บทที่ 2

ปริทัศน์ผลงานศึกษาที่เกี่ยวข้อง

ในบทของปริทัศน์ผลงานศึกษาที่เกี่ยวข้องนี้ได้แบ่งเนื้อหาออกเป็น 2 ส่วน คือ ปริทัศน์ถึงบุคคลต่างๆ ที่ได้กล่าวถึงผลของการละทิ้งตัวแปรที่เกี่ยวข้องออกไปจากแบบจำลอง และปริทัศน์ผลงานวิจัยต่างๆ ที่ได้มีการละทิ้งตัวแปรที่เกี่ยวข้องออกไปจากแบบจำลอง

2.1 ปริทัศน์ผลของการละทิ้งตัวแปรที่เกี่ยวข้อง

Dhrymes (1978 : 220-227) กล่าวว่า การประมาณค่าพารามิเตอร์ที่มีการละทิ้งตัวแปรที่เกี่ยวข้องออกไปจากแบบจำลองคือ

$$\hat{\beta}^* = \left(X^{*\prime} X^*\right)^{-1} X^{*\prime} y \qquad (2.1)$$

ซึ่งจะทำให้เกิดผลดังนี้

1. มีการเอนเอียงจากการประมาณค่าพารามิเตอร์คือ $(P-I)\beta$ เมื่อ $P = \left(X^{*\prime} X^*\right)^{-1} X^{*\prime} X$ ซึ่งเมตริกซ์ P จะเป็นสัมประสิทธิ์ของการถดถอยของตัวแปรที่ถูกละทิ้งไปกับตัวแปรที่อยู่ในแบบจำลองทั้งหมด แต่ถ้าในกรณีที่ตัวแปรที่ถูกละทิ้งไปนั้นไม่มีความสัมพันธ์กับตัวแปรที่อยู่ในแบบจำลอง หรือเป็นเมตริกซ์ของตัวแปรที่ถูกละทิ้งเป็น Orthogonal กับตัวแปรที่อยู่ในแบบจำลอง ($X^{*\prime} X = 0$) การประมาณค่าพารามิเตอร์จะไม่เกิดการเอนเอียง

2. เกิดความไม่แนนนัย (inconsistent) คือ $\left(\overline{P}-I\right)\beta$ เมื่อ

$$\overline{P} = \underset{T\to\infty}{p\lim}\left(X^{*\prime} X^*\right)^{-1} X^{*\prime} X$$

3. ค่า MSE จะเป็น $MSE\left(\hat{\beta}^*\right) = (P-I)\beta\beta'(P-I) + \sigma^2\left(X^{*\prime} X^*\right)^{-1}$ (ยกเว้นกรณีที่เมตริกซ์ของตัวแปรที่ถูกละทิ้งเป็น Orthogonal กับตัวแปรที่อยู่ในแบบจำลอง ($X^{*\prime} X = 0$) จะทำให้ $MSE\left(\hat{\beta}^*\right) = \sigma^2\left(X^{*\prime} X^*\right)^{-1}$ และ $\underset{T\to\infty}{p\lim} MSE\left(\hat{\beta}^*\right) = \left(\overline{P}-I\right)\beta\beta'(P-I)$

Johnston (1972 : 168-169) กล่าวถึงการละทิ้งตัวแปรที่เกี่ยวข้องว่า ถ้ามีการละทิ้งตัวแปรที่เกี่ยวข้องออกไปจากแบบจำลองอย่างน้อย 1 ตัว เช่น X_k จะทำให้การประมาณค่าพารามิเตอร์ตามวิธี OLS จะเป็น

$$b=\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}y \tag{2.2}$$

และจะได้เป็น

$$b=\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}X\beta+\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}u$$

ดังนั้นจะเกิดการเอนเอียงคือ

$$E(b)-\beta=p_k\beta_k \tag{2.3}$$

เมื่อ $p=\left(X^{*\prime}X^{*}\right)^{-1}X^{*\prime}X_k$ เป็นสัมประสิทธิ์ของการถดถอย X_k กับตัวแปรที่เกี่ยวข้องตัวอื่นที่อยู่ในแบบจำลอง ซึ่งผลของความเอนเอียงจะขึ้นอยู่กับค่าสหสัมพันธ์ระหว่างตัวแปรที่อยู่ในแบบจำลองกับตัวแปรที่ถูกละทิ้งไป และยังขึ้นอยู่กับค่า β_k ด้วย

จากผลที่ได้จะเห็นว่าการละทิ้งตัวแปรที่เกี่ยวข้องจะทำให้เกิดความเอนเอียงในการประมาณค่าพารามิเตอร์ และก็ยังไม่มีความแม่นยำด้วย เนื่องจากความแปรปรวนของส่วนที่เหลือจะมีความเอนเอียงไปในทางที่เพิ่มขึ้น

Greene (2000 : 334-337) กล่าวถึงการละทิ้งตัวแปรที่เกี่ยวข้องว่า ถ้าสมมุติให้แบบจำลองที่ถูกต้องคือ

$$y=X_1\beta_1+X_2\beta_2+\varepsilon$$

แต่ถ้ามีการละทิ้งตัวแปร X_2 จะทำให้การประมาณค่าพารามิเตอร์เป็น

$$b_1=\beta_1+\left(X_1'X_1\right)^{-1}X_1'X_2\beta_2+\left(X_1'X_1\right)^{-1}X'\varepsilon$$

