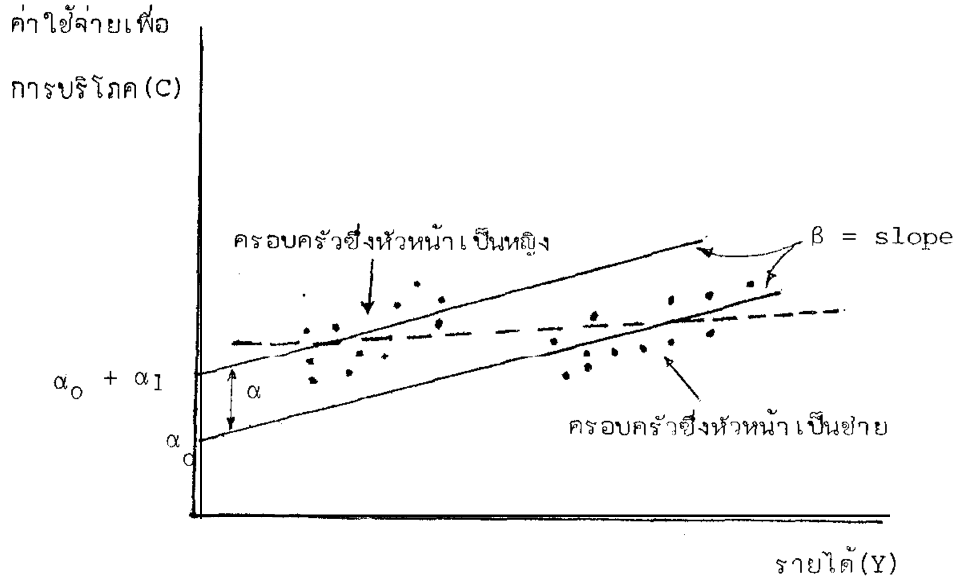


รูปที่3.3 ความสัมพันธ์ของค่าใช้จ่ายเพื่อการบริโภคในครอบครัวกับรายได้

<u>ตัวอย่างที่ 3.2</u>

จงประมาณความสัมพันธ์ระหว่างค่าใช้จ่ายเพื่อการบริโภคของบุคคลและรายได้จากข้อมูลปี 1930 -49 ซึ่งในช่วงนี้จะรวมปีที่เกิดสงครามโลกครั้งที่ 2 ด้วย (1942 - 45) และเราคาดว่าความโน้มเอียงในการบริโภคเฉลี่ยในภาวะสงครามจะต่ำกว่าปกติเพราะในภาวะสงครามมักจะมีการปันส่วนและการควบคุมการบริโภค ดังนั้น จึงกำหนดให้ Dummy Variable แทนผลจากภาวะสงคราม โดยให้ D =1 ในภาวะสงคราม และ D = 0 ในภาวะปกติ

ตารางที่ 3.4

ปี	บริโภค (C)	รายได้ (Y)	(D)				
	Y	X	Z	x	z	.	.
1930	C_o	Y_o	0	.	.	.	
.	C_1	Y_1	0	.	.	.	.
		.	0	.	.	.	
			0				
1942			1	.	.	.	
1943			1	.	.	.	
1944			1	.	.	.	
1945			1	.	.	.	
.			[illegible]	.	.	.	
๑๙๔๘	C_{48}	Y_{48}	0	.	.	.	
1949	C_{49}	Y_{49}	0	.	.	.	

สมการ regression :

$$C = a + bY + cD + e$$

การคำนวณ

คำนวณตามวิธี Multiple Regression กำลังสองน้อยที่สุด ตามที่เรียนมาแล้ว (C_0, C_1 ... และ Y_0, Y_1 .... ต้องเป็นข้อมูลตัวเลข) โดยหาสัมประสิทธิ a, b, c และค่าสถิติต่าง ๆ สมมุติว่าได้สมการ regression และค่าสถิติดังนี้

$$C = 2.656 + 0.907\ Y - 22.746\ D \qquad \ldots\ldots\ldots\ldots(1)$$

$$\quad (2.393) \quad (0.014) \qquad (1.7)$$

$$R^2 = 0.9960, \quad \bar{R}^2 = 0.9955$$

$$SEE = 2.863$$

$$\Sigma e^2 = 136.72$$

จะเห็นว่า ภาวะสงครามจะมีผลต่อการบริโภค คือ ทำให้การบริโภคลดลง (c เป็นลบ) ตามที่คาดไว้ อาจจะมีผู้แย้งว่า การเปลี่ยนแปลงนี้เกิดขึ้นโดยบังเอิญ เราก็อาจจะทดสอบได้ด้วย ratio ของค่า c กับ Standard Error ของมัน ($\hat{t}_c$) จะได้ $\frac{22.746}{1.7} = 13.38$ (> ตาราง t) ซึ่งเป็นไปไม่ได้อย่างยิ่งว่า C = 0 แสดงว่า ภาวะสงครามมีผลต่อการบริโภคจริง

การทดสอบนัยสำคัญของสัมประสิทธิของ D (c) อาจจะทำได้โดย F -testซึ่งคำตอบก็จะได้เหมือน t - test ($F = t^2$)

