอยอย่างมาก

(5) การละทิ้งตัวแปรสำคัญจะทำให้ค่าความแปรปรวนของการประมาณสมการไม่คงที่

(6) ความเบ้ (Skewness) ของข้อมูล เช่น การกระจายของรายได้และความมั่งคั่งของกลุ่มคนบนสุดมักจะมีจำนวนน้อย ความเบ้ ดังแสดงในภาพที่ 11

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com



ภาพที่ 11 Normality Skewness and Kurtosis

ที่มา: http://www.pfc.cfs.nrcan.gc.ca/profiles/wulder/mvstats/normality_e.html

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

(7) สาเหตุอื่นๆ เช่น การ Transformation Dada ที่ไม่ถูกต้อง การกำหนดรูปแบบฟังก์ชันไม่ถูกต้อง เป็นต้น นอกจากนี้ลักษณะที่เกี่ยวกับ ขนาดของตัวแปรที่อาจจะแบ่งเป็น ขนาดเล็ก ขนาดกลาง ขนาดใหญ่ เช่น กลุ่มผู้มีขนาดรายได้สูง กลุ่มผู้มีขนาดรายได้ปานกลาง กลุ่มผู้มีขนาดรายได้ต่ำ กรณีผู้ประกอบการหรือธุรกิจ อาจจะมีขนาดใหญ่ ขนาดกลาง ขนาดเล็ก หากลักษณะดังกล่าวเกิดขึ้นกับข้อมูลที่เก็บมาแบบรวมก็เป็นไปได้ว่าอาจจะเกิดปัญหา Heteroscedasticity

ผลของการเกิด HETEROSCEDASTICITY

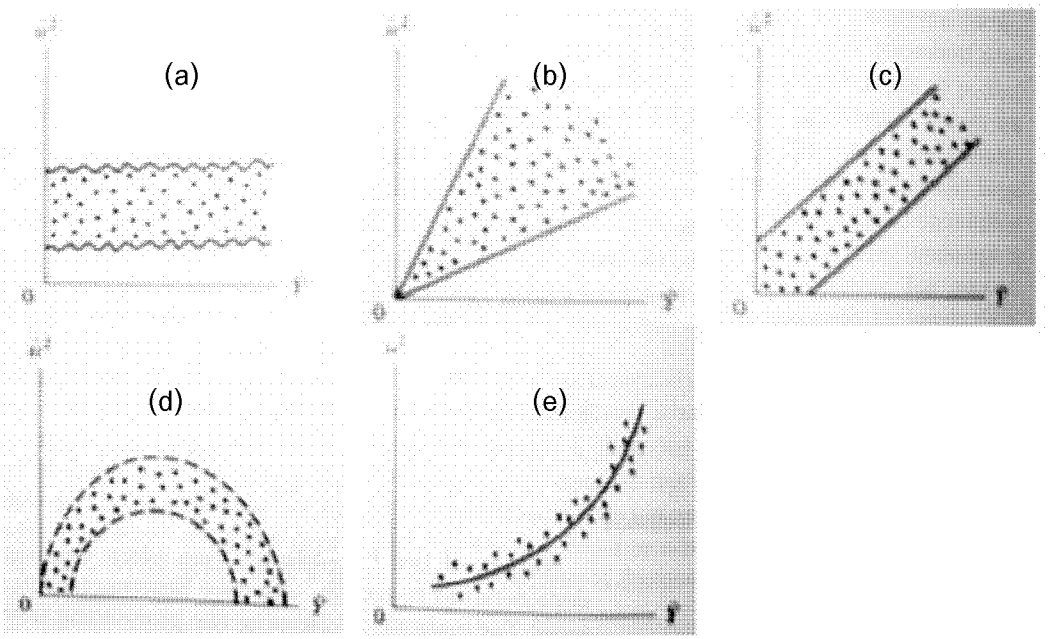
การใช้ OLS กรณีการเกิดปัญหา Heteroscedasticity นั้นตัวประมาณยังคง Unbiasedness และ Consistency แต่ตัวประมาณเหล่านั้นจะไม่เป็น Minimum Variance หรือตัวประมาณเหล่านั้นไม่มี efficiency นั่นเอง หรือกล่าวได้ว่าไม่เป็น BLUE (เหมือนกับกรณีปัญหา Autocorrelation) เนื่องจากค่าความแปรปรวนไม่ใช่ค่าต่ำสุด ดังนั้นการทดสอบสถิติ เช่น t-test และ F-test จะทำให้เกิดความเข้าใจผิดได้

การตรวจสอบการเกิด HETEROSCEDASTICITY

การค้นหาว่ามีปัญหา Heteroscedasticity หรือไม่นั้นกระทำได้ 2 วิธี คือ แบบไม่เป็นทางการและเป็นทางการ ซึ่งแบบไม่เป็นทางการ ได้แก่ พิจารณาจากลักษณะของข้อมูล การใช้วิธีแสดงกราฟ ส่วนแบบเป็นทางการได้แก่ Goldfeld-Quandt Test (เมื่อทราบว่าตัวแปรอิสระเพียง 1 ตัวที่มีความสัมพันธ์กับ σ_i^2 และ u_i มีการแจกแจงแบบปกติ), Breusch-Pagan-Godfrey (BPG)Test (ใช้ได้ในกรณีมีตัวแปรอิสระหลายตัวและไม่ทราบว่าปัญหา Heteroscedasticity เกิดจากตัวแปรอสระตัวใด รวมทั้ง u_i ต้องมีการแจกแจงแบบปกติ), White's General Heteroscedasticity Test (ใช้ได้เหมือนกับกรี BPG Test และไม่ต้องสมมติว่า u_i ต้องมีการแจกแจงแบบปกติ) โดยวิธีต่างๆ ที่กล่าวข้างต้นจะได้กล่าวในรายละเอียด ดังนี้