เมื่อใส่ค่าคาดหวังจะได้เป็น

$$E\left(b_1\right)=\beta_1+P_{1,2}\beta_2 \tag{2.4}$$

เมื่อ $P_{1,2}=\left(X_1'X_1\right)^{-1}X_1'X_2$ เป็นคอลัมพ์ของความชัน (slopes) ของการถดถอยกำลัง 2 น้อยที่สุดของคอลัมพ์ X_2 กับคอลัมพ์ X_1 และจะเห็นได้ว่าเกิดความเอนเอียงขึ้น ยกเว้นในกรณีที่ $X_1'X_2=0$ หรือ $\beta_2=0$ จะไม่เกิดความเอนเอียง

ค่าความแปรปรวนของพารามิเตอร์ b_1 จะมีค่าเป็น $\operatorname{var}(b_1)=\frac{\sigma^2}{s_{11}}$ $;s_{11}=\sum(x_{i1}-\bar{x}_1)^2$ แต่ถ้ามีการปรับการถดถอยให้ถูกต้อง (คือมีการใส่ X_2 เข้าไป) แล้วความชันของ X_1 จะไม่เอนเอียง จะทำให้ความแปรปรวนของ b_1 ที่ปรับแล้วกลายเป็น

$$\operatorname{var}(b_{1.2})=\sigma^2\left[X_1'X_1-X_1'X_2\left(X_2'X_2\right)^{-1}X_2'X_1\right]^{-1}$$

$$=\frac{\sigma^2}{s_{11}\left(1-r_{12}^2\right)} \qquad (2.5)$$

เมื่อ r_{12} เป็นค่าสหสัมพันธ์ของ X_1 กับ X_2 ซึ่งการที่ค่า r_{12} มีค่ามากจะทำให้ $\operatorname{var}(b_{1.2})$ มากกว่า $\operatorname{var}(b_1)$ โดยเปรียบเทียบ ดังนั้นแล้วมีความเป็นไปได้ที่ b_1 จะเป็นตัวประมาณค่าที่อยู่บนพื้นฐานของ MSE (means-squared-error) จะเห็นได้ว่ามีความเป็นไปได้ที่ b_1 จะมีความแน่นอนมากกว่า $b_{1.2}$

ส่วนการประมาณค่า σ^2 จะหาได้จาก $s^2=\frac{e_1'e_1}{n-k_1}$ เมื่อใส่ค่าคาดหวังใน $e'e$ และให้ $E\left(X_1'\varepsilon\right)=0$ แล้วจะได้

$$E\left(e_1'e_1\right)=\beta_2'X_2'M_1X_2\beta_2+\sigma^2\left(n-k_1\right) \qquad (2.6)$$

ซึ่งในเทอมแรกจะแสดงถึงการเพิ่มขึ้นในผลรวมของส่วนที่เหลือยกกำลัง 2 (residual sum of squares) ที่จะมีค่าเป็นบวก นั้นคือ s^2 จะมีค่าเพิ่มขึ้นถึงแม้ว่า $X_1'X_2=0$ (orthogonal) ก็ตาม

Kmenta (1971 : 391-395) กล่าวว่า ในการพิจารณาการเกิดความคลาดเคลื่อนจากการสร้างแบบจำลอง อันเนื่องมาจากการละทิ้งตัวแปรที่เกี่ยวข้อง สามารถพิจารณาได้ดังนี้

สมมุติให้สมการการถดถอยที่ถูกต้องเป็น

$$Y_i=\beta_1+\beta_2X_{i2}+\beta_3X_{i3}+\varepsilon_i \qquad (2.7)$$

แต่ว่ากลับไปประมาณค่าในสมการ

$$Y_i=\beta_1+\beta_2X_{i2}+\varepsilon^*_i \qquad (2.8)$$

ดังนั้นการประมาณค่า $\hat{\beta}_2$ คือ

$$E\left(\hat{\beta}_2\right)=E\left[\frac{\sum\left(X_{i2}-\bar{X}_2\right)\left(Y_i-\bar{Y}\right)}{\sum\left(X_{i2}-\bar{X}_2\right)^2}\right]$$

แต่รู้ว่า $\left(Y_i-\bar{Y}\right)=\beta_2\left(X_{i2}-\bar{X}_2\right)+\beta_3\left(X_{i3}-\bar{X}_3\right)+\left(\varepsilon_i-\bar{\varepsilon}\right)$ ดังนั้น

$$E\left(\hat{\beta}_2\right)=\beta_2+\beta_3 d_{32} \qquad (2.9)$$

เมื่อ

$$d_{32}=\frac{\sum\left(X_{i2}-\bar{X}_2\right)\left(X_{i3}-\bar{X}_3\right)}{\sum\left(X_{i2}-\bar{X}_2\right)^2}$$

และ

$$E\left(\hat{\beta}_1\right)=\beta_1+\beta_3 d_{31} \qquad (2.10)$$

เมื่อ $d_{31}=\bar{X}_3-d_{32}\bar{X}_2$

ซึ่ง d_{31} และ d_{32} จะอยู่ในรูปของ $X_{i3}=d_{31}+d_{32}X_{i2}+residual$ แสดงว่าเป็นค่าประมาณพารามิเตอร์ของการถดถอยตัวแปรที่ถูกละทิ้งไปกับตัวแปรที่อยู่ในแบบจำลอง