ถ้าหากเราไม่พิจารณาภาวะสงคราม (โดยไม่ใช้ Dummy Variable) เส้น regression ที่คำนวณจากการบริโภคและรายได้จะได้เป็น :

$$C = 5.044 + 0.839\ Y \qquad \ldots\ldots\ldots(2)$$

$$(4.898) \quad (0.043)$$

$$R2 = 0.9540 , \quad \bar{R}^2 = 0.9514$$

$$SEE = 9.361 \qquad \Sigma e^2 = 1577.31$$

จะเห็นว่า Marginal Propensity to Consume จะต่ำกว่า MPC ที่แท้จริง (b ของ (2) < ของ (1) ความคลาดเคลื่อนมาตรฐานของค่าประมาณ (SEE) และผลบวกกำลังสองของความคลาดเคลื่อน (Σe^2) ของ (2) ก็สูงกว่า (1) ซึ่งแสดงว่า ค่าพยากรณ์ของ (2) จะคลาดเคลื่อนมาก

ดังนั้น จะเห็นว่า การใช้ตัวแปรดัมมี่จึงมีความสำคัญมากและมีประโยชน์ในการวิเคราะห์พฤติกรรมของตัวแปรเศรษฐกิจที่มีทั้งตัวแปรที่วัดได้และวัดไม่ได้ ดังจะเห็นจากตัวอย่างที่กล่าวข้างต้น

3. <u>ข้อจำกัดของวิธีกำลังสองน้อยที่สุด</u>

การคำนวณเส้นถดถอยตามวิธีกำลังสองน้อยที่สุด (OLS) จะให้ความแปรปรวนน้อยที่สุด แต่วิธีกำลังสองน้อยที่สุดนั้นมีข้อจำกัด คือ การคำนวณจะต้องเป็นไปตามข้อสมมุติพื้นฐานใหญ่ ๆ 4 ประการที่กล่าวแล้วในบทที่ 2 คือ $E(u_i) = 0$, $Cov.(u_i, u_j) = 0$ ถ้า $i \neq j$, $Var.(u_i) = \sigma_u^2$ และ $E(\emptyset, u_i) = 0$ ซึ่งข้อสมมุติเหล่านี้อาจจะอธิบายให้เข้าใจง่าย ๆ ได้ว่า

1. ความคลาดเคลื่อนที่รวมอยู่ในตัวแบบต้องเป็นตัวแปรอิสระสุ่ม มีการกระจายปกติ (Normal distribution) มีค่าเฉลี่ยเลขคณิต = 0
2. ไม่มีสหสัมพันธ์ในตัว
3. ตัวคลาดเคลื่อนและตัวแปรอิสระต้องเป็นอิสระต่อกัน
4. ตัวแปรอิสระแต่ละตัวต้องไม่มีปัญหา Multicollinearity
5. การวัดตัวแปรอิสระต้องไม่ผิดพลาด
6. ความแปรปรวนของตัวคลาดเคลื่อนคงที่ตลอดทั้งเส้น (σ_u^2 คงที่) เรียกว่า homoscedasticity(แต่ความแปรปรวนแต่ละตัวไม่จำเป็นต้องเท่ากัน) แต่ถ้าความแปรปรวนไม่คงที่ เรียกว่า heteroscedasticityก็จะทำเส้นถดถอยที่คำนวณได้เกิด bias ได้
7. ผลรวมของตัวแปรทั้งหมดต้องถูกต้อง

8. ฟังชั่นที่คำนวณจะต้องเข้าใจได้ว่าเป็นอะไร

9. รายการรายละเอียดของตัวแบบจะต้องถูกต้อง

ถ้าหากเส้นถดถอยไม่เป็นไปตามข้อสมมุติก็จะทำให้เกิดปัญหาดังต่อไปนี้ คือ

3.1 ปัญหา Autocorrelation

ข้อสมมุติพื้นฐานข้อหนึ่งของ regression model คือ ค่าของตัวคลาดเคลื่อน (disturbance) จะต้องเป็นอิสระต่อค่าของมันเองในอีกเวลาหนึ่งจนทำให้

$$\text{Cov.}(u_i, u_j) = 0 \qquad i \neq j$$

แต่ถ้า u ไม่เป็นอิสระตามข้อสมมุติดังกล่าวก็จะทำให้เกิดสหสัมพันธ์ในตัว (Autocorrelation) ขึ้น

การเกิด Autocorrelation:

ประการแรก ปัญหา Autocorrelation มักเกิดกับข้อมูลเศรษฐกิจที่เป็นอนุกรมเวลาซึ่งมักจะไม่สุ่ม เช่น ราคา ผลผลิต สินค้าคงคลัง เงินออม และตัวแปรเศรษฐกิจอื่น ๆ ในปีที่กำหนดมักจะไม่เป็นอิสระ ปกติจะมีความสัมพันธ์กับตัวแปรเดียวกันในปีก่อน เรียกว่าไม่เป็นอิสระต่อกัน คำว่า Autocorrelationจะถูกใช้เพื่ออธิบายความสัมพันธ์ที่ล่าช้า (lag) ของอนุกรมเวลาโดยเฉพาะกับตัวของมันเอง ซึ่งล่าช้าไปหนึ่งหน่วยเวลา ตัวแปร X ของข้อมูลอนุกรมเวลาจะมีสหสัมพันธ์ในตัวถ้า X_t (X ในเวลา t) สัมพันธ์กับ X_{t-1} (X ในเวลาt-1) หรือ X_t สัมพันธ์กับ X_{t-2}, X_{t-3} เป็นต้น