วิธีแสดงกราฟ ซึ่งเป็นการแสดงกราฟความสัมพันธ์ระหว่าง $\hat{Y}_i$ กับ $\hat{u}_i^2$ ซึ่งถ้าผลปรากฏรูปแบบดังภาพที่ 12 (b) (c) (d) และ (e) ก็แสดงว่าเกิดปัญหา Heteroscedasticity ส่วนภาพที่ 12 (a) นั้นไม่เกิดปัญหา Heteroscedasticity

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com



ภาพที่ 12 กราฟแสดงความสัมพันธ์ ระหว่าง $\hat{Y}_i$ กับ $\hat{u}_i^2$

ที่มา: Gujarati (2003: 402)

สำหรับวิธีที่เป็นทางการจะนำเสนอเพียงวิธี Breusch-Pagan-Godfrey (BPG)Test และ White's General Heteroscedasticity Test

Breusch-Pagan-Godfrey (BPG)Test

สมมติว่ารูปแบบฟังก์ชันที่จะประมาณ คือ

$$Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + u_i \qquad (5)$$

และสมมติว่า error variance σ_i^2 นั้นมีลักษณะในรูปแบบฟังก์ชันในสมการที่ (6) โดยที่ Z คือ nonstochastic variables หรือ X บางตัวหรือทั้งหมดสามารถใช้เป็น Z ได้ หลักการวิธีการนี้ คือ การทดสอบสัมประสิทธิ์ของสมการที่ (6) ว่า $\alpha_2 = \alpha_3 = \ldots = \alpha_m = 0$ เท่ากับศูนย์หรือไม่ ถ้าเท่ากับศูนย์แสดงว่าก็จะเหลือแค่ α_1 ซึ่งเป็นค่าคงที่ แสดงว่า σ_i^2 เป็น homoscedastic

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \ldots + \alpha_m Z_{mi} \qquad (6)$$

วิธีการเพื่อทดสอบมีดังนี้

ขั้นตอนที่ 1 ประมาณสมการที่ (5) เพื่อให้ได้ residual ของแต่ละ observation

ขั้นตอนที่ 2 หา $\tilde{\sigma}^2 = \dfrac{\sum \hat{u}_i^2}{n}$

ขั้นตอนที่ 3 สร้างตัวแปร p_i ในแต่ละ observation จากเทอม $p_i = \dfrac{\hat{u}_i^2}{\tilde{\sigma}^2}$

ขั้นตอนที่ 4 ประมาณสมการ $p_i = \alpha_1 + \alpha_2 Z_{2i} + \ldots + \alpha_m Z_{mi} + v_i$ แล้วหาค่า ESS

(ซึ่งหาจาก $ESS = \sum (\hat{P}_i - \overline{P})^2$)

ขั้นตอนที่ 5 จากขั้นตอนที่ 4 เราจะได้ค่า ESS แล้วนำมาหาค่า $\Theta = \dfrac{1}{2} ESS \sim \chi^2_{m-1}$ ซึ่งต้องมีตัวอย่างขนาดใหญ่ เมื่อได้ค่า Θ แล้วให้นำมาเทียบกับค่าในตารางไคสแควร์ ซึ่ง df=m-1 ถ้าค่า Θ เกินช่วงวิกฤติแสดงว่าปฏิเสธสมมติหลัก นั่นคือ มีปัญหา Heteroscedasticity

ตัวอย่างการใช้วิธี Breusch-Pagan-Godfrey โดยใช้ข้อมูลดังแสดงในตารางที่ 18 โดยดาวน์โหลดข้อมูลได้ที่ http://www.kaset51.com (คลิ๊กหัวข้อเอกสารประกอบการสอน) ในหัวข้อ Breusch-Pagan-Godfrey Test

ตารางที่ 18 Data on Consumption Expenditure and Income

OBS	Y	X	RY	RX	OBS	Y	X	RY	RX
1	55	80	55	80	16	115	180	115	180
2	65	100	70	85	17	140	225	130	185
3	70	85	75	90	18	120	200	135	190
4	80	110	65	100	19	145	240	120	200
5	79	120	74	105	20	130	185	140	205
6	84	115	80	110	21	152	220	144	210
7	98	130	84	115	22	144	210	152	220
8	95	140	79	120	23	175	245	140	225
9	90	125	90	125	24	180	260	137	230
10	75	90	98	130	25	135	190	145	240
11	74	105	95	140	26	140	205	175	245
12	110	160	108	145	27	178	265	189	250
13	113	150	113	150	28	191	270	180	260
14	125	165	110	160	29	137	230	178	265
15	108	145	125	165	30	189	250	191	270

