การตรวจสอบความเหมาะสม ของสมการถดถอยแบบง่าย (Visual Test)

เอกรินทร์ ยลระบิล[1]

บทคัดย่อ

การวิเคราะห์การถดถอย (Regression Analysis) เป็นเครื่องมือเชิงคณิตศาสตร์ที่ใช้ช่วยสำหรับการพยากรณ์ที่สำคัญ เนื่องจากเป็นการศึกษาถึงความสัมพันธ์ของตัวแปรต่าง ๆอย่างมีระบบ โดยอาศัยการศึกษาถึงความสัมพันธ์ที่เป็นไปได้ของค่าตัวแปรที่เกี่ยวข้องจากข้อมูลในอดีตที่ผ่านมา และศึกษาความสัมพันธ์ของตัวแปรเหล่านั้นว่ามีรูปแบบความสัมพันธ์กันมากน้อยเพียงใดหรือไม่ โดยอาศัยเครื่องมือทางคณิตศาสตร์และสถิติศาสตร์ เข้ามาใช้เป็นเครื่องมือที่สำคัญในการประมาณการความสัมพันธ์ที่เกิดขึ้นมานั้น ๆ ทั้งนี้ในปัจจุบันมีการพัฒนาโปรแกรมคอมพิวเตอร์ทางด้านสถิติให้สามารถใช้ได้โดยง่าย อาทิเช่น SPSS MINITAB SAS รวมทั้งการเพิ่มฟังก์ชันทางสถิติลงในโปรแกรมที่ใช้ทั่วไป เช่น Microsoft Excel ให้สามารถนำการวิเคราะห์ทางสถิติไปใช้ได้อย่างสะดวกสำหรับการใช้งานทางธุรกิจ ด้วยเหตุนี้ การสร้างตัวแบบแสดงความสัมพันธ์ของตัวแปรที่สนใจ โดยการวิเคราะห์การถดถอยจึงง่ายแค่ปลายนิ้วคลิ๊ก ซึ่งจริง ๆแล้วการได้สมการแสดงความสัมพันธ์ดังกล่าวเป็นเพียงแค่จุดเริ่มต้นของการวิเคราะห์เท่านั้น ความง่ายและความสะดวกของการใช้โปรแกรมคอมพิวเตอร์มาช่วยในการสร้างสมการแสดงความสัมพันธ์ทำให้ผู้ใช้ละเลยที่จะศึกษาถึงวิธีการคิด วิธีคำนวณและตรรกวิทยาเบื้องหลังการได้มาซึ่งสมการดังกล่าว รวมไปถึงข้อสมมติฐานต่าง ๆของการใช้สมการดังกล่าว ทั้งนี้ก่อนการนำผลจากสมการถดถอยมาใช้ประโยชน์จะต้องมีการตรวจสอบความเหมาะสมของตัวแบบที่ได้ว่าสอดคล้องกับสมมติฐานที่สำคัญหรือไม่ และการตรวจสอบแบบง่าย (Visual Test) ก็คือ การตรวจสอบด้วยโดยการวิเคราะห์จากกราฟของค่าความคลาดเคลื่อนนั่นเอง

[1] อาจารย์ประจำสาขาวิชาบริหารอุตสาหการและปฏิบัติการ คณะพาณิชยศาสตร์และการบัญชี มหาวิทยาลัยธรรมศาสตร์

บทนำ

หน้าที่หลักอย่างหนึ่งของผู้บริหารในทุกระดับขององค์กรคือ การวางแผน ไม่ว่าจะเป็นการวางแผนระยะสั้น ระยะกลาง หรือระยะยาว การวางแผนคือ การเตรียมการหรือการเตรียมทรัพยากรให้เพียงพอสำหรับเหตุการณ์ที่จะเกิดขึ้นในอนาคต ปัจจัยที่สำคัญของกระบวนการวางแผนก็คือ ข้อมูลหรือเหตุการณ์ที่คาดคะเนว่าจะเกิดขึ้นในอนาคต ปัญหาก็คือ ไม่มีใครคาดการณ์เกี่ยวกับอนาคตได้อย่างแน่นอน ดังนั้นผู้บริหารจึงต้องทำการคาดการณ์สิ่งที่จะเกิดขึ้นในอนาคตอย่างมีหลักการหรือเรียกอย่างเป็นทางการว่า "*การพยากรณ์*"