บางครั้งเราเรียก Autocorrlation ว่า Serial Correlation [2/]

ประการที่สอง ตัวคลาดเคลื่อนใน model เป็นผลจากที่ไม่ได้นำตัวแปรบางตัวเข้ามาใน model ดังนั้น ควรจะพิจารณาสมการ (equation) ใหม่อีกครั้ง

ประการที่สาม อาจจะสร้างรูปฟังชั่นที่ผิด เช่น ฟังชั่นเส้นโค้งสร้างเป็นฟังชั่นเส้นตรง ฉะนั้น ตัวคลาดเคลื่อนก็จะแสดง Autocorrelation ออกมา

เมื่อเกิดมี Autocorrelation ขึ้นเช่นนี้ จะทำให้ตัวประมาณค่า (estimator) มี Variance ที่ไม่ต่ำสุด ฉะนั้นก่อนที่จะทำการประมาณค่าต่าง ๆ ควรจะมีการทดสอบว่า ข้อมูลมี Autocorrelationหรือไม่ วิธีการทดสอบมีหลายวิธีเช่น วิธีการของ Durbin - Watson เรียกว่า ค่าสถิติเดอร์บิน - วัตสัน (The Durbin- Watson Statistic)วิธีของ Von - Neuman เรียกว่าอัตราส่วนฟอนนิวแมน (The Von Neuman Ratio)วิธีของ B.I.Hart เรียก The mean square successive different method และยังมีวิธีอื่น ๆ อีกหลายวิธี ในที่นี้จะกล่าวถึง เฉพาะค่าสถิติของ เดอร์บิน - วัตสัน ซึ่ง เป็นวิธีที่ใช้กันแพร่หลาย และ เพื่อความ เข้าใจเบื้องต้นในเรื่อง Autocorrelation

การทดสอบ Autocorrelation โดยค่าสถิติ Durbin- Watson :

ให้ γ เป็นสัมประสิทธิของ Autocorrelation ซึ่ง γ จะวัดสหสัมพันธ์ระหว่างความคลาดเคลื่อน u

---

2/ บางท่านให้ความหมายของ Autocorrelation กับ Serial Correlation แตกต่างกัน เช่น Autocorrelation เป็นสหสัมพันธ์ของอนุกรมข้อมูลที่ล่าช้ากับตัวของมันเอง ส่วน Serial Correlationเป็นสหสัมพันธ์ที่ล่าช้าระหว่างอนุกรมข้อมูล2 อนุกรม หรือ บางท่านกล่าวว่า Autocorrelationเป็นสหสัมพันธ์ในตัวของข้อมูลที่มีประชากร (Population) เป็นพื้นฐาน ส่วน Serial Correlationเป็นสหสัมพันธ์ในตัวของข้อมูลที่มีตัวอย่าง (Sample)เป็นพื้นฐาน อย่างไรก็ตาม โดยมากมักจะใช้แทนกัน

สมมุติฐาน H_0 : $\gamma = 0$ (ความคลาดเคลื่อนไม่มี Autocorrelation)

H_a : $\gamma \neq 0$ (ความคลาดเคลื่อนมี Autocorrelation)

ค่าสถิติของ Durbin - Watson ใช้สัญลักษณ์ว่า d

$$\hat{d} = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

หรือเขียนสั้น ๆ ได้ว่า $\hat{d} = \frac{\Sigma (\Delta \hat{e})^2}{\Sigma \hat{e}^2}$

e เป็นค่าความคลาดเคลื่อนของข้อมูลตัวอย่างและเป็นตัวแทนของ u ซึ่งเป็นค่าความคลาดเคลื่อนของประชากร การหา e หาตามวิธีการเส้นถดถอย คือ ต้องหา $\hat{Y} = \hat{a} + \hat{b}X + e$ ก่อนแล้วนำ $\hat{Y}$ ไปหักจาก Y_i จะได้ e $(e = Y_i - \hat{Y})$

$e_{t-1} = e_t$ ที่ล่าช้าไป1 ระยะ

ค่า $\hat{d}$ เป็นค่าที่คำนวณจาก e_t ซึ่ง e_t ขึ้นอยู่กับค่า X จึงทำให้การทดสอบไม่มีค่าวิกฤตเดียว Durbin - Watson ได้เสนอค่าวิกฤตเป็นช่วง คือ ขอบเขตขั้นสูง (Upper bound : d_u) และขอบเขตขั้นต่ำ (Lower bound : d_L) ดังตารางที่ 3.5 โดยมีจำนวนตัวอย่างข้อมูล (n) เริ่มจาก15 ถึง 100 ตัว มีจำนวนตัวแปรอิสระไม่เกิน 5 ตัว