Y = Consumption Expenditure, $

X = Income, $

RY = Consumption Expenditure Ranked by X Values

RX = Ranked Income

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

จากสมการที่ (5) ได้ผลการวิเคราะห์ดังแสดงในตารางที่ 19 ซึ่งในขั้นตอนที่ 1 นี้เราจะได้ residual ของแต่ละ observation พร้อมทั้งได้ในขั้นตอนที่ 2 ด้วย นั่นคือ ได้ $\sum \hat{u}_i^2$ หรือ Sum squared resid ใน E-Views ซึ่งเท่ากับ 2,361.153 และได้ $\hat{\sigma}^2 = (\sum u_i^2)/n = 2,361.153/30 = 78.7051$

ตารางที่ 19 ผลการวิเคราะห์ BPG Test ขั้นตอนที่ 1

Dependent Variable: Y
Method: Least Squares
Date: 02/21/07 Time: 13:57
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.290307	5.231386	1.775879	0.0866
X	0.637785	0.028617	22.28718	0.0000

R-squared	0.946638	Mean dependent var	119.7333
Adjusted R-squared	0.944732	S.D. dependent var	39.06134
S.E. of regression	9.182968	Akaike info criterion	7.336918
Sum squared resid	2361.153	Schwarz criterion	7.430332
Log likelihood	-108.0538	F-statistic	496.7183
Durbin-Watson stat	1.702261	Prob(F-statistic)	0.000000

ในขั้นตอนที่ 3 นำค่า 78.7051 ไปหารกับ $\hat{u}_i$ ก็จะได้ p_i ของแต่ละ observation

ในขั้นตอนที่ 4 ประมาณสมการ $p_i = \alpha_1 + \alpha_2 Z_{2i} + ... + \alpha_m Z_{mi} + v_i$ จะได้ผลตามตารางที่ 20 และได้ ESS ของการประมาณสมการนี้เท่ากับ 10.4280 ซึ่งค่านี้ใน E-Views หาได้โดยการใช้คำสั่ง Forecast ซึ่งจะได้ค่า $\hat{p}_i$ ซึ่งอาจจะตั้งชื่อตัวแปรเป็น pif จากนั้นนำค่านี้ใช้คำสั่ง Generate Series โดยสร้างตัวแปรใหม่ เช่น pif2=(pif-1)^2 เมื่อหาผลรวมของ pif2 ก็จะได้ค่า ESS วิธีการหาผลรวมของตัวแปร pif2 ใช้คำสั่ง

Quick Group Statistics Descriptive Statistics Comon sample

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ตารางที่ 20 ผลการวิเคราะห์ BPG Test ขั้นตอนที่ 4

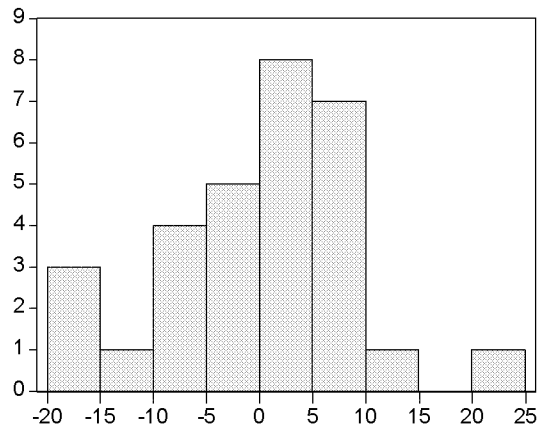
Dependent Variable: PI
Method: Least Squares
Date: 02/21/07 Time: 15:58
Sample: 1 30
Included observations: 30

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.742614	0.752927	-0.986302	0.3324
X	0.010063	0.004119	2.443327	0.0211

R-squared	0.175740	Mean dependent var	1.000000
Adjusted R-squared	0.146302	S.D. dependent var	1.430432
S.E. of regression	1.321659	Akaike info criterion	3.459993
Sum squared resid	48.90992	Schwarz criterion	3.553406
Log likelihood	-49.89990	F-statistic	5.969846
Durbin-Watson stat	0.778824	Prob(F-statistic)	0.021114

ขั้นตอนที่ 5 เมื่อเปิดค่าในตารางไคสแควร์ ซึ่ง df=m-1 =2-1=1 ที่ระดับนัยสำคัญร้อยละ 1 หรือใช้คำสั่งใน E-Views โดยพิมพ์ **=@qchisq(.99,1)** ในช่องสีขาวยาว ในหน้าจอเมนูหลัก (ที่มีเมนู File) ซึ่งมีค่าเท่ากับ 6.6349 ซึ่งแสดงอยู่มุมด้านล่างซ้ายของหน้าจอหลัก ดังนั้นจะเห็นว่าค่าสถิติที่คำนวณได้ นั่นคือ $\Theta = \frac{1}{2}ESS = (10.4280)(0.5) = 5.2140$ มีค่าไม่ตกในเขตวิกฤติทำให้ยอมรับสมมติฐานหลัก ดังนั้นแสดงว่าไม่มีปัญหา Heteroscedasticity