จากผลที่ได้นี้แสดงว่า $\hat{\beta}_2$ จะเกิดความเอนเอียงขึ้นมา นอกจากว่าค่า d_{32} จะมีค่าเป็น 0 (X_{i2} กับ X_{i3} ไม่มีความสัมพันธ์กัน) และถ้า β_3 กับ d_{32} มีเครื่องหมายเหมือนกัน แล้วความเอนเอียงใน $\hat{\beta}_2$ จะมีค่าเป็นบวก ถ้าเครื่องหมายตรงกันข้ามก็จะมีค่าเป็นลบ ส่วน $\hat{\beta}_1$ จะยังคงเอนเอียงอยู่ถึงแม้ว่า d_{32} จะเท่ากับ 0 แต่ถ้า $\bar{X}_3$ เท่ากับ 0 ด้วยแล้ว ค่า $\hat{\beta}_1$ จะไม่เอนเอียง

ถ้าค่าสหสัมพันธ์ระหว่าง X_{i2} กับ X_{i3} ไม่หายไปเมื่อตัวอย่างมีขนาดใหญ่ขึ้น ($\lim_{n\to\infty} d_{32}\neq 0$) แล้ว $\hat{\beta}_2$ จะเกิดความไม่แนบนัย (inconsistent) เช่นเดียวกันกับ $\hat{\beta}_1$ (เมื่อ $\lim_{n\to\infty}\left(\bar{X}_3-d_{32}\bar{X}_2\right)\neq 0$)

ส่วนการประมาณค่า $\operatorname{var}\left(\hat{\beta}_2\right)$ ในสมการที่ 2.8 คือ

$$\operatorname{var}\left(\hat{\beta}_2\right)=\frac{\sigma^2}{\sum\left(X_{i2}-\bar{X}_2\right)^2}$$

แต่การประมาณค่า $\operatorname{var}\left(\hat{\beta}_2\right)$ ในสมการที่ 2.7 คือ

$$s_{\hat{\beta}_2}^2=\frac{s^2}{\sum\left(X_{i2}-\bar{X}_2\right)^2}=\frac{\sum\left[\left(Y_i-\bar{Y}\right)-\hat{\beta}_2\left(X_{i2}-\bar{X}_2\right)\right]^2/(n-2)}{\sum\left(X_{i2}-\bar{X}_2\right)^2}$$

ดังนั้นค่าคาดหวังของ $s_{\hat{\beta}_2}^2$ คือ

$$E\left(s_{\hat{\beta}_2}^2\right)=\operatorname{var}\left(\hat{\beta}_2\right)+\frac{\beta_3^2\sum\left(X_{i3}-\bar{X}_3\right)^2}{(n-2)\sum\left(X_{i2}-\bar{X}_2\right)^2} \qquad (2.11)$$

ซึ่งแสดงว่า $\operatorname{var}\left(\hat{\beta}_2\right)$ ในสมาการที่ 2.8 จะมีการประมาณที่เอนเอียงที่เป็นลบ ซึ่งจะทำให้เกิดการลงความเห็นที่ผิดพลาดในการทดสอบสมมุติฐานได้

Gujarati (1995 : 203-207) ได้แสดงให้เห็นถึงผลของการละทิ้งตัวแปรจากการที่ได้ศึกษาการประมาณค่าเส้น Expectations-Augmented Phillips ของประเทศสหรัฐอเมริการะหว่างปี 1970-1982 ซึ่งจะได้สมการที่ถูกต้องสำหรับเส้น Expectations-Augmented Phillips คือ

$$\hat{Y}_t = 7.1933 - 1.3925X_{2t} + 1.4700X_{3t} : R^2=0.8766$$
$$(1.5948) \quad (0.3050) \quad (0.1758)$$

เมื่อ Y_t คือ อัตราเงินเฟ้อที่แท้จริง ณ เวลาที่ t

X_{2t} คือ อัตราการว่างงาน ณ เวลาที่ t

X_{3t} คือ อัตราเงินเฟ้อที่คาดหวัง ณ เวลาที่ t

(...) ตัวเลขในวงเล็บคือค่าเบี่ยงเบนมาตรฐานของพารามิเตอร์

ซึ่งตามทฤษฎีทางเศรษฐศาสตร์มหภาค ค่าสัมประสิทธิ์ของ X_{2t} จะต้องเป็นลบ และค่าสัมประสิทธิ์ของ X_{3t} จะต้องเป็นบวก แต่ถ้าหากว่าการประมาณค่าเส้น Expectations-Augmented Phillipsได้มีการละทิ้งตัวแปร X_{3t} ออกไปจากสมการแล้ว จะทำให้เกิดความเอนเอียงของการประมาณค่าพารามิเตอร์ โดยผลที่ได้คือ

$$\hat{Y}_t = 6.1272 + 0.2448X_{2t} : R^2=0.0135$$
$$(4.2853) \quad (0.6304)$$

จะเห็นได้ว่าค่าสัมประสิทธิ์ที่ประมาณค่ามาได้นั้นเกิดความเอนเอียง ไม่มีนัยสำคัญทางสถิติ และสัมประสิทธิ์ของ X_{2t} ยังมีเครื่องหมายที่ผิดไปจากทฤษฎีทางเศรษฐศาสตร์มหภาคด้วย Gujarati ได้กล่าวว่าความเอนเอียงที่เกิดขึ้นนั้นจะมาจาก

$$E(\widehat{\beta}_2{}^*) = \beta_2 + \beta_3 d_{32}$$

เมื่อ $\widehat{\beta}_2{}^*$ คือ สัมประสิทธิ์ที่ประมาณค่ามาได้ของตัวแปร X_{2t} จากแบบจำลองที่ผิด