ถ้า Σe_t^2 และ Σe_{t-1}^2 มีค่าเท่ากันโดยประมาณหรืออาจจะต่างกันเพียงเล็กน้อย เราจะได้ความสัมพันธ์ระหว่าง d กับ γ คือ

$$d = 2(1-\gamma)$$

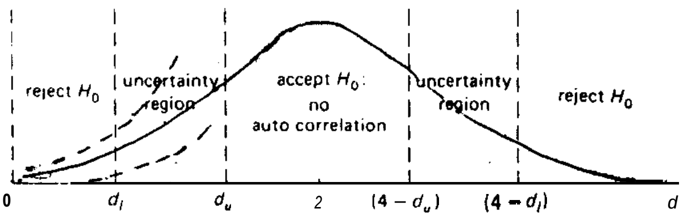
ถ้าความคลาดเคลื่อน u เป็นอิสระต่อกันไม่มีสหสัมพันธ์กัน จะทำให้ $\gamma = 0$ และจากสมการแสดงความสัมพันธ์ d และ γ เมื่อ $\gamma = 0$ จะให้ d = 2 นั่นคือ ไม่มีปัญหาสหสัมพันธ์ในตัว แต่ถ้า γ มีค่าใกล้ 1 ค่า d มีแนวโน้มเข้าสู่ 0 ซึ่งเป็นกรณี Positive Autocorrelation เมื่อ γ มีค่าเข้าใกล้ -1 ค่า d มีแนวโน้มเข้าใกล้ 4 ซึ่งเป็นกรณี negative Autocorrelation โดยสรุป

$$\gamma = 0, \quad d = 2$$

$$\gamma = 1, \quad d = 0$$

$$\gamma = -1, \quad d = 4$$

จากความสัมพันธ์นี้ ก็จะได้ช่วงยอมรับและช่วงปฏิเสธดังรูปที่3.4 ถ้าใช้ตัวอย่างขนาดใหญ่ (n มาก) d จะมีค่าใกล้ 2



รูปที่3.4ช่วงนัยสำคัญของค่าสถิติ d

ภายใต้ข้อสมมุติฐานข้างต้น นำค่าสถิติ $\hat{d}$ ที่คำนวณได้ไปเปรียบเทียบกับ d ในตาราง ถ้า $\hat{d}$ ตกอยู่นอกช่วงค่าวิกฤตก็สรุปได้ว่ามีปัญหาสหสัมพันธ์ในตัวหรือไม่ และเป็นชนิดใด ดังนี้คือ

1. เมื่อ $d < 4 - d_u$: ยอมรับ H_0 แต่ปฏิเสธ H_a ซึ่งแสดงว่าไม่มีปัญหาสหสัมพันธ์ในตัว

2. เมื่อ $d > 4 - d_L$: ยอมรับ H_a ปฏิเสธ H_0 มีสหสัมพันธ์ในตัว (Negative Correlation)

3. เมื่อ $d \leqslant d_L$ ยอมรับ H_a ปฏิเสธ H_0 มีสหสัมพันธ์ในตัว (Positive autocorrelation)

4. เมื่อ $d_L \leqslant d \leqslant d_u$ และ $4 - d_u \leqslant d \leqslant 4 - d_L$ ไม่สามารถสรุปผลการทดสอบได้

ตารางที่ 3.5

ILLUSTRATIVE CRITICAL POINTS OF THE DURBIN-WATSON TEST FOR AUTOCORRELATION

Sample Size	Error Probability, α	Number of Explanatory Variables: 1		2		3		4	
		d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	.01	.81	1.07	.70	1.25	.59	1.46	.49	1.70
	.05	1.08	1.36	.95	1.54	82	1.75	69	1.97
20	01	.95	1.15	.86	1.27	.77	1.41	.68	1.57
	.05	1.20	1.41	1.10	1.54	1.00	1.68	.90	1 83
25	.01	1.05	1.21	.98	1.30	.90	1.41	.83	1.52
	.05	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77
30	.01	1.13	1.26	1.07	1.34	1.01	1.42	.94	1.51
	.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74
40	.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52
	.05	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72
50	.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1 54
	.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72
60	.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56
	.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73
80	.01	1.47	1.51	1.44	1.54	1.42	1.57	1.39	1.60
	.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74
100	.01	1.51	1.56	1.50	1.58	1.48	1.60	1.46	1.63
	.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76

ตัวอย่างที่ 3.3

จงคำนวณหาว่าค่าสถิติ d ของข้อมูลต่อไปนี้

Y	1	6	18	19	25	36	40	58	66	71
X	13	17	32	38	45	56	58	75	85	91

วิธีทำ

1)คำนวณหาเส้น regression $Y = a + bX$ ก่อน

ตารางที่ 3.6

n	Y	X	$y = Y - \bar{Y}$	$x = X - \bar{X}$	x^2	xy
1	1	13	-33	-38	1,444	1,254
2	6	17	-28	-34	1,156	952
3	18	32	-16	-19	361	304
4	19	38	-15	-13	169	195
5	25	45	-9	-6	36	54
6	36	56	2	5	25	10
7	40	58	6'	7	49	42
8	58	75	24	24	576	576
9	66	85	32	34	1.156	1,088
10	71	91	37	40	1.600	1,480
	340	510			6,572	5.955