ข้อสังเกต เนื่องจากตัวอย่างมีจำนวนน้อย การทดสอบจะอ่อนไหวต่อการแจกแจงของ error term ซึ่งวิธีนี้ต้องใช้ตัวอย่างขนาดใหญ่ จึงต้องพิจารณาด้วยว่าค่า error term มีการแจกแจงปกติหรือไม่ โดยใช้คำสั่ง View Rsidual Test Histogram-Normality Test ซึ่งปรากฏดังภาพที่ 13 ซึ่ง Residual ของสมการที่ (5) นั้นมีการแจกแจงแบบปกติ



Series: Residuals
Sample 1 30
Observations 30

Mean	-1.86E-14
Median	0.880779
Maximum	20.26355
Minimum	-18.98076
Std. Dev.	9.023252
Skewness	-0.359065
Kurtosis	2.977931
Jarque-Bera	0.645247
Probability	0.724246

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ภาพที่ 13 การทดสอบการแจกแจงของค่า Residual ของสมการถดถอย

White's General Heteroscedasticity Test

วิธีนี้ไม่จำเป็นต้องมีข้อสมมติเรื่อง error term ที่ต้องมีการแจกแจงปกติ และสามารถได้กับตัวแปรอิสระหลายตัว โดยสมมติเบื้องต้นว่าสมการถดถอยที่ใช้ คือ สมการที่ (7) และมีขั้นตอนดังนี้

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \tag{7}$$

ขั้นตอนที่ 1 ประมาณสมการที่ (7) เพื่อให้ได้ $\hat{u}_i$ และ $\hat{u}_i^2$

ขั้นตอนที่ 2 ประมาณสมการ Auxiliary Regression ดังสมการที่ (8) เพื่อจะได้ค่า R^2

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + u_i \tag{8}$$

ขั้นตอนที่ 3 คำนวณหาค่า $n \cdot R^2 \sim \chi^2_{\text{df}}$ โดยที่ df = จำนวนตัวแปรอิสระ

ขั้นตอนที่ 4 เปิดตารางไคสแควร์เพื่อนำมาเปรียบเทียบกับค่าสถิติที่คำนวณได้ ซึ่งถ้าหากค่าสถิติมีค่าตกอยู่ในเขตวิกฤติแสดงว่าปฏิเสธสมมติฐานหลัก หรือมีปัญหา Heteroscedasticity

ตัวอย่างการใช้วิธี White's General Heteroscedasticity โดยใช้ข้อมูลดังแสดงในตารางที่ 21 โดยดาวน์โหลดข้อมูลได้ที่ http://www.kaset51.com (คลิ๊กหัวข้อเอกสารประกอบการสอน) ในหัวข้อ White's General Heteroscedasticity Test

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ตารางที่ 21 **สัดส่วนภาษีจากการส่งออกและนำเข้าต่อรายได้ของรัฐบาล (Y) กับ สัดส่วนการส่งออกนำเข้าต่อ GDP (X1) และ GDP per Capita (X2)**

OBS	Y	X1	X2	OBS	Y	X1	X2
1	14.36	21.96	490	21	18.54	11.24	73
2	7.47	29.78	1316	22	22.42	30.97	726
3	5.76	40.84	670	23	3.71	21.71	306
4	6.21	55.83	1196	24	22.44	12.35	144
5	5.53	14.93	293	25	34.46	50.32	362
6	41.96	37.86	57	26	53.47	81.16	356
7	10.41	31.99	1947	27	26.29	21.61	262
8	55.09	62.92	129	28	10.07	66.82	836
9	16.66	23.11	379	29	6.9	52.12	1130
10	30.72	23.57	263	30	28.64	14.25	70
11	57.84	46.44	357	31	30.6	33.96	329
12	44.44	30.01	189	32	20.38	39.4	179
13	57.95	43.06	219	33	30.06	21.54	220
14	13.26	41.04	794	34	24.71	37.92	224
15	2.41	19.99	943	35	38.22	33.01	60
16	68.59	42.65	172	36	5.39	42.97	1380
17	2.78	26.34	340	37	28.8	44.86	1428
18	47.09	34.29	189	38	48.7	35.14	96
19	65.18	24.65	105	39	15.26	43.54	395
20	50.95	36.79	194	40	1.12	6.78	2577
				41	21.28	56.72	648

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

จากตารางที่ 21 นำข้อมูลไปประมาณสมการถดถอยในขั้นตอนที่ 1 จะได้ตารางที่ 22 โดยก่อนการประมาณจะต้องจัดข้อมูลตัวอย่างให้อยู่ในรูป squared term and cross-product term ได้แก่ $X_{1i}^2, X_{2i}^2, X_{1i}X_{2i}$ นั่นคือต้องใช้คำสั่ง Generate Series เพื่อให้ได้ตัวแปรใหม่ตามที่ต้องการ

ตารางที่ 22 ผลการวิเคราะห์ White's General Heteroscedasticity Test ในขั้นตอนที่ 1

Dependent Variable: Y
Method: Least Squares
Date: 02/21/07 Time: 18:56
Sample: 1 41
Included observations: 41

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	25.03375	6.431405	3.892423	0.0004
X1	0.344455	0.155559	2.214307	0.0329
X2	-0.019439	0.004439	-4.378898	0.0001