เทคนิคการพยากรณ์แบ่งเป็น 2 ประเภท คือ

1. **การพยากรณ์เชิงคุณภาพ** คือ การพยากรณ์โดยใช้ดุลยพินิจของผู้พยากรณ์ หรือกลุ่มผู้ทำการพยากรณ์ เช่น เทคนิคเดลไฟ (Delphi Method) เป็นต้น
2. **การพยากรณ์เชิงปริมาณ** คือ เทคนิคการพยากรณ์ที่มีการนำเอาข้อมูลทางตัวเลขมาศึกษาวิเคราะห์เพื่อหารูปแบบการเปลี่ยนแปลงมาสร้างเป็นตัวแบบเพื่อใช้ในการพยากรณ์ ซึ่งเทคนิคนี้แบ่งได้เป็น 2 ประเภท คือ
   2.1 การพยากรณ์โดยการวิเคราะห์อนุกรมเวลา (Time Series Analysis) เป็นการศึกษาเพื่อหารูปแบบการเปลี่ยนแปลงของตัวแปรที่สนใจ เมื่อเวลาเปลี่ยนไปหรือการศึกษาความสัมพันธ์ระหว่างตัวแปรที่สนใจกับเวลา
   2.2 การพยากรณ์โดยการวิเคราะห์การถดถอย (Regression Analysis) ซึ่งเป็นการศึกษาความสัมพันธ์ของตัวแปรสองประเภท คือ ตัวแปรที่สนใจที่จะศึกษารูปแบบการเปลี่ยนแปลง หรือที่เรียกว่าตัวแปรตาม (Dependent Variable: Y) ซึ่งเปลี่ยนแปลงตามการเปลี่ยนแปลงของตัวแปรอีกประเภทหนึ่งหรือตัวแปรอิสระ (Independent variable: X) เพื่อใช้รูปแบบความสัมพันธ์ที่ได้ทำการพยากรณ์หรือควบคุมค่าตัวแปรตามจากการกำหนดหรือทราบค่าตัวแปรอิสระ

การวิเคราะห์ความถดถอย (Regression Analysis)

การวิเคราะห์ความถดถอย (Regression Analysis) มีด้วยกันหลายประเภท เช่น การวิเคราะห์การถดถอยเชิงลิเนียร์ (Linear Regression Analysis) การวิเคราะห์การถดถอยแบบไม่ใช่ลิเนียร์ (Non-linear Regression Analysis) หรือ การวิเคราะห์การถดถอยแบบลอจีสติก (Logistic Regression Analysis) เป็นต้น แต่เทคนิคที่ใช้กันโดยทั่วไป คือ การวิเคราะห์การถดถอยเชิงลิเนียร์ เนื่องจากเป็นตัวแบบที่เข้าใจได้ง่าย โดยไม่ต้องมีความรู้ทางคณิตศาสตร์มากนัก ประกอบกับปัจจุบันมีการพัฒนาโปรแกรมคอมพิวเตอร์ทางด้านสถิติให้สามารถใช้ได้โดยง่าย อาทิเช่น SPSS MINITAB SAS รวมทั้งการเพิ่มฟังก์ชันทางสถิติลงในโปรแกรมที่ใช้ทั่วไป เช่น Microsoft Excel ให้สามารถนำการวิเคราะห์ทางสถิติไปใช้ได้อย่างสะดวกสำหรับการใช้งานทางธุรกิจ ด้วยเหตุนี้ การสร้างตัวแบบแสดงความสัมพันธ์ของตัวแปรที่สนใจ โดยการวิเคราะห์การถดถอยจึงง่ายแค่ปลายนิ้วคลิ๊ก เราก็สามารถใช้สมการแสดงความสัมพันธ์ของข้อมูลตัวอย่างที่นำมาศึกษา ในรูปของสมการ

$$\hat{y} = b_0 + b_1 x_i$$

โดยที่ b_0 = จุดตัดบนแกน Y หรือค่าของ $\hat{y}$ เมื่อ x_i เป็นศูนย์

b_1 = ความชันของสมการนี้หรืออัตราการเปลี่ยนแปลงของ $\hat{y}$ เมื่อ x_i เปลี่ยนแปลง ไป 1 หน่วย