$\bar{Y} = 34$ $\bar{X} = 51$

$$\hat{b} = \frac{\Sigma xy}{\Sigma x^2} = \frac{5,955}{6,572} = 0\ .9$$

$$\hat{a} = \bar{Y} - \hat{b}\bar{X} = 34 - 0.9(51) = -12$$

เส้น regression จะเป็น

$$\hat{Y} = -12 + 0.9\ x$$

2) คำนวณค่าสถิติ d

ตารางที่ 3.7

$\hat{Y} = -12+0.9x$	$e_t = Y - \hat{Y}$	$\Delta e = e_t - e_{t-1}$	e_t^2	$(\Delta e)^2$
-.3	1.3		1.69	
3.3	2.7	1.4	7.29	1.96
16.8	1.2	-1.5	1.44	2.25
22.2	-3.2	-4.4	10.24	19.36
28.5	-3.5	-0.3.	12.25	0.09
38.4	-2.4	1.1	5.76	1.21
40.2	-0.2	2.2	0.04	4.84
55.2	2.5	2.1	6.25	7.29
64.5	1.5	-1.0	2.25	1.00
69.9	1.1	-0.4	1.21	0.16
			48.42	38.16

$$\hat{d} = \frac{\Sigma(\Delta\ e)^2}{\Sigma\ \hat{e}^2}$$

$$\hat{d} = \frac{=38.36}{48.42} \qquad 0.79$$

ถ้าต้องการจะทดสอบควรจะใช้ n ไม่ต่ำกว่า15 แล้วดำเนินการทดสอบตามวิธีที่กล่าว อย่างไรก็ตาม ค่า $\hat{d}$ ที่ได้ข้างบนค่อนข้างต่ำ มีแนวโน้มที่จะเกิด Autocorrelation ได้

3 . 2 วิธีกำลังสองสองขั้น (Two Stage Least **Square** Method:2SLS)

3.2.1 กรณีสมการเดี่ยว (Single Equation)

ใน model ความสัมพันธอย่างง่ายของข้อมูลประชากร

$$Y_i = \alpha + \beta X_i + u_i \qquad (i = 1,2,.....n)$$

โดยมี u_i เป็นอิสระและกระจายปกติ มี mean = 0 , Variance = σ_u^2 แต่ข้อสมมุติพื้นฐานที่ว่า $E(\emptyset . u_i) = \emptyset\ E(u_i) = 0$ จะใช้ไม่ได้ถ้า X_i มีความสัมพันธ์กับ u ตัวประมาณค่าโดยวิธี OLS จะให้ค่าที่ลำเอียง (bias) และไม่สอดคล้อง (inconsistent) วิธีกำลังสองสองขั้น (2SLS) จะช่วยขจัดความสัมพันธระหว่าง X กับ u จนกระทั่งทำให้สมการเป็นไปตามข้อสมมุติของ regression model วิธีการประมาณคือ

ขั้นที่ 1 จะขจัดตัวแปรอิสระซึ่งสัมพันธ์กับความคลาดเคลื่อน นั่นก็คือมีค่าที่แก้ไขใหม่สำหรับตัวแปรอิสระที่สงสัยจะมีความสัมพันธ์กับความคลาดเคลื่อน ค่าที่แก้ไขเหล่านี้จะไม่สัมพันธ์กับความคลาดเคลื่อน ค่าที่นำมาแก้ไขเรียกว่า "ตัวแปรเครื่องมือ" (Instrumental Variable) โดยให้ตัวแปรเครื่องมือทำหน้าที่เป็นตัวแปรอิสระไม่มีความสัมพันธ์กับ e และให้ตัวแปรอิสระที่สงสัยเป็นตัวแปรตาม แล้วคำนวณหาเส้นถดถอยโดยวิธีกำลังสองน้อยที่สุดอย่างง่าย (OLS) เช่น

$Y = a + bX + e$ เราสงสัยว่า X จะสัมพันธ์กัน e จึงกำหนดให้ Z ซึ่งไม่มีความสัมพันธ์กับ e เป็นตัวแปร เครื่องมือทำหน้าที่เป็นตัวแปรอิสระในสมการโดยมี X เป็นตัวแปรตาม จะได้เส้นถดถอยเป็น

$$\hat{X} = \hat{c} + \hat{d}Z$$

ขั้นที่ 2 จะเป็นการประมาณค่าพารามิเตอร์ธรรมดาตามวิธีกำลังสองน้อยที่สุด (OLS) อีกครั้งหนึ่ง โดยมี $\hat{X}$ เป็นตัวแปรอิสระและไม่มีความสัมพันธ์กับ e Y เป็นตัวแปรตาม จะได้

$$\hat{Y} = \hat{a} + \hat{b}\hat{X} + e$$

สำหรับสมการเชิงซ้อน ซึ่งมี model เป็น

$$\hat{Y} = \alpha + \beta_1 x_{1i} + \beta_2 X_{2i} + .. + \beta_n X_{ki} + u_i$$

การขจัดความความสัมพันธ์ระหว่างตัวแปรอิสระหนึ่งหรือหลายตัวที่มีความสัมพันธ์กับความคลาดเคลื่อน u_i มีวิธีการเหมือนกับสมการถดถอยอย่างง่าย

<u>ตัวอย่างที่ 3.4</u>

จงใช้วิธีกำลังสองน้อยที่สุดประมาณค่าพารามิเตอร์ของ X Y และ Z โดยให้ Z เป็นตัวแปรเครื่องมือ