R-squared	0.384796	Mean dependent var	26.73463
Adjusted R-squared	0.352417	S.D. dependent var	19.45120
S.E. of regression	15.65288	Akaike info criterion	8.409542
Sum squared resid	9310.478	Schwarz criterion	8.534925
Log likelihood	-169.3956	F-statistic	11.88406
Durbin-Watson stat	1.971886	Prob(F-statistic)	0.000098

ในขั้นตอนที่ 2 นำ $\hat{u}_i^2$ มาประมาณสมการกับ $x_1, x_2, x^2_1, x^2_2, x_1x_2$ จะได้ตารางที่ 23 ซึ่งค่า R^2 = 0.070753

ตารางที่ 22 ผลการวิเคราะห์ White's General Heteroscedasticity Test ในขั้นตอนที่ 2

Dependent Variable: RES1SQ
Method: Least Squares
Date: 02/21/07 Time: 19:39
Sample: 1 41
Included observations: 41

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	50.33761	252.4941	0.199362	0.8431
X1	12.21823	11.83796	1.032123	0.3091
X2	-0.282371	0.521458	-0.541503	0.5916
X1SQ	-0.114353	0.130912	-0.873506	0.3883
X2SQ	0.000163	0.000176	0.929111	0.3592
X1X2	-0.001503	0.007027	-0.213941	0.8318

R-squared	0.070753	Mean dependent var	227.0848
Adjusted R-squared	-0.061996	S.D. dependent var	268.6660
S.E. of regression	276.8689	Akaike info criterion	14.21942
Sum squared resid	2682974.	Schwarz criterion	14.47019
Log likelihood	-285.4982	F-statistic	0.532983
Durbin-Watson stat	1.798823	Prob(F-statistic)	0.749791

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ในขั้นตอนที่ 3 หาค่า $n \cdot R^2 = 41 \cdot (0.070753) = 2.901$

ในขั้นตอนที่ 4 หาค่าไคสแควร์ ที่ df = 5 ที่ระดับนัยสำคัญร้อยละ 1 ซึ่งมีค่าเท่ากับ 15.0863 ดังนั้นยอมรับสมมติฐานหลัก นั่นคือ ไม่มีปัญหา Heteroscedasticity

การแก้ปัญหา HETEROSCEDASTICITY

การแก้ปัญหา Heteroscedasticity มีสองกรณีคือ กรณีที่หนึ่งเมื่อทราบค่า σ_i^2 และกรณีที่สองเมื่อไม่ทราบ σ_i^2

กรณีทราบค่า σ_i^2 สามารถใช้วิธี Weighted Least Squares (WLS) เพื่อทำให้ OLS เป็น BLUE โดยการสร้าง series ใหม่ที่จะใช้ weighted ข้อมูลซึ่งก็คือ σ_i โดยพิจารณาตัวอย่างต่อไปนี้ โดยใช้ข้อมูลในตารางที่ 23 ซึ่งสามารถดาวน์โหลดข้อมูลได้ที่ http://www.kaset51.com (คลิ๊กหัวข้อเอกสารประกอบการสอน) ในหัวข้อ WLS Table 11.4

ตารางที่ 23 Average Compensation, Employment Size and Standard Deviation of Compensation

Y	X	SIGMA	YSIGMA	XSIGMA
3396	1	743.7	4.5664	0.0013
3787	2	851.4	4.448	0.0023
4013	3	727.8	5.5139	0.0041
4104	4	805.06	5.0978	0.005
4146	5	929.9	4.4585	0.0054
4241	6	1080.6	3.9247	0.0055
4387	7	1243.2	3.5288	0.0056
4538	8	1307.7	3.4702	0.0061
4834	9	1112.5	4.3532	0.0081

Data on Average Compensation, Employment Size and Standard Deviation of Compensation

Y = Average Compensation

X = Employment Size

where:

1=1 to 4 Employees
2=5 to 9 Employees
3=10 to 19 Employees
4=20 to 49 Employees
5=50 to 99 Employees
6=100 to 249 Employees
7=250 to 499 Employees
8=500 to 999 Employees
9=1000 to 2499 Employees

SIGMA = Standard Deviation of Compensation, YSIGMA = Y/SIGMA , XSIGMA = X/SIGMA

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

จากข้อมูลในตารางที่ 23 เราจะประมาณสมการ Y กับ X ดังปรากฏผลในตารางที่ 24

ตารางที่ 24 ผลการวิเคราะห์โดยใช้ OLS

Dependent Variable: Y
Method: Least Squares
Date: 02/21/07 Time: 21:26
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3419.833	80.43523	42.51661	0.0000
X	148.1667	14.29372	10.36586	0.0000
R-squared	0.938838	Mean dependent var		4160.667
Adjusted R-squared	0.930101	S.D. dependent var		418.7798
S.E. of regression	110.7186	Akaike info criterion		12.44499
Sum squared resid	85810.33	Schwarz criterion		12.48882
Log likelihood	-54.00246	F-statistic		107.4511
Durbin-Watson stat	1.215000	Prob(F-statistic)		0.000017