β_2, β_3 คือ สัมประสิทธิ์ของตัวแปร X_{2t} และ X_{3t} จากแบบจำลองที่ถูกต้อง

d_{32} คือ สัมประสิทธิ์ที่ได้จากการถดถอย X_{3t} กับตัวแปร X_{2t}

จะเห็นได้ว่าถ้า $\beta_3 d_{32}$ ไม่เท่ากับ 0 ก็จะเกิดความเอนเอียงขึ้น โดยที่ถ้า $\beta_3 d_{32}$ มีค่าเป็นบวกแสดงว่ามีความเอนเอียงในทางบวก แต่ถ้า $\beta_3 d_{32}$ มีค่าน้อยกว่า 0 แสดงว่ามีความเอนเอียงในทางที่ลดลง และนอกจากความเอนเอียงที่เกิดขึ้นแล้ว ค่าพารามิเตอร์ที่ประมาณค่ามาได้นั้นก็จะไม่มีความแนบนัย (inconsistent) ด้วย ในกรณีที่ d_{32} =0 หรือ X_{2t} ไม่มีความสัมพันธ์กับ X_{3t} (r_{23}=0) แล้ว การประมาณค่า $\widehat{\beta}_2^*$ จะไม่เอนเอียงแต่ $\widehat{\beta}_1^*$ จะยังคงเอนเอียงอยู่

นั้นคือ ผลของตัวแปร X_{2t} ที่มีต่อ Y ทั้งหมดเมื่อมีการละทิ้งตัวแปร X_{3t} คือผลทางตรงของ X_{2t} ที่มีต่อ Y (β_2) รวมกับ ผลทางอ้อมของ X_{2t} ที่มีต่อ Y ($\beta_3 d_{32}$) โดยผลทางอ้อมนี้จะเป็นความเอนเอียงที่เกิดขึ้นกับตัวแปร X_{2t} ด้วย

$$\begin{aligned} &\textit{Gross effect of } X_2 \textit{ on } Y (= \widehat{\beta}_2^*) \\ &\quad = \textit{direct effect of } X_2 \textit{ on } Y (= \beta_2) \\ &\quad\quad + \textit{indirect effect of } X_2 \textit{ on } Y (= \beta_3 d_{32}) \end{aligned} \tag{2.12}$$

ส่วนค่า d_{32} หาได้จากการถดถอย X_{3t} กับตัวแปร X_{2t} ที่เรียกว่าเป็นสมการการถดถอยช่วย จะได้สมการคือ

$$X_{3t} = 0.7252 + 1.1138 X_{2t}$$

ดังนั้น d_{32} คือ 1.1138 หมายความว่าถ้า X_{2t} เพิ่มขึ้น 1 หน่วยจะทำให้ X_{3t} เพิ่มขึ้นประมาณ 1.11 หน่วย และการเพิ่มขึ้นของ X_{3t} จะไปทำให้มีผลต่อค่า Y คือ (1.47)(1.1138)= $\beta_3 d_{32}$ =1.6373 หรือ X_{2t} มีความเอนเอียงไปจากค่าที่แท้จริงอยู่ 1.6373

ส่วนการประมาณความแปรปรวนของพารามิเตอร์ก็จะเกิดความเอนเอียงด้วย โดยที่ $\operatorname{var}\left(\widehat{\beta}_2^*\right)$ คือ

$$\operatorname{var}\left(\widehat{\beta}_2^*\right) = \frac{\sigma^2}{\sum x_{2i}^2} \tag{2.13}$$

แต่ $\operatorname{var}\left(\widehat{\beta}_2\right)$ เป็น

$$\operatorname{var}\left(\hat{\beta}_2\right)=\frac{\sigma^2}{\sum x_{2i}^2\left(1-r_{23}^2\right)} \tag{2.14}$$

จะเห็นได้ว่า แม้ว่า $\hat{b}_2$ จะเป็นตัวประมาณค่าที่มีความเอนเอียง แต่ว่าจะมีความแปรปรวนน้อยกว่าตัวประมาณค่าที่แท้จริง ถ้า σ^2 ของทั้ง 2 แบบจำลองไม่ต่างกันมากนัก โดยมีความเป็นไปได้ที่ σ^2 ของแบบจำลองที่มีการละทิ้งตัวแปรจะมีขนาดใหญ่กว่า

Pindyck and Rubinfeld (1998 : 184-187) กล่าวถึงการละทิ้งตัวแปรที่เกี่ยวข้องออกไปจากแบบจำลองว่า สมมุติแบบจำลองที่แท้จริงในรูปผลต่างของค่าเฉลี่ยของตัวเอง (deviations form) เป็น $y_i=\beta_2x_{2i}+\beta_3x_{3i}+\varepsilon_i$ ถ้ามีการละทิ้งตัวแปร X_3 ออกไปแล้ว แบบจำลองที่จะทำการประมาณค่าคือ $y_i=\beta_2^*x_{2i}+\varepsilon_i^*$ ดังนั้นแล้วการประมาณค่าพารามิเตอร์คือ

$$\hat{\beta}_2^*=\frac{\sum x_{2i}y_i}{\sum x_{2i}^2} \tag{2.15}$$

เมื่อแทนค่า y_i จากแบบจำลองที่แท้จริงก็จะได้เป็น

$$\hat{\beta}_2^*=\beta_2+\beta_3\frac{\sum x_{2i}x_{3i}}{\sum x_{2i}^2}+\frac{\sum x_{2i}\varepsilon_i}{\sum x_{2i}^2}$$

ในรูปของค่าคาดหวังคือ

$$E\left(\hat{\beta}_2^*\right)=\beta_2+\beta_3\frac{\sum x_{2i}x_{3i}}{\sum x_{2i}^2}\beta_2+\beta_3\frac{\operatorname{cov}(x_2,x_3)}{\operatorname{var}(x_2)} \tag{2.16}$$