Y	10	19	16	22	23	45	48	61	70	76
X	5	8	6	7	10	12	13	15	20	24
Z	2	4	3	5	8	11	12	15	14	16

วิธีทำ

ขั้นที่ 1 คำนวณ X on Z จะได้

$$\hat{X} = \hat{c} + \hat{d}Z$$

ตารางที่ 3.8

X	Z	$x = (X-\bar{X})$	$z = (Z-\bar{Z})$	xz	z^2	$\hat{X} = 2.208 + 1.088\ Z$
5	2	-7	-7	49	49	4.83
8	4	-4	-5	20	25	6.56
6	4	-6	-6	36	36	5.47
7	5	-5	-4	20	16	7.65
10	8	-2	-1	2	1	10.91
12	11	0	2	0	4	14.18
13	12	1	3	3	9	15.26
15	15	3	6	18	36	18.53
20	14	8	5	40	25	17.44
24	16	12	7	48	49	19.62
120	90			272	250	
$\bar{X} = 12$	$\bar{Z} = 9$					

$$\hat{c} = \bar{X} = 12$$

$$\hat{d} = \frac{\Sigma xz}{\Sigma z^2} = \frac{272}{250} = 1,088$$

ถ้าใช้สูตรสัมประสิทธิ 2 สูตรข้างบนนี้ สมการถดถอยจะเป็น

$$\hat{X} = 12 + 1.088\ (z)$$

เปลี่ยนรูปเดิมโดยแทนค่า z

$$\hat{X} = 12 + 1.088\ (Z - \bar{Z})$$

$$= 12 + 1.088\ Z - 1.088\ (9)$$

$$\hat{X} = 2.208 + 1.088\ Z$$

<u>ขั้นที่ 2</u> คำนวณ Y on $\hat{X}$ จะได้

$$\hat{Y} = \hat{a} + \hat{b}\hat{X} + e$$

ตารางที่ 3.9

Y	$\hat{X}$	$y = Y - \bar{Y}$	$\hat{x} = \hat{X} - \bar{X}$	$\hat{x}^2$	$\hat{x}y$
10	4.38	-30	-7.62	59.99	228.5
19	6.56	-21	-5.44	29.58	114.2
16	5.47	-24	-6.54	42.60	156.6
22	7.65	-18	-4.35	l.8.93	78.3
33	10.91	- 7	-1.09	1.18	7.6
4.5	14.18	5	2.18	4.74	10.9
48	15.26	8	3.27	lb.66	26.1
61	18.53	21	6.53	42.63	137.1
70	17.44	30	5.44	29.57	163.4
76	19.62	36	7.62	58.02	274.2
400				295.90	1,196.7

$\bar{Y}=40$

$$\hat{a} = \bar{Y} = 40$$

$$\hat{b} = \frac{\Sigma x\hat{y}}{\Sigma \bar{x}^2} = \frac{1196.1}{295.9} = 4.0441$$

$$\hat{Y} = 40 + 4.0441\,(\hat{X} - \bar{\hat{X}})$$

$$= 40 + 4.0441\,\hat{X} - 4.0441\,(12.00)$$

$$= -8.5292 + 4.0441\,\hat{X}$$

3.2.2 กรณีสมการเกี่ยวเนื่อง (Simultaneous Equation)

ตัวแบบทางเศรษฐศาสตร์บางครั้งประกอบด้วยสมการมากกว่าหนึ่งสมการ ซึ่งเป็นสมการเกี่ยวเนื่อง (Simulteneous Equation) วิธีกำลังสองน้อยที่สุดสองขั้นนอกจากจะใช้ประมาณค่าของ Single Equation ซึ่งตัวแปรอิสระมีความสัมพันธ์กับความคลาดเคลื่อนแล้วยังเป็นวิธีการหาค่าสมการ Simulteneous Equation ซึ่งตัวแปรตามหรือตัวแปรอิสระมีความสัมพันธ์กับความคลาดเคลื่อน โดยให้ตัวแปรบางตัวที่เป็นตัวแปรอิสระหรือตัวแปรภายนอก (Exogeneous Variable) ของสมการเกี่ยวเนื่องนั้นเป็นตัวแปรเครื่องมือ (Instrumental Variable) เช่น

(1) สมการรายได้ประชาชาติ :

$$Y = C + I$$

$$C = a + bY + e$$

Y = รายได้ประชาชาติ

C = ค่าใช้จ่ายในการบริโภค

I = ค่าใช้จ่ายในการลงทุน

a = การบริโภคที่ไม่ได้ขึ้นอยู่กับรายได้ (autonomous consumption)

b = MPC

e = error term

เราจะหาค่าใช้จ่ายในการบริโภค (C) โดยใช้สมการ $C = a + bX + e$ ตามวิธี OLS ไม่ได้ ถึงแม้ C จะถูกกำหนดโดย Y แต่ C ก็อาจจะกำหนด Y ได้ดังสมการ $Y = C + I$