จากตารางที่ 24 เราต้องการดูว่าเกิดปัญหา Hetroscedasticity หรือไม่ ในตัวอย่างนี้จะใช้ White's General Heteroscedasticity Test ซึ่งเราจะได้ค่า squared residual จากตารางที่ 24 แล้วจากนั้นก็จะนำมาประมาณสมการกับ X และ X^2 ได้ผลดังตารางที่ 25 จากนั้นนำ $R^2 = 0.583$ มาใช้หา nR^2 ซึ่งเท่ากับ 5.247 และค่าไคสแควร์ df=2 ณ ระดับนัยสำคัญร้อยละ 10 (หรือระดับความเชื่อมั่น ร้อยละ 90) เท่ากับ 4.605 ดังนั้นปฏิเสธสมมติฐานหลัก นั่นคือมีปัญหา Hetroscedasticity

ตารางที่ 25 White's General Heteroscedasticity Test ขั้นตอนที่ 2

Dependent Variable: RES1SQ
Method: Least Squares
Date: 02/21/07 Time: 21:33
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	33161.15	9179.040	3.612704	0.0112
X	-9087.276	4214.658	-2.156113	0.0745
XSQ	688.7276	411.0476	1.675542	0.1448
R-squared	0.582980	Mean dependent var		9534.481
Adjusted R-squared	0.443973	S.D. dependent var		9674.301
S.E. of regression	7213.855	Akaike info criterion		20.86660
Sum squared resid	3.12E+08	Schwarz criterion		20.93234
Log likelihood	-90.89968	F-statistic		4.193898
Durbin-Watson stat	3.137812	Prob(F-statistic)		0.072522

เมื่อทราบว่าเกิดปัญหา Heteroscedasticity และเราทราบ σ_i จึงสามารถใช้ WLS ดังผลปรากฏในตารางที่ 26 ซึ่งรูปแบบสมการที่ใช้ประมาณ คือ $Y_i/\sigma_i = \beta_1(1/\sigma_i) + \beta_2(X_i/\sigma_i) + U_i/\sigma_i$ โดยจะไม่มี Intercept term

ตารางที่ 26 ผลการวิเคราะห์โดยใช้ WLS

Dependent Variable: YSIGMA
Method: Least Squares
Date: 02/22/07 Time: 01:31
Sample: 1 9
Included observations: 9

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SIG_INVERSE	3411.280	79.31392	43.00986	0.0000
XSIGMA	153.4786	16.64418	9.221162	0.0000
R-squared	0.965669	Mean dependent var		4.373500
Adjusted R-squared	0.960764	S.D. dependent var		0.671417
S.E. of regression	0.132994	Akaike info criterion		-1.003889
Sum squared resid	0.123813	Schwarz criterion		-0.960062
Log likelihood	6.517502	Durbin-Watson stat		1.206883

กรณีไม่ทราบค่า σ_i^2 ซึ่งโดยส่วนมากมักจะเป็นเช่นนั้น วิธีที่นิยมใช้ คือการใช้ White Heteroskedasticity-Consistent Standard Errors & Covariance และต้องใช้กับตัวอย่างที่มีขนาดใหญ่ นอกจากข้อมูลที่มีขนาดใหญ่แล้ว การใช้ White Heteroskedasticity-Consistent Standard Errors & Covariance อาจจะไม่มีประสิทธิภาพเท่ากับการ Transform model ซึ่งการ Transform model มีหลายแบบ ได้แก่ การ weighted ด้วย $\frac{1}{X_i}$, $\frac{1}{\sqrt{X_i}}$, $\frac{1}{\hat{Y}_i}$, การทำให้อยู่ในรูป log-log, อย่างไรก็ตามการจะใช้วิธี Transform data ก็ต้องคำนึงถึงสิ่งต่อไปนี้

(1) เราอาจจะไม่ทราบว่าตัวแปรอิสระใดที่จะต้องนำมาใช้ในการ Transform model อย่างไรก็ตามเราสามารถทำได้โดยการแสดงกราฟระหว่าง Squared residual กับตัวแปรอิสระทุกตัวแล้วดูว่าเป็นความสัมพันธ์แบบใดหรือไม่มีความสัมพันธ์

(2) Log-Log Transform ไม่สามารถใช้ได้กับกรณีที่ค่าตัวแปรตามหรือตัวแปรอิสระเท่ากับศูนย์ แต่ก็สามารถเลี่ยงได้โดยการ บวกค่าคงที่ K ซึ่งมีค่ามากกว่าศูนย์เข้าไปในตัวแปรทุกตัว เพื่อทำให้ค่าของ Y หรือ X ที่เป็นศูนย์ให้มีค่ามากกว่าศูนย์

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

(3) การ transform model อาจจะทำให้เกิด Spurious Correlation เช่นกรณี $Y_i / X_i = \beta_1(1/X_i) + \beta_2$ ซึ่ง Y_i / X_i มักจะสัมพันธ์กับ $1/X_i$

(4) ในการ transform model จะต้องตีความผลด้วยความระมัดระวังในกรณีที่มีขนาดตัวอย่างไม่มาก

ตัวอย่างการ transform model กรณีเกิดปัญหา Heteroscedasticity พิจารณาจากข้อมูลในตารางที่ 27 ซึ่งสามารถดาวน์โหลดข้อมูลได้ที่ http://www.kaset51.com (คลิ๊กหัวข้อเอกสารประกอบการสอน) ในหัวข้อ Transform Model Table 11.5