ดังนั้น ถ้าเรารู้ค่า x_i เราก็สามารถประมาณค่า $\hat{y}$ ได้จากสมการแสดงความสัมพันธ์นี้ ถ้าจุดมุ่งหมายของเรามีเพียงเพื่อการสร้างสมการถดถอยอธิบายความสัมพันธ์ของข้อมูลที่ได้เก็บมาเป็นตัวอย่าง การวิเคราะห์สมการถดถอยเชิงลิเนียร์ก็คงจะเสร็จสิ้นเพียงเท่านี้ แต่เนื่องจากสิ่งที่ต้องการจริง ๆไม่ใช่แค่นั้น จุดมุ่งหมายที่แท้จริงคือการนำสมการที่คำนวณได้จากข้อมูลตัวอย่าง $\hat{y} = b_0 + b_1 x_i$ ไปใช้อ้างอิงถึงค่าที่แท้จริงของประชากรหรือข้อมูลที่เป็นไปได้ทั้งหมด คือ สมการ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ เพื่อจะได้รู้ว่าค่า b_0 และ b_1 ประมาณค่า β_0 และ β_1 ได้ดีเพียงใด หรือค่าตัวแปรตามที่คำนวณได้จากสมการ $\hat{y}$ เข้าใกล้ค่าจริงของตัวแปรตาม Y_i เพียงใด เมื่อเป็นเช่นนี้ การได้สมการแสดงความสัมพันธ์ดังกล่าวเป็นเพียงแค่จุดเริ่มต้นของการวิเคราะห์เท่านั้น ความง่ายและความสะดวกของการใช้โปรแกรมคอมพิวเตอร์มาช่วยในการสร้างสมการแสดงความสัมพันธ์อาจจะเป็นดาบสองคมก็ได้ กล่าวคือ เมื่อการใช้โปรแกรมคอมพิวเตอร์ มาสร้างสมการถดถอยทำได้ง่าย อาจทำให้ผู้ใช้ละเลยที่จะศึกษาถึงวิธีการคิด วิธีคำนวณและตรรกวิทยาเบื้องหลังการได้มาซึ่งสมการดังกล่าว รวมไปถึงข้อสมมติฐานต่าง ๆของการใช้สมการดังกล่าวประกอบกับข้อมูลที่นำมาสร้างสมการดังกล่าวเป็นข้อมูลตัวอย่างเท่านั้นไม่ใช่ข้อมูลของประชากรทั้งหมด ดังนั้นการจะนำผลการศึกษาจากตัวอย่างไปสรุปลักษณะของประชากรทั้งหมดเลย คงต้องใช้วิธีการทางสถิติเชิงอ้างอิง (Inferential Statistics) กล่าวคือ ต้องมีการศึกษาถึงเทคนิคการสุ่มตัวอย่าง (Sampling Method) การแจกแจงของตัวอย่าง (Sampling Distribution) เทคนิคการประมาณค่า (Estimation) การทดสอบสมมติฐาน (Hypothesis Testing) รวมไปถึงเทคนิคที่โปรแกรมคอมพิวเตอร์ในการคำนวณสมการ คือ เทคนิคกำลังสองน้อยที่สุด (Ordinary Least Square Method: OLS) ซึ่งมีข้อสมมติที่แน่นอนเกี่ยวกับลักษณะของตัวแปรอิสระ x_i และค่าความผิดพลาด $(Y - \hat{y})$ หรือค่า ε_i ด้วย เนื่องจากสมการที่เราต้องการอ้างอิง คือ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ดังนั้นค่าของ Y_i จึงขึ้นกับ X_i และ ε_i ถ้าเราไม่สามารถระบุการเกิดขึ้นของ X_i และ ε_i ได้ เราก็ไม่สามารถที่จะอ้างอิงค่า Y_i, β_0, β_1 ได้ เมื่อเป็นเช่นนี้ ข้อสมมติฐานเกี่ยวกับ X_i และ ε_i จึงเป็นสิ่งสำคัญในการแปลความหมายค่าประมาณที่คำนวณได้จากสมการถดถอย

OLS เป็นเทคนิคที่ใช้กันมากในการสร้างสมการถดถอย เนื่องจากเป็นเทคนิคที่ง่ายในการคำนวณ แม้ไม่ใช้โปรแกรมคอมพิวเตอร์ก็ตาม และเทคนิคนี้มีหลักในการเลือกค่า b_i ที่ทำให้ค่าความผิดพลาดมีค่าน้อยที่สุด หรืออาจถือได้ว่า OLS เป็น เทคนิคที่ดีที่สุดในการประมาณค่าสมการถดถอยเชิงลิเนียร์ (*B*est *L*inear *U*nbiased *E*stimator: BLUE) คือ มีคุณสมบัติ 4 ข้อดังนี้

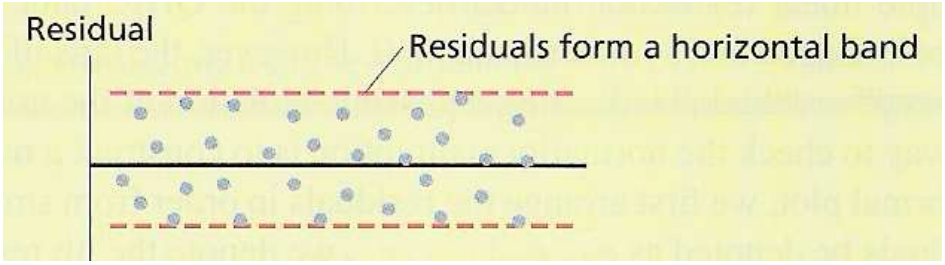
1. ความไม่เอนเอียง (Unbiased) คือ ค่า $E(b_i) = \beta_i$
2. ความมีประสิทธิภาพ (Efficient หรือ Best) คือ เทคนิคนี้จะทำให้ได้ค่าความแปรปรวนของค่า b_i จากค่า β_i มีค่าน้อยที่สุด
3. ความสม่ำเสมอ (Consistent) คือ ถ้าจำนวนตัวอย่างเพิ่มขึ้น ค่าความแปรปรวนจะลดลง มีผลทำให้ค่าประมาณ เข้าใกล้ค่าจริงขึ้นเรื่อย ๆ
4. การแจกแจงแบบปกติ (Normal Distribution) คือ การแจกแจงของค่า b_i เป็นการแจกแจงแบบปกติทำให้สามารถทำการทดสอบทางสถิติของค่า β_i ได้