จากผลที่ได้แสดงว่าการประมาณค่าพารามิเตอร์จะเกิดการเอนเอียงขึ้น และการเอนเอียงจะไม่หายไปเมื่อขนาดตัวอย่างเพิ่มขึ้น นั้นคือเป็นตัวพารามิเตอร์ที่ไม่แนบนัย (inconsistent) แต่ถ้า $\operatorname{cov}(x_2,x_3)$ มีค่าเท่ากับ 0 แล้ว ค่าพารามิเตอร์ที่ถูกประมาณก็จะไม่เกิดการเอนเอียง ซึ่งการเอนเอียงในทางที่เพิ่มขึ้น หรือลดลงนั้นก็จะขึ้นอยู่กับทิศทางของค่าสหสัมพันธ์ (r) ของตัวแปรที่ถูกละ

ทิ้งกับตัวแปรที่อยู่ในแบบจำลอง กับทิศทางของค่าความชัน (slope) ของตัวแปรที่ถูกละทิ้ง ยิ่งค่า r ระหว่าง X_2 กับ X_3 ก็จะแสดงถึงการที่สัมประสิทธิ์ของ X_2 จะรวมผลกระทบของตัวแปร X_3 ด้วย

ส่วนการประมาณค่าความแปรปรวนของพารามิเตอร์นั้น $\mathrm{var}(\hat{\beta}_2^*)$ จะมีความเอนเอียงไปจากค่า $\mathrm{var}(\hat{\beta}_2)$ โดยที่ $\mathrm{var}(\hat{\beta}_2^*)$ จะมีค่าน้อยกว่า $\mathrm{var}(\hat{\beta}_2)$

Studenmund (1992 : 177-180) กล่าวถึงการละทิ้งตัวแปรที่เกี่ยวข้องออกไปจากแบบจำลองว่า สมมุติให้ $Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i^*$ เป็นแบบจำลองที่ได้มีการละทิ้งตัวแปรที่เกี่ยวข้อง X_{2i} ออกไป ดังนั้นแล้ว $\varepsilon_i^* = \beta_2 X_{2i} + \varepsilon_i$ จะเห็นได้ว่า ε_i^* จะไม่ได้เป็นอิสระกับ X_{2i} ซึ่งถ้า X_{2i} มีความสัมพันธ์กันกับ X_{1i} แล้ว การเปลี่ยนแปลงใน X_{2i} ก็จะไปทำให้ X_{1i} และ ε_i^* เปลี่ยนแปลงไปด้วย ซึ่งความสัมพันธ์ของ X_1 กับ X_2 สามารถวัดได้จากสหสัมพันธ์อย่างง่าย (simple correlation coefficient : r_{12}) ดังนั้นการประมาณค่าแบบ OLS ก็จะไม่เป็น Best Linear Unbiased Estimator (BLUE) ทำให้

$$E\left(\hat{\beta}_1\right) = \beta_1 + \beta_2 . f(r_{12}) \qquad (2.17)$$

เกิดความเอนเอียงขึ้นมา ซึ่ง $\hat{\beta}_1$ ที่ได้มาจาก OLS นี้ จะรวมผลกระทบของ X_1 ต่อ Y และผลกระทบของ X_2 ต่อ Y ด้วย แต่ถ้า X_1 กับ X_2 ไม่มีความสัมพันธ์ต่อกัน แล้วการประมาณค่า $\hat{\beta}_1$ จะยังคงไม่เอนเอียง เพราะค่า $\hat{\beta}_1$ ที่ประมาณมาได้นั้นจะไม่รวมผลกระทบของ X_2 ต่อ Y ส่วน $f(r_{12})$ เป็นฟังก์ชั่นของ r_{12} ซึ่งก็คือสัมประสิทธิ์การถดถอย X_2 บน X_1 นั้นเอง

2.2 ปริทัศน์ผลงานวิจัยต่าง ๆ ที่เกี่ยวข้อง

Grilliches (1957) ได้ศึกษาความเอนเอียงที่เกิดขึ้นในการประมาณค่าของฟังก์ชั่นการผลิต โดยได้พยายามที่จะอธิบายวิธีการของการสืบหาผลของการเอนเอียงที่เกิดขึ้นจากการประนีประนอมของการใช้ตัวแปรที่ดีที่สุดรองลงมา (the second best) หรือการละทิ้งตัวแปรบางตัวออกไป หรือยอมรับการประมาณค่าใกล้เคียง หรือข้อมูลแบบผลรวม (aggregate) ซึ่งถ้ารู้ผลของการประนีประนอมที่จะเกิดขึ้นเหล่านั้น ก็อาจจะสามารถอธิบายผลการวิเคราะห์ได้ดีขึ้นก็ได้ ในการประมาณค่าฟังก์ชันการผลิตแบบ Cobb-Douglas ได้ใช้นิยามของผลตอบแทนต่อขนาด (returns to scale) เท่ากับผลบวกของสัมประสิทธิ์ที่อยู่ข้างบนปัจจัยการผลิตในฟังก์ชันการผลิต

แบบ Cobb-Douglas ซึ่งผลบวกนี้ใช้สัญลักษณ์ $e = \sum \beta_i$ โดยจะเรียกว่า "ความยืดหยุ่นของการผลิต (elasticity of production)" หรือ "สัมประสิทธิ์แห่งฟังก์ชัน (function coefficient)" ซึ่งผลที่เกิดขึ้น