ตารางที่ 27 R&D Expenditure in the USA, 1988 ($,millions)

OBS	SALES	RD	PROFITS
1	6375.3	62.5	185.1
2	11626.4	92.9	1569.5
3	14655.1	178.3	276.8
4	21869.2	258.4	2828.1
5	26408.3	494.7	225.9
6	32405.6	1083	3751.9
7	35107.7	1620.6	2884.1
8	40295.4	421.7	4645.7
9	70761.6	509.2	5036.4
10	80552.8	6620.1	13869.9
11	95294	3918.6	4487.8
12	101314.1	1595.3	10278.9
13	116141.3	6107.5	8787.3
14	122315.7	4454.1	16438.8
15	141649.9	3163.8	9761.4
16	175025.8	13210.7	19774.5
17	230614.5	1703.8	22626.6
18	293543	9528.2	18415.4

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

จากตารางที่ 27 เราจะประมาณสมการระหว่าง RD กับ SALES ซึ่งปรากฏผลดังตารางที่ 28

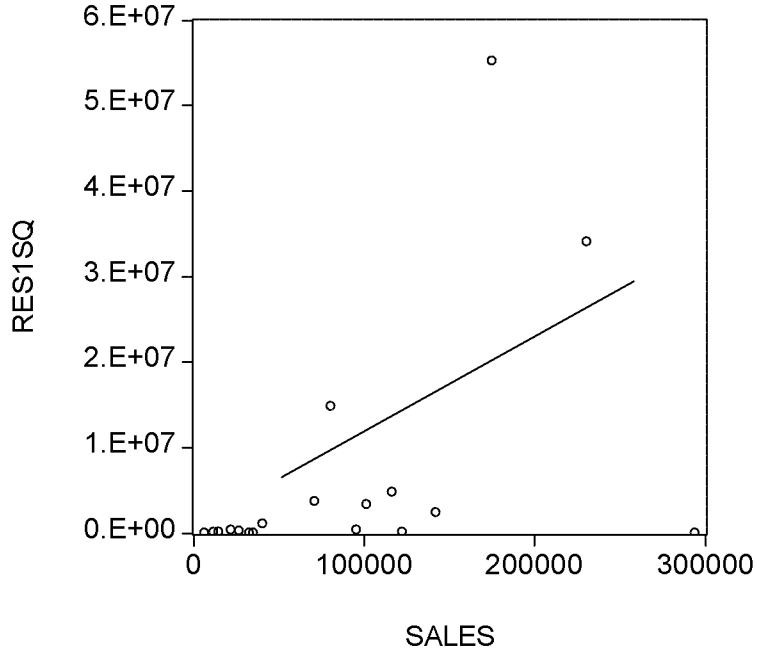
ตารางที่ 28 ผลการวิเคราะห์สมการถดถอยระหว่าง RD กับ SALES

Dependent Variable: RD
Method: Least Squares
Date: 02/21/07 Time: 22:59
Sample: 1 18
Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	192.9931	990.9858	0.194749	0.8480
SALES	0.031900	0.008329	3.830033	0.0015

R-squared	0.478303	Mean dependent var	3056.856
Adjusted R-squared	0.445697	S.D. dependent var	3705.973
S.E. of regression	2759.153	Akaike info criterion	18.78767
Sum squared resid	1.22E+08	Schwarz criterion	18.88660
Log likelihood	-167.0891	F-statistic	14.66916
Durbin-Watson stat	3.015607	Prob(F-statistic)	0.001476

จากผลการประมาณในตารางที่ 28 เราสามารถพิจารณาจากกราฟระหว่าง squared residual กับตัวแปร Sales ได้ ดังแสดงในภาพที่ 14 ซึ่งพบว่าสองตัวแปรมีรูปแบบของความสัมพันธ์



ภาพที่ 14 กราฟแสดงความสัมพันธ์ squared residual กับ Sales

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

ต่อมา จะทดสอบด้วยวิธีที่เป็นทางการเพื่อให้มั่นใจว่ามีปัญหา Heteroscedasticity จริงหรือไม่ โดยใช้ White's General Heteroscedasticity Test ดังนี้ จากตารางที่ 28 เราได้ square residual โดยจะนำมาประมาณสมการกับตัวแปร Sales และ Sales^2 เพื่อให้ได้ค่า R^2 ดังแสดงในตารางที่ 29

ตารางที่ 29 White's General Heteroscedasticity Test ขั้นตอนที่ 2

Dependent Variable: RES1SQ
Method: Least Squares
Date: 02/21/07 Time: 23:17
Sample: 1 18
Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6219665.	6459809.	-0.962825	0.3509
SALES	229.3508	126.2197	1.817077	0.0892
SALES_SQ	-0.000537	0.000449	-1.194952	0.2507

R-squared	0.289583	Mean dependent var	6767046.
Adjusted R-squared	0.194861	S.D. dependent var	14706011
S.E. of regression	13195639	Akaike info criterion	35.77968
Sum squared resid	2.61E+15	Schwarz criterion	35.92808
Log likelihood	-319.0171	F-statistic	3.057178
Durbin-Watson stat	1.694567	Prob(F-statistic)	0.076975