แต่อย่างไรก็ตามเทคนิค OLS จะเป็น BLUE ได้ก็ต่อเมื่อสมการถดถอยที่สร้างขึ้นโดยเทคนิค OLS นี้เป็นไปตามสมมติฐานของ OLS เท่านั้น ดังนั้นการที่จะนำสมการถดถอยที่สร้างขึ้นไปใช้ได้อย่างมีประสิทธิภาพต้องมีการตรวจสอบว่าเป็นไปตามสมมติฐานดังกล่าว โดยการคำนวณค่าทางสถิติต่าง ๆมาประกอบการตัดสินใจ ซึ่งต้องใช้ความรู้ทางด้านสถิติระดับหนึ่ง แต่ในที่นี้เพื่อไม่เป็นการสร้างอุปสรรคต่อผู้ที่ไม่มีความรู้พื้นฐานทางด้านสถิติมากนักในการนำสมการถดถอยไปใช้ จึงขอนำเสนอรูปแบบการตรวจสอบความถูกต้องของสมการถดถอยโดยพิจารณาจากกราฟของค่าความผิดพลาด (Residuals Plot) ซึ่งเทคนิคนี้มีข้อดีคือ ไม่ต้องคำนวณค่าสถิติต่าง ๆเพื่อมาทำการทดสอบสมมติฐานประกอบการตัดสินใจ แต่ก็มีข้อจำกัดในการใช้คือ จะใช้ได้อย่างมีประสิทธิภาพในกรณีที่การละเมิดสมมติฐานนั้นมีค่อนข้างชัดเจน

คุณสมบัติของค่าความผิดพลาด (Residuals: e_i)

ค่าความผิดพลาด (Residuals: e_i) คือ ผลต่างระหว่างค่าจริงของตัวแปรตาม (Y_i) กับค่าตัวแปรตามที่คำนวณได้จากสมการถดถอย $\hat{y} = b_0 + b_1 x_1$ หรือ $e_i = Y_i - \hat{y}_i$ ดังนั้น ด้วยสมมติฐานของ OLS ถ้าสมการถดถอยที่ได้มีความเหมาะสม ค่าความผิดพลาดควรมีค่าเฉลี่ยเป็นศูนย์ ค่าแปรปรวนคงที่เท่ากับ σ^2 มีการแจกแจงแบบปกติ (Normal Distribution) และไม่มีรูปแบบของความสัมพันธ์ใด ๆในอนุกรมความผิดพลาดนั้น ๆ ซึ่งคุณสมบัติดังกล่าวสามารถตรวจสอบได้โดยกราฟแสดงความสัมพันธ์ระหว่างค่าความผิดพลาดกับค่าของตัวแปรอิสระ (Residuals Plot) และกราฟตรวจสอบการแจกแจงแบบปกติของค่าความผิดพลาด (Normal Probability Plot)

ในกรณีที่ค่าความผิดพลาดมีค่าเฉลี่ยเป็นศูนย์ ค่าความแปรปรวนคงที่และเป็นอิสระจากกันนั้น ผลของ Residual Plot ควรมีรูปการแจกแจงเหมือนรูปที่ 1 กล่าวคือ มีการแจกแจงอยู่ใกล้ค่าศูนย์ มีความเบี่ยงเบนจากค่าศูนย์ค่อนข้างคงที่ และไม่มีรูปแบบของการแจกแจงในอนุกรมค่าผิดพลาด



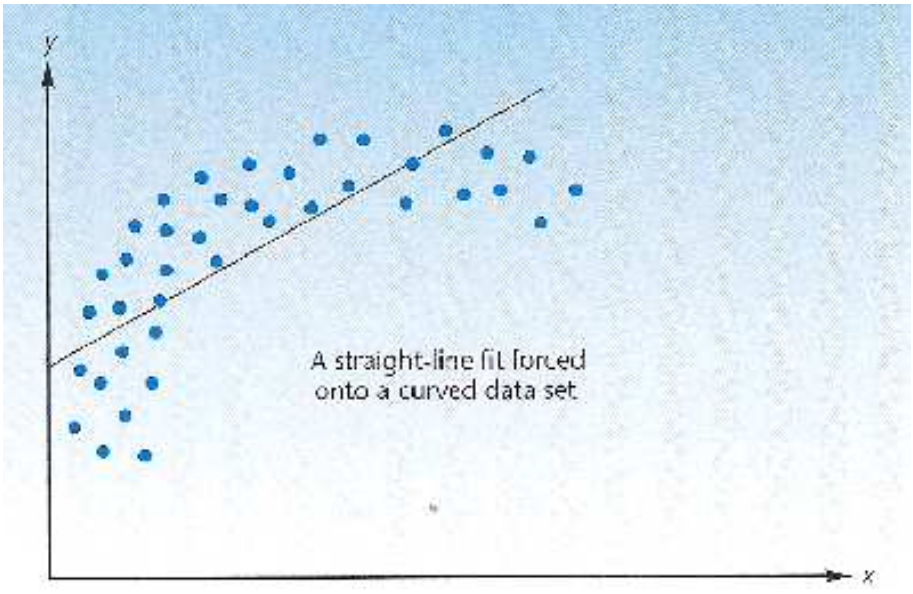
รูปที่ 1 แสดง Residual Plot ของสมการถดถอยที่เหมาะสม[2]