สามารถที่จะแสดงให้เห็นว่าส่วนที่เอนเอียง (biased) ในค่า ประมาณการของผลตอบแทนต่อขนาดจะมีค่าเท่ากับ

$$E(\hat{e}-e) = E\left(\sum_{i=1}^{k-1} b_i - \sum_{i=1}^{k} \beta_i\right)$$

$$= \beta_k\left(\sum_{i=1}^{k-1} P_{ik} - 1\right) \tag{2.18}$$

นั่นคือ การที่เราประมาณค่าน้อยไปกว่าความเป็นจริงหรือมากกว่าความเป็นจริงขึ้นอยู่กับผลบวกของสัมประสิทธิ์จาก การถดถอยช่วย (auxiliary regression) เราจะประมาณค่า e ต่ำกว่าความเป็นจริงถ้าผลบวกดังกล่าวมีค่าน้อยกว่าหนึ่ง สูงกว่าความเป็นจริงถ้าผลบวกมีค่ามากกว่าหนึ่ง และจะไม่มีส่วนที่เอนเอียงเลยถ้าผลบวกเท่ากับหนึ่งพอดี นั่นคือ

ถ้า	$\sum_{i=1}^{k-1} P_{ik}$	> 1	ดังนั้น	biased	> 0
		$= 1$		biased	$= 0$
		< 1		biased	< 0

การอธิบายอย่างง่ายๆ เกี่ยวกับผลบวกของสัมประสิทธิ์นี้เป็นไปได้ โดยอาศัยการถดถอยช่วยและถ้าเป็นกรณี Cobb-Douglas สัมประสิทธิ์เหล่านี้ก็จะเป็น

$$X_k = X_1^{p1} X_2^{p2} \ldots X_{k-1}^{pk-1} e^{v} \tag{2.19}$$

ผลบวกของ p's ในกรณีนี้ก็มีค่าเท่ากับ "ความยืดหยุ่นของการผลิต (elasticity of production)" หรือผลตอบแทนต่อขนาด (returns to scale) ในสมการถดถอยช่วย ส่วนที่เอนเอียงในค่าประมาณของผลตอบแทนต่อขนาดจะขึ้นอยู่กับว่า X_k ได้เปลี่ยนแปลงไปอย่างไร ในข้อมูลตัวอย่างของเราเมื่อ X's อื่นๆ ทุกตัวได้เปลี่ยนไปกับขนาดการผลิต จะประมาณค่า e ต่ำไป ถ้าการเปลี่ยนแปลงของอัตราส่วนในปัจจัยที่อยู่ในแบบจำลองนั้นมีความเกี่ยวพันกันกับตัวแปรที่เราละ

ทิ้งไป ซึ่งเปลี่ยนแปลงในอัตราส่วนที่น้อยกว่าในข้อมูลตัวอย่างของเรา จะไม่มีการเอนเอียง ถ้าตัวแปรที่เราละทิ้งไปนั้นโดยเฉลี่ยได้เปลี่ยนแปลง ในอัตราส่วนเดียวกันกับการเปลี่ยนแปลงในตัวแปรที่อยู่ในแบบจำลองทั้งหมด และจะประมาณค่าผลตอบแทนต่อขนาดมากกว่าความเป็นจริง ถ้าการเปลี่ยนแปลงของตัวแปรที่ละทิ้งไปมากกว่าการเปลี่ยนแปลงของอัตราส่วนของตัวแปรที่อยู่ในแบบจำลอง

การละทิ้งตัวแปรทางด้านบริหารของผู้ประกอบการหรือการจัดการ จะพบว่าเกษตรกรที่มีการผลิตเป็นสองเท่าของเพื่อนบ้าน ก็จะใช้ความเป็นผู้ประกอบการเป็นสองเท่า ถ้าข้อสมมุตินี้ถูกต้อง ผลบวกของสัมประสิทธิ์ในสมการช่วยจะมีค่าน้อยกว่าหนึ่ง และเราก็จะได้รับค่าประมาณของผลตอบแทนต่อขนาดต่ำกว่าความเป็นจริง

มีความเป็นไปได้ในธุรกิจที่มีระดับของการประกอบการสูง หรือการจัดการสูง ก็มักจะมีการใช้ต้นทุนสูงด้วย นั้นคืออาจจะคาดการณ์ได้ว่า มีความสัมพันธ์ในทางบวกระหว่าง การจัดการในปัจจัยการผลิตกับระดับของปัจจัยทุน ถ้าเป็นเช่นนั้นสัมประสิทธิ์ของปัจจัยทุนในการถดถอยช่วยที่มีการจัดการเป็นตัวแปรตาม และปัจจัยการผลิตอื่นๆ เป็นตัวแปรอิสระจะมีเครื่องหมายเป็นบวก ซึ่งจะทำให้ค่าประมาณของความยืดหยุ่นของผลผลิตเมื่อคำนึงถึงปัจจัยทุน มีค่าสูงกว่าความเป็นจริง ซึ่งก็จะทำให้ เราประมาณค่าผลตอบแทนต่อปัจจัยทุนสูงเกินความจริง

ส่วนการไม่คำนึงถึงความแตกต่างทางด้านคุณภาพ ก็เป็นอีกกรณีหนึ่งที่จะทำให้เกิดความเอนเอียง (biased) ในการประมาณค่าได้ สมมุติให้ q_iL_i เป็นหน่วยวัดปัจจัยแรงงานที่แท้จริงโดยที่ L_i คือ ปริมาณเวลา และ q_i คือ ตัวคูณที่เกี่ยวข้องกับปัจจัยแรงงานที่ i ซึ่งจะเป็นตัวแปรค่าหรือถ่วงน้ำหนักเพื่อให้ได้ "หน่วยของแรงงานที่มีลักษณะเท่ากัน" ตัวที่ q_i นี้อาจจะมีค่าแตกต่างกันไปตามแต่ละคนหรือแต่ละฟาร์ม ซึ่งแทนที่เราจะใช้ q_iL_i ในการประมาณค่าฟังก์ชันการผลิตสมมุติว่าเราใช้แต่เพียง L_i เป็นการวัดแรงงาน ดังนั้นผลที่ได้คือ

แรงงานที่มีคุณภาพสูงจะเกี่ยวข้องกับ การใช้ปัจจัยทุนสูงตามไปด้วย เพราะว่าจะเป็นการเพิ่มผลิตภาพส่วนเพิ่มของทุน และเนื่องจากเกษตรกรที่มีคุณภาพของแรงงานอยู่ในระดับที่สูงกว่าก็จะมีข้อจำกัดเกี่ยวกับการควบคุมทุนในระดับที่น้อยกว่า และเป็นไปได้ว่าคุณภาพนั้นเป็นสิ่งทดแทนปริมาณได้ และฟาร์มที่มีคุณภาพของแรงงานที่สูงกว่าจะใช้แรงงานน้อยกว่า จะทำให้ประมาณค่าตอบแทนต่อทุนสูงไปกว่าความเป็นจริง และผลตอบแทนต่อแรงงานน้อยไปกว่าความเป็นจริง ดังนั้นอาจเป็นคำอธิบายว่า ในการศึกษาเรื่องฟังก์ชันการผลิตเรามักจะได้ผลลัพธ์ออกมาว่า ผลตอบแทนแรงงานนั้นต่ำอย่างไม่มีเหตุผล และผลตอบแทนต้นทุนก็สูงมากอย่างไม่มีเหตุผล

Mark Roger และ Yi-Ping Tseng (2000) ได้ศึกษาการวิเคราะห์ผลิตภาพของระดับแรงงานในหน่วยธุรกิจ โดยใช้การสำรวจข้อมูล ที่มีการแบ่งชั้นของข้อมูลที่อยู่บนพื้นฐานของชนิดของอุตสาหกรรมและขนาดของธุรกิจ โดยจะมีการพิจารณาว่าควรมีการถ่วงน้ำหนักในการการถดถอย ซึ่งได้มีการขยายความในการประมาณค่าของฟังก์ชั่นการผลิตแบบ Cobb-Douglas ที่จะมีการใส่ตัวแปรหุ่น (dummies) เข้าไปด้วย เมื่อให้ตัวแปรหุ่นเป็นตัวแทนของระดับความรู้ที่สำคัญที่มีอยู่แล้วของหน่วยธุรกิจนั้นคือ

$$y = AK^{\lambda}C^{\alpha}L^{\beta}$$

เมื่อ Y คือ ค่าที่ถูกเพิ่ม C คือ ต้นทุนทางกายภาพ L คือ แรงงาน และ K คือ ระดับความรู้ของทุน A เป็นค่าคงที่ ซึ่งค่า K จะแสดงถึงความรู้ความสามารถที่มี เทคนิคในการจัดการ และต้นทุนมนุษย์ (เช่น การศึกษา,การฝึกฝน,ประสบการณ์การทำงาน ซึ่งโดยทั่วไปแล้วข้อมูลเหล่านี้ยากที่จะหาออกมาได้ ดังนั้นจึงมีการใช้ตัวแปรหุ่นมาแทนในค่า K ดังนั้นแล้วก็จะได้สมการคือ

$$\ln\frac{V}{L} = \ln A + \alpha \ln\left[\frac{C}{L}\right] + \beta_i D_i \qquad (2.20)$$

เมื่อ D แทนตัวแปรหุ่นแสดงถึงองค์ความรู้ที่มีอยู่ในระดับต่างๆ ในส่วนการประมาณค่าพารามิเตอร์ จะพิจารณาว่าควรจะมีการใช้ WLS (weighted least square) หรือ OLS ในการประมาณค่า โดยจะมีการทดสอบที่ $\hat{\beta}_{ols} = \hat{\beta}_{wls}$ หรือไม่ ถ้าปฏิเสธสมมุติฐานก็จะแสดงว่า ค่าพารามิเตอร์ที่ประมาณค่ามาได้นั้นจะมีความแตกต่างกันในแต่ละชั้น หรืออาจมีการละทิ้งตัวแปรที่เกี่ยวข้องออกไปจากแบบจำลอง ถ้าเป็นแบบนี้ก็จะแสดงว่ามีความสัมพันธ์เกิดขึ้นในระหว่างค่าความคลาดเคลื่อนกับค่าตัวแปรอธิบาย และระดับชั้น

ผลลัพธ์ที่ได้จากการถดถอยแสดงให้เห็นว่า จะมีความแตกต่างอย่างมีนัยสำคัญระหว่างพารามิเตอร์ที่ประมาณค่ามาได้จาก WLS (weighted least square) กับพารามิเตอร์ที่ประมาณค่ามาได้จาก OLS โดยเฉพาะในตัวแปรที่แสดงถึงความหนาแน่นของการรวมกลุ่ม ที่ WLS ได้แสดงถึงการมีความสัมพันธ์กันในทางบวกกับค่าผลิตภาพของแรงงาน ส่วนการใช้ OLS ไม่ได้แสดงถึงการมีความสัมพันธ์ แต่ OLS ได้แสดงถึงค่าใช้จ่ายในการฝึกฝนต่างๆ จะเป็นปัจจัยที่สำคัญกว่าเมื่อเทียบกับการใช้ WLS