ดังนั้นทั้ง Y และ C ต่างก็เป็นตัวแปรตามหรือตัวแปรภายใน (Endogeneous Variable) ซึ่งต่างก็กำหนดซึ่งกันและกันและต่างก็ถูกกำหนดโดย I และ e Y จึงสัมพันธ์กับe I เป็นตัวแปรภายนอกซึ่งถูกกำหนดโดยอิทธิพลภายนอกสมการ, I และ e ไม่มีความสัมพันธ์กัน I จึงเป็นตัวแปรเครื่องมือค่าใช้จ่ายในการบริโภค จะหาได้ดังนี้

<u>ขั้นที่1</u>

หาเส้นถดถอย Y on I โดยวิธี OLS จะได้

$$\hat{Y} = b_0 + b_1 I$$

<u>ขั้นที่ 2</u>

ให้ $\hat{Y}$ แทน Y ในสมการ C , $\hat{Y}$ จะไม่สัมพันธ์กับ e ก็จะประมาณค่า a และb ได้ตามวิธี OLS

$$\hat{C} = \hat{a} + \hat{b}\hat{Y} + e$$

(2) สมการ demand และ supply :

$$Q_d = a + bP + cT + dY + e$$

$$Q_s = f + gP + hR + v$$

$$Q_d = Q_s$$

Q_d = Quantity demand

Q_s = Quantity supply

P, T, Y, R = ตัวแปรหมายถึง ราคา รสนิยม รายได้ **น้ำฝน**

a, b, c, d, f, g, h = parameters

e, v = ความคลาดเคลื่อน

เราสามารถหาตัวแปรภายนอกมาใส่ในสมการที่ต้องการหาค่าพารามิเตอร์ได้ โดยตัวแปรภายนอกนั้นไม่มีความสัมพันธ์กับความคลาดเคลื่อนของอีกสมการหนึ่ง และตัวแปรภายนอกตัวนั้นจะถูกใช้เป็นตัวแปรเครื่องมือ สมการ supply (Q_s) มีความสัมพันธ์กับปริมาณน้ำฝน (R) แต่ปริมาณน้ำฝนไม่สัมพันธ์กับความคลาดเคลื่อน (e) ในสมการ demand (Q_d) R จึงเป็นตัวแปรภายนอกที่ใช้เป็นตัวแปรเครื่องมือ ทำนองเดียวกัน รสนิยม (T) และรายได้ (Y) ไม่สัมพันธ์กับความคลาดเคลื่อน (v) ในสมการ Supply, T Y จึงเป็นตัวแปรภายนอกที่นำมาเป็นตัวแปรเครื่องมือ ดังนั้นระบบสมการนี้จึงมีตัวแปรเครื่องมือทั้งหมดคือ T, Y และ R

เมื่อได้ตัวแปรเครื่องมือทั้งหมดแล้ว เราก็หาตัวประมาณค่าหรือพารามิเตอร์ได้

ขั้นที่ 1 หา เส้นถดถอยเชิงซ้อน $\hat{P}$ on T,Y และ R ตามวิธี OLS คือ

$$\hat{P} = b_0 + b_1T + b_2Y + b_3R$$

T, Yและ R เป็นอิสระจากความคลาดเคลื่อน v ดังนั้นP ก็เป็นอิสระจาก v เหมือนกัน

ขั้นที่ 2

ถ้าจะหาสมการ demand ใส่ $\hat{P}$ แทน P ในสมการ Q_d แล้วประมาณค่าพารามิเตอร์ a, b, c และ d โดยวิธี OLS จะได้

$$Q_d = a + bP' + cT + dY + e$$

ถ้าจะหาสมการ supply ใส่ $\hat{P}$ แทน P ในสมการนั้น ค่า f g และ h จะประมาณได้โดยวิธี OLS คือ

$$Q_s = f + g\hat{P} + hR + v$$

4. ปัญหาการเลือกตัวแปร

ปัญหาการเลือกตัวแปรเป็นปัญหาทางปฏิบัติ กล่าวคือ ก่อนที่จะสร้าง regression model ผู้ศึกษาต้องยึดทฤษฎีในการตัดสินใจเกี่ยวกับตัวแปรตามและตัวแปรอิสระซึ่งสามารถจะอธิบายทฤษฎีได้ดีที่สุด แต่การดำเนินงานอาจจะมีการผิดพลาดขึ้นได้โดยเฉพาะเกี่ยวกับตัวแปรอิสระ

ประการแรก อาจจะไม่ได้นำเอาตัวแปรอิสระหรือตัวแปรที่ทำให้เกิดการถดถอย (Regressor) ที่สำคัญเข้าไว้ใน model โดยมองข้ามความสำคัญของมันไป ผลก็คือ เมื่อพยากรณ์ค่าตัวแปรตามหรือตัวแปรถดถอย (Regressand)จะทำให้ตัวประมาณค่าให้ค่าที่ลำเอียง (bias)

ไม่สอดคล้อง (inconsistant) และไม่มีประสิทธิภาพ (inefficiency) ซึ่งไม่ใช่คุณสมบัติของการประมาณค่าที่ดี

ประการที่สอง อาจจะรวมเอาตัวแปรบางตัวเข้ามาใน model ทั้ง ๆ ที่ไม่มีความสำคัญ นั่นคือ model จะมีตัวแปรอิสระมากเกินไป จะทำให้ความแปรปรวนของตัวประมาณค่าพารามิเตอร์เพิ่มขึ้น ตามจำนวนตัวแปรอิสระที่เพิ่มขึ้น เมื่อความแปรปรวนมาก ค่าสัมประสิทธิหรือพารามิเตอร์จะมีความถูกต้องน้อยลง ฉะนั้นเราควรจะเลือกตัวแปรด้วยเหตุผลที่เชื่อว่ามีอิทธิพลต่อตัวแปรตาม