ในกรณีที่ Residual Plot ของสมการถดถอยใดมีการแจกแจงต่างไปจากรูปที่ 1 แสดงว่ามีการละเมิดสมมติฐานเกิดขึ้นกับสมการถดถอยนั้น ผู้ใช้จะต้องพึงระวังในการนำผลของสมการถดถอยไปใช้ อาจจะต้องมีการแปลงข้อมูลก่อนสร้างตัวแบบใหม่หรือปรับปรุงตัวแบบเดิม โดยรูปแบบการละเมิดสมมติฐานที่สามารถตรวจด้วย Residual Plot มีรูปแบบต่าง ๆ ดังนี้

[2] ที่มา: Berenson, Levine, and Krehbiel (2004), **Basic Business Statistics: Concept and Applications,** 9th Edition, Prentice Hall.

รูปแบบที่ 1 ความสัมพันธ์ของสมการถดถอยไม่ใช่ความสัมพันธ์เชิงลีเนียร์

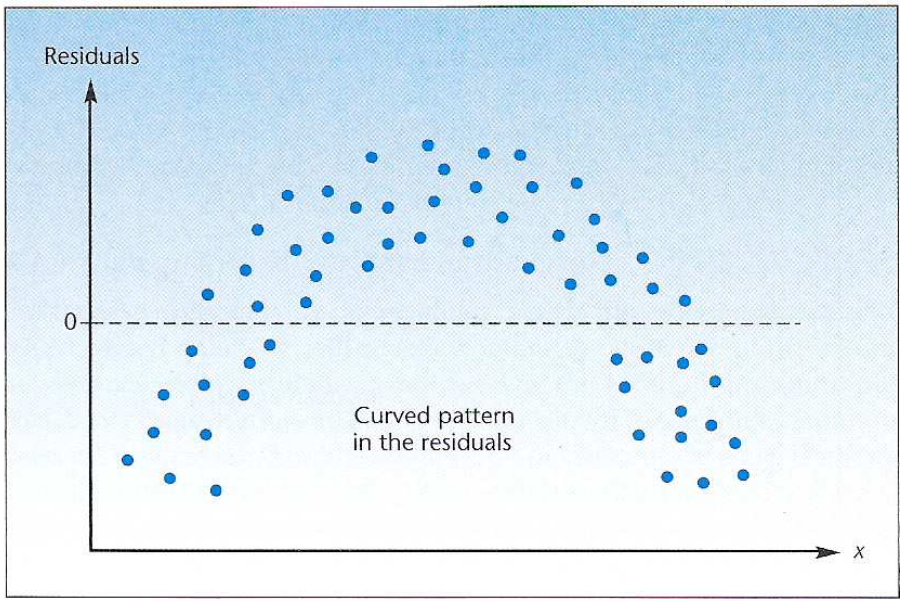
เนื่องจากเทคนิค OLS นั้น เป็นการพัฒนาสมการถดถอยเชิงลีเนียร์เพื่อแสดงความสัมพันธ์ของตัวแปร x และตัวแปร y ดังนั้นถ้าความสัมพันธ์ของตัวแปร x และตัวแปร y เป็นรูปแบบอื่น การประมาณค่า y จะมีความผิดพลาดที่มีรูปแบบเกิดขึ้น ดังตัวอย่างในรูปที่ 2



รูปที่ 2 แสดง Scatter Plot ของความสัมพันธ์แบบเส้นโค้ง[3]