จากผลที่ต่างกันนี้จะแสดงถึงการคงอยู่ของตัวแปรที่ยังไม่ได้สำรวจที่มีความสัมพันธ์กันกับระดับชั้น ซึ่งความแตกต่างเหล่านี้จะเป็นสาเหตุหนึ่งในความแตกต่างของพารามิเตอร์ในแต่ละชั้น

ในความหมายทางทฤษฎีแล้ว จะหมายถึงการใช้ฟังก์ชั่นการผลิตอย่างเดียวกันในแต่ละหน่วยธุรกิจนั้นจะไม่มีความเหมาะสม ดังนั้นแล้วจึงควรที่จะมีการประมาณค่าพารามิเตอร์ในแต่ละส่วนหรือแต่ละชั้นของกลุ่มตัวอย่าง

Chow (1966 : 111-131) กับ Taylor และ Newhouse (1969 : 851-856) ได้ทำการศึกษาเกี่ยวกับ อุปสงค์ของเงินในระยะสั้นและระยะยาว ซึ่ง Chow ได้ทำการประมาณสมการของเงินในระยะยาวคือ

$$\widehat{M}_t = 0.1365 + 1.069Y_{pt} - 0.01321Y_t - 0.7476R_t : R^2 = 0.9965$$

เมื่อ M คือ ค่าลอกการิทึ่มธรรมชาติของปริมาณเงินทั้งหมด

Y_{pt} คือ ค่าลอกการิทึ่มธรรมชาติของรายได้ถาวร

Y_t คือ ค่าลอกการิทึ่มธรรมชาติของรายได้ปัจจุบัน

R_t คือ ค่าลอกการิทึ่มธรรมชาติของอัตราดอกเบี้ย

Chow ได้สรุปว่ารายได้ถาวรจะมีความสำคัญมากกว่ารายได้ในปัจจุบันเมื่อคำนึงถึงข้อจำกัดของทรัพย์สินของบุคคลนั้นในระยะยาว และตัวแปรของรายได้ในปัจจุบันจะไม่มีนัยสำคัญทางสถิติ แต่อย่างไรก็ตาม L. D. Taylor และ J. P. Newhouse ก็ได้แย้งเกี่ยวกับการประมาณเส้นอุปสงค์เงินของ Chow ว่า เกิดปัญหาการระบุแบบจำลองที่ผิดพลาดมีการประมาณพารามิเตอร์ที่เอนเอียง ซึ่งได้ประมาณเส้นอุปสงค์เงินในระยะยาวขึ้นมาใหม่คือ

$$\widehat{M}_t = 0.3067 + 0.06158Y_{pt} + 0.3274Y_t - 0.3325R_t + 0.5878M_{t-1} \;: R^2 = 0.9988$$

ซึ่งความเอนเอียงที่เกิดขึ้นของการประมาณค่าพารามิเตอร์ของตัวแปรต่างๆ จะเท่ากับค่าสัมประสิทธิ์ของการถดถอยช่วยระหว่าง M_{t-1} กับตัวแปร Y_{pt}, Y_t, และ R_t คูณกับค่าสัมประสิทธิ์ของตัวแปร M_{t-1} และ Taylor กับ Newhouse ได้สรุปว่า ตัวแปร M_{t-1} จะมีเครื่องหมายเป็นบวกและมีนัยสำคัญทางสถิติ ในขณะที่ Y_{pt} ไม่มีนัยสำคัญทางสถิติส่วนตัวแปร Y_t กับมีนัยสำคัญทางสถิติเป็นอย่างมากต่อเส้นอุปสงค์เงินในระยะยาว ดังนั้นแล้วถ้าหากว่าตัวแปร M_{t-1} ไม่สามารถที่จะหาได้ก็จะแสดงค่าเกิดความเอนเอียงของการประมาณค่าพารามิเตอร์ ซึ่งจะทำให้มีการลงความเห็นทางนโยบายที่ผิดพลาดได้

จากที่กล่าวมาทั้งหมดนี้จะเห็นได้ว่า เมื่อมีความจำเป็นต้องละทิ้งตัวแปรที่เกี่ยวข้องออกไป จะทำให้มีการประมาณค่าพารามิเตอร์เกิดความเอนเอียง ดังนั้นถ้าทราบถึงขนาดและอัตราส่วนของการเอนเอียงว่ามีเป็นเท่าไร เมื่อตัวแปรที่อยู่ในแบบจำลองกับตัวแปรที่ตัดออกมีความสัมพันธ์กันในระดับต่างๆ ก็สามารถนำเอาผลลัพธ์ที่ได้มานั้น ไปพิจารณาค่าที่แท้จริงของพารามิเตอร์ในแบบจำลองที่จำเป็นต้องละทิ้งตัวแปรที่เกี่ยวข้อง และจะพิจารณาในค่าตัดแกน หรือค่าคงที่ (intercept term) ว่าความเอนเอียงที่เกิดขึ้นนั้นอาจจะมีค่าเท่ากับ $\beta_k \overline{X}_k$ หรือมีค่าอื่นใดอีกที่เป็นตัวกำหนดความเอนเอียงอีกในอัตราส่วนเท่าไรบ้าง และจะพิจารณาถึงค่าความแปรปรวนของความคลาดเคลื่อน และความแปรปรวนของค่าสัมประสิทธิ์ซึ่งอาจจะมีค่าเพิ่มขึ้นก็ได้ อันเนื่องมาจากการเกิด heteroscedasticity และ autocorrelation