อีกทางหนึ่งที่จะแก้ปัญหาการเลือกตัวแปร ซึ่งนักวิจัยมักจะใช้กันคือ จะทดลองตัวแปรที่คิดว่าสำคัญเข้ามาในสมการแล้วทดสอบนัยสำคัญทางสถิติ ถ้าปรากฏว่าไม่มีนัยสำคัญต่างไปจาก 0 ก็จะทิ้งมันไปจากสมการ ตัวแปรที่ใช้ได้ก็จะเก็บเอาไว้ ทำเช่นนี้ไปเรื่อยจนกว่าจะหมดทุกตัวแปร ก็จะได้ตัวแปรที่สำคัญรวมเข้าไว้ใน model ตามต้องการ วิธีนี้เรียกว่า Stepwise - Regression สำหรับวิธีการคำนวณในรายละเอียดจะไม่กล่าวถึง

แบบฝึกหัด

1. ถ้า $n = 20$

$\bar{Y} = 60$	$\Sigma(Y - \bar{Y})^2 = 215.5$	$\Sigma(Y - \bar{Y})(X - \bar{X}) = -80$
$\bar{X} = 31$	$\Sigma(X - \bar{X})^2 = 75$	$\Sigma(Y - \bar{Y})(Z - \bar{Z}) = -10$
$\bar{Z} = 10$	$\Sigma(Z - \bar{Z})^2 = 100$	$\sigma(X - \bar{X})(Z - \bar{Z}) = 50$

1.1 จงคำนวณหาสมการ regression Y on X on Z โดยวิธีกำลังสองน้อยที่สุด ($Y = 100 - 1.5X + 0.65Z$)

1.2 หา R^2, $\bar{R}^2$ (0.527, 0.471)

1.3 จงหา s^2, S_b (6, 0.35)

1.4 ค่าสถิติ t และทดสอบ (4.286)

1.5 ตาราง ANOVA ค่าสถิติ F และทดสอบ (9.46)

1.6 SEE (2.45)

2. ข้อมูลเกี่ยวกับปริมาณซื้อ ราคา และค่าโฆษณา ของบริษัทแห่งหนึ่งเป็นดังนี้

Q_d	55	70	90	100	90	105	80	110	125	115	130	130
P	100	90	80	70	70	70	70	65	60	60	55	50
A	5.50	6.30	7.20	7.00	6.30	7.35	5.60	7.15	7.50	6.90	7.15	6.50

จงคำนวน

2.1 เส้น demand (Q_d = 117.532 - 1.326 P + 11.237 A)

2.2 S^2 (20.586)

2.3 S_a, S_b, S_c (21.264, 0.112, 2.403)

2.4 ค่าสถิติ t

2.5 ทดสอบช่วงความเชื่อมั่น a, $\hat{b}$, $\hat{c}$, ที่ระดับ 95 % โดยทดสอบแบบสองข้าง

$$(69.43 < \hat{a} < 165.63)$$

$$(-1.58 < \hat{b} < -1.07)$$

$$(5.80 < c < 16.67).$$

2.6 สัมประสิทธิการตัดสินใจ (0.981)

2.7 สัมประสิทธิการตัดสินใจปรับปรุง

3. สมมุติรายจ่ายค่าอาหารของครอบครัวรายสัปดาห์ (F) เป็นฟังชั่นเส้นตรงของรายได้รายสัปดาห์ (Y) และขนาดของครอบครัว (N) จากตัวอย่าง 10 ครอบครัว ดังข้อมูลต่อไปนี้

รายจ่ายค่าอาหาร (บาท)	รายได้ (บาท)	ขนาดครอบครัว (คน)
25	80	2
36	70	4
40	100	4
47	150	3
49	180	3
50	200	2
53	190	3
55	130	5
60	225	4
65	175	5

3.1 จงคำนวณหาสมการ regression $F = a + bY + cN$, R^2, $\bar{R}^2$, SEE และความคลาดเคลื่อนของสัมประสิทธิทั้งสอง

3.2 การจ่ายค่าอาหารเป็นเศษส่วนเท่าไรของรายได้ที่เพิ่มขึ้น (0.1824)

3.3 ถ้าขนาดของครอบครัวเพิ่มขึ้น 1 คน โดยที่รายได้ไม่เปลี่ยนแปลง รายจ่ายค่าอาหารจะเปลี่ยนแปลงไปเท่าไร (6.2437)

3.4 จงหารายจ่ายค่าอาหารต่อรายได้โดยเฉลี่ย ($\frac{480}{1,500} = 0.32$)

คำตอบ 3.1 $F = -1.2072 + 0.1824\,Y + 6.2437\,N$

$(4.0625) \quad (0.0170) \quad (0.8360)$

$R^2 = 0.9590 \qquad \bar{R}^2 = 0.9473$

$SEE = 2.7056 \qquad \Sigma e^2 = 51.2539$