ในรูปที่ 2 นี้ เป็นการสร้างกราฟเปรียบเทียบระหว่างค่า y จริง กับค่า $\hat{y}$ ที่คำนวณจากสมการถดถอยเชิงลีเนียร์ หรือ Predicted Y จะเห็นได้ว่าการใช้สมการถดถอยเชิงลีเนียร์ประมาณค่า y ที่มีความสัมพันธ์กับ x แบบไม่ใช่ลิเนียร์ จะมีความผิดพลาดที่มีรูปแบบ กล่าวคือ ค่า $\hat{y}$ จะสูงกว่าค่า y จริงมาก ณ ระดับค่า x น้อย และความผิดพลาดจะลดลงเมื่อค่า x สูงขึ้น จากนั้นการประมาณค่าจะต่ำกว่าค่าจริงในช่วงค่า x ณ ระดับถัดมา แล้วค่าประมาณดังกล่าว จะสูงกว่าค่าจริงอีก ซึ่งรูปแบบนี้สามารถมองได้ชัดเจนขึ้นโดยการใช้ Residual Plot ดังรูปที่ 3

[3] ที่มา: Aczel (2002), **Complete Business Statistics,** 5th Edition, McGraw-Hill.



รูปที่ 3 แสดง Residual Plot กรณีความสัมพันธ์เป็นแบบเส้นโค้ง[4]

จากรูปที่ 3 จะเห็นรูปแบบการแจกแจงของค่าความผิดพลาดอย่างชัดเจนว่า มีลักษณะเป็นเส้นโค้ง แสดงว่าการใช้สมการถดถอยเชิงลิเนียร์ดังกล่าวมาอธิบายการเปลี่ยนแปลงของค่า y ไม่เหมาะสม ควรจะเปลี่ยนรูปแบบของความสัมพันธ์ดังกล่าว โดยการปรับฟังก์ชันความสัมพันธ์ของตัวแปร

รูปแบบที่ 2 กรณีค่าความแปรปรวนของค่าความผิดพลาดไม่คงที่ (Heteroscedasticity)

การละเมิดสมมติฐานข้อนี้จะทำให้ค่าความแปรปรวนของสมการถดถอยที่คำนวณจากเทคนิค OLS จะไม่ใช่ค่าที่ต่ำที่สุดอีกต่อไป จึงเป็นการลดความมีประสิทธิภาพของสมการถดถอยดังกล่าว คุณสมบัติของค่าความแปรปรวนของค่าความผิดพลาด คือ ค่าความผิดพลาดต้องมีค่าความแปรปรวนคงที่ ในทุกระดับค่าของ X_i หรือเรียกว่า Homoscedasticity คือ

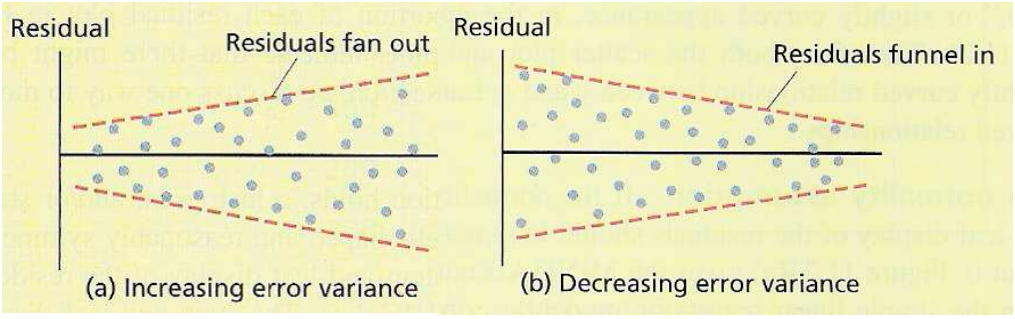
$$Var(\varepsilon_i | X_i) = \sigma^2$$

แต่ถ้าความแปรปรวนของค่าความผิดพลาดเพิ่มหรือลด ในขณะที่ค่า X เพิ่มขึ้น จะเรียกว่าเกิด Heteroscedasticity หรือ

$$Var(\varepsilon_i | X_i) = \sigma_i^2$$

สำหรับปัญหา Heteroscedasticity สามารถตรวจสอบจาก Residual Plot กล่าวคือ ถ้าการกระจายของค่าความผิดพลาดมีการเปลี่ยนแปลงไม่ว่าจะเพิ่มขึ้นหรือลดลงเมื่อค่า x เพิ่มขึ้น แสดงถึงกรณีค่าความแปรปรวนของค่าความผิดพลาดไม่คงที่ ดังรูปที่ 4

[4] ที่มา: Aczel (2002), **Complete Business Statistics**, 5th Edition, McGraw-Hill.



รูปที่ 4 แสดง Residual Plot กรณีความแปรปรวนไม่คงที่[5]

จากรูปที่ 4 (a) จะเห็นได้ว่าค่าความเบี่ยงเบนของค่าความผิดพลาดมีการเพิ่มขึ้น เมื่อ ค่า x เพิ่มขึ้น ในขณะที่รูปที่ 4 (b) ค่าความเบี่ยงเบนของค่าความผิดพลาดลดลง เมื่อ ค่า x เพิ่มขึ้น ทั้งสองกรณีนี้เรียกว่าเกิด Heteroscedasticity หรือค่าความแปรปรวนของค่าความผิดพลาดไม่คงที่ ในกรณีนี้การใช้เทคนิค OLS จะไม่เหมาะสม เนื่องจากค่าความแปรปรวนที่ได้ไม่ใช่ค่าที่ต่ำที่สุด ทำให้ไม่สามารถสรุปได้ว่าค่าประมาณที่ได้เป็น BLUE เพื่อแก้ไขปัญหาดังกล่าวต้องใช้เทคนิคที่ซับซ้อนมากขึ้น กล่าวคือ ต้องมีการปรับค่าของตัวแปรตามหรือตัวแปรอิสระ เรียกเทคนิคนั้นว่า Weighted Least Square (WLS)

รูปแบบที่ 3 ค่าความผิดพลาดไม่เป็นอิสระจากกัน (Autocorrelation)

ส่วนใหญ่เกิดขึ้นในกรณีข้อมูลเป็นแบบอนุกรมเวลา (Time Series) ตามสมมติฐานของ OLS แล้วค่าความผิดพลาด ε_i ของค่าสังเกตที่แตกต่างกันจะเป็นอิสระต่อกัน ซึ่งหมายความว่า

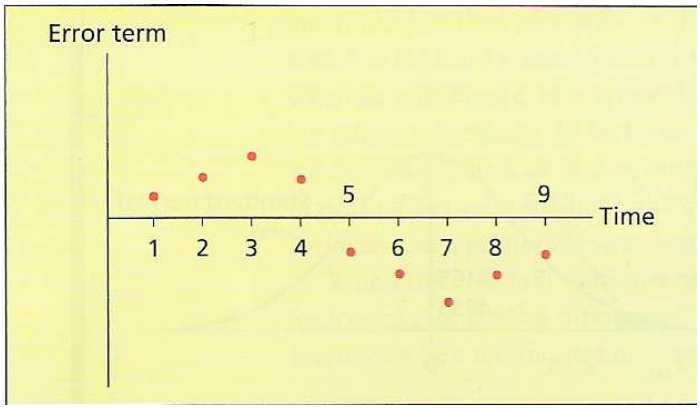
$$Cov(\varepsilon_i, \varepsilon_j) = 0$$

กล่าวคือ ไม่มีสหสัมพันธ์ในตัวเอง (Autocorrelation) ระหว่างค่า ε_i ใดๆ ซึ่งหมายความว่า ณ ค่า X_i ความเบี่ยงเบนของ Y_i จากค่าเฉลี่ย 2 ตัวใดๆ จะไม่มีรูปแบบ

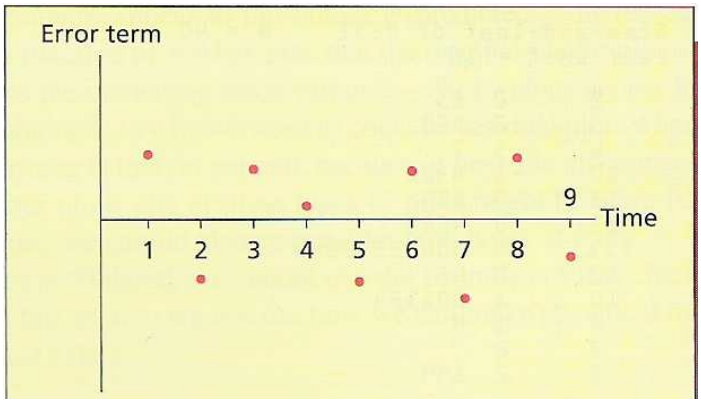
เนื่องจากสมการถดถอยที่แท้จริงคือ $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ ถ้า ε_i และ ε_j มีความสัมพันธ์กัน ก็หมายความว่า Y นั้นไม่ได้ขึ้นอยู่กับค่า X_i เท่านั้น แต่ยังขึ้นกับค่า ε_j ด้วย เนื่องจากค่า ε_j มีส่วนกำหนดค่า ε_i ดังนั้นในการกำหนดสมมติฐานนี้ก็เพื่อศึกษาเฉพาะผลกระทบที่เป็นระบบของ X_i ต่อ Y_i โดยไม่คำนึงถึงผลกระทบจากปัจจัยอื่นๆ ที่มีต่อ Y_i

การตรวจสอบการมีสหสัมพันธ์ในตัวเอง (Autocorrelation) ระหว่างค่าความผิดพลาดใด ทำโดยการสร้างกราฟแสดงความสัมพันธ์ระหว่างค่าความผิดพลาดกับเวลา ถ้าพบรูปแบบการแจกแจง กล่าวคือ ค่าความผิดพลาดที่เกิดขึ้นก่อนสามารถอธิบายค่าความผิดพลาดที่เกิดขึ้นหลังได้ ดังรูปที่ 5 และรูปที่ 6

[5] ที่มา: Berenson, Levine, and Krehbiel (2004), **Basic Business Statistics: Concept and Applications**, 9th Edition, Prentice Hall.



รูปที่ 5 แสดงสหสัมพันธ์ในตัวเองที่มีค่าเป็นบวก[6]



รูปที่ 6 สหสัมพันธ์ในตัวเองที่มีค่าเป็นลบ[7]

จากรูปที่ 5 การแจกแจงของค่าความผิดพลาดแสดงถึงความสัมพันธ์ระหว่างค่าความผิดพลาดเป็นไปในทิศทางเดียวกัน ส่วนรูปที่ 6 นั้นการแจกแจงของค่าความผิดพลาดแสดงถึงความสัมพันธ์ระหว่างค่าความผิดพลาดที่เป็นทิศทางตรงข้าม กรณีเช่นนี้ถือว่าการแจกแจงของค่าความผิดพลาดไม่เป็นอิสระจากกัน ซึ่งอาจเกิดจากการที่ความสัมพันธ์ของตัวแปรไม่ใช้ความสัมพันธ์แบบลิเนียร์ หรือการละเลยตัวแปรที่สำคัญจากสมการถดถอย (Omitted Variable) ต้องทำการตรวจสอบถึงสาเหตุที่แท้จริงเสียก่อนจึงจะสามารถปรับสมการถดถอยได้

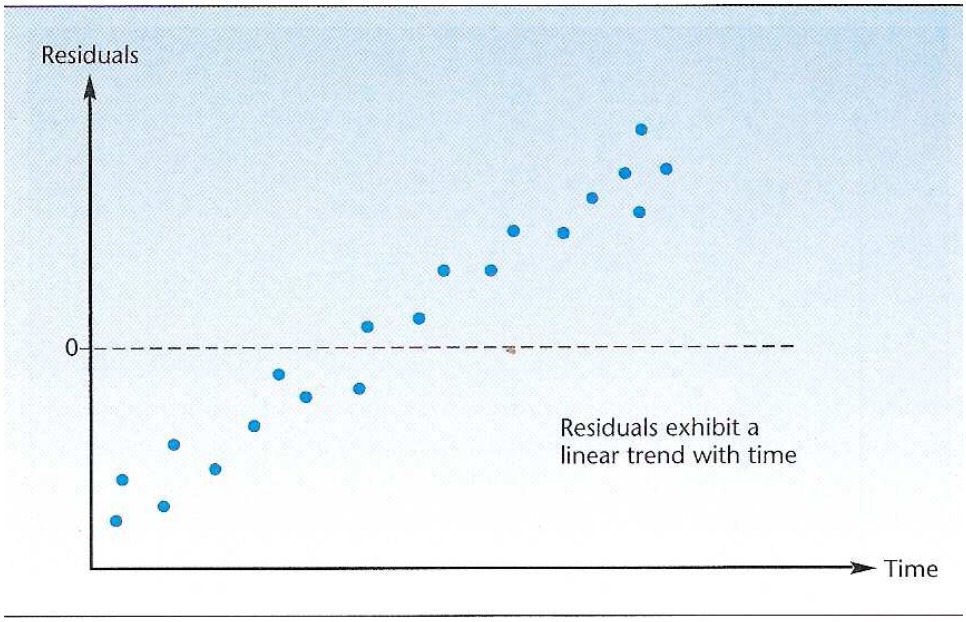
รูปแบบที่ 4 การขาดตัวแปรที่สำคัญในสมการถดถอย (Omitted Variables)

กรณีนี้เป็นการละเลยตัวแปรสำคัญที่สามารถอธิบายการเปลี่ยนแปลงของตัวแปรตาม y ผลลัพธ์ที่เกิดขึ้น คือ จะทำให้เกิดการเอนเอียง (Biased) ของค่า b_0 และ b_1 ไปจากค่าพารามิเตอร์ β_0 และ β_1 ที่แท้จริง โดยสามารถตรวจสอบโดยการสร้างกราฟแสดงความสัมพันธ์ระหว่างค่าความผิดพลาดกับตัวแปรอื่นที่คาดว่าน่าจะใส่เข้าไปในสมการถดถอย ถ้ามีความสัมพันธ์เกิดขึ้น คือ การแจกแจงของค่าความผิดพลาดมีรูปแบบอย่างใดอย่างหนึ่ง แสดงว่าควรใส่ตัวแปรนั้นเข้าไปในสมการถดถอย เพราะตัวแปรนั้นสามารถอธิบายการเปลี่ยนแปลงของความผิดพลาดที่เกิดขึ้นได้ ดังรูปที่ 7

---

[6] ที่มา: Bowerman, O'Connell, and Hand (2001), **Business Statistics in Practice**, 2nd Edition, McGraw-Hill.

[7] ที่มา: Bowerman, O'Connell, and Hand (2001), **Business Statistics in Practice**, 2nd Edition, McGraw-Hill.