ต่อมานำค่า R^2 = 0.289583 มาหาค่า nR^2= (18)(0.289583) = 5.2125 และค่าไคสแควร์ df=2 ระดับนัยสำคัญร้อยละ 10 เท่ากับ 4.6052 ซึ่งตกอยู่ในเขตวิกฤติ ดังนั้นปฏิเสธสมมติฐานหลัก นั่นคือ มีปัญหา Heteroscedasticity เพื่อบรรเทาปัญหาดังกล่าวเราจะใช้การคาดเดาในลักษณะของ error variance เพราะเราไม่ทราบค่า error variance ที่แท้จริงได้และก็ไม่สามารถใช้ WLS ได้ ดังนั้นจึต้องใช้ Transform Model แทน จากภาพที่ 14 พอจะบอกได้ว่า error variance นั้น มีความสัมพันธ์เป็นสัดส่วนกับตัวแปร Sales นั่นคือ $E(u_i^2) = \sigma^2 X_i$ ดังนั้นจึง weighted ด้วย $\frac{1}{\sqrt{X_i}}$ เพราะว่า

$$E(v_i^2) = E(\frac{u_i}{\sqrt{X_i}})^2 = E(\frac{u_i^2}{X_i}) = \frac{1}{X_i}E(u_i^2)$$

$$= \frac{1}{X_i} \cdot \sigma^2 X_i = \sigma^2$$

ดังนั้นเราจะ weighted สมการด้วย $\frac{1}{\sqrt{X_i}}$ โดยที่การประมาณสมการจะต้องไม่มี Intercept term เนื่องจากตัวที่นำไป weighted ค่าคงที่ในสมการดั้งเดิมนั้นได้หายไปและเปรียบเสมือนเป็นตัวแปรอิสระอีกตัวหนึ่งที่เราต้องประมาณหาค่าสัมประสิทธิ์ ดังนั้นสรุปว่าสมการที่เราต้องประมาณคือ

นิติพงษ์ ส่งศรีโรจน์ econ555@gmail.com, www.kaset51.com

$\frac{rd}{\sqrt{sales_i}} = \beta_1 \frac{1}{\sqrt{sales_i}} + \beta_2 \sqrt{sales_i} + \frac{U_i}{\sqrt{sales_i}}$ ในโปรแกรม E-Views ต้องมีการ Generate Series ใหม่เพื่อใช้ในการประมาณสมการข้างต้น ซึ่งปรากฏผลดังตารางที่ 30 เมื่อเทียบกับตารางที่ 28 พบว่าค่าสัมประสิทธิ์ความชันนั้นค่อนข้างจะเหมือนกัน แต่ Standard Errors ของความชันนั้นต่างกันโดยที่กรณีในตารางที่ 28 นั้นจะมีค่า Standard Errors เท่ากับ 0.0083 ส่วนในตารางที่ 30 นั้น ค่า Standard Errors เท่ากับ 0.0071

ตารางที่ 30 Transform Model โดย Weighted ด้วย $1/\sqrt{Sales_i}$

Dependent Variable: RD_SROOT
Method: Least Squares
Date: 02/22/07 Time: 00:09
Sample: 1 18
Included observations: 18

Variable	Coefficient	Std. Error	t-Statistic	Prob.
SALE_SROOT_INVERSE	-246.6769	381.1285	-0.647228	0.5267
SALES_SROOT	0.036798	0.007114	5.172315	0.0001
R-squared	0.364889	Mean dependent var		8.855264
Adjusted R-squared	0.325195	S.D. dependent var		8.834378
S.E. of regression	7.257134	Akaike info criterion		6.906286
Sum squared resid	842.6560	Schwarz criterion		7.005217
Log likelihood	-60.15658	Durbin-Watson stat		2.885313

นอกจากวิธี Transform Model แล้ว หากใช้วิธี White Heteroskedasticity-Consistent Standard Errors & Covariance ก็ปรากฏผลดังแสดงในตารางที่ 31 ซึ่งเทียบกับสมการดั้มเดิมในตารางที่ 28 พบว่าค่าสัมประสิทธิ์นั้นเหมือนกันเพียงแต่ค่า Standard Errors ของความชันมีค่าเพิ่มขึ้นเล็กน้อย ส่วน Standard Errors ของ Intercept มีค่าลดลงแต่อย่างไรก็ตามตารางที่ 31 นั้นมีตัวอย่างจำนวนน้อยจึงต้องระมัดระวังในการตีความและนำผลไปใช้

ตารางที่ 31 White Heteroskedasticity-Consistent Standard Errors & Covariance

Dependent Variable: RD
Method: Least Squares
Date: 02/22/07 Time: 00:48
Sample: 1 18
Included observations: 18
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	192.9931	533.9317	0.361457	0.7225
SALES	0.031900	0.010147	3.143815	0.0063

R-squared	0.478303	Mean dependent var	3056.856
Adjusted R-squared	0.445697	S.D. dependent var	3705.973
S.E. of regression	2759.153	Akaike info criterion	18.78767
Sum squared resid	1.22E+08	Schwarz criterion	18.88660
Log likelihood	-167.0891	F-statistic	14.66916
Durbin-Watson stat	3.015607	Prob(F-statistic)	0.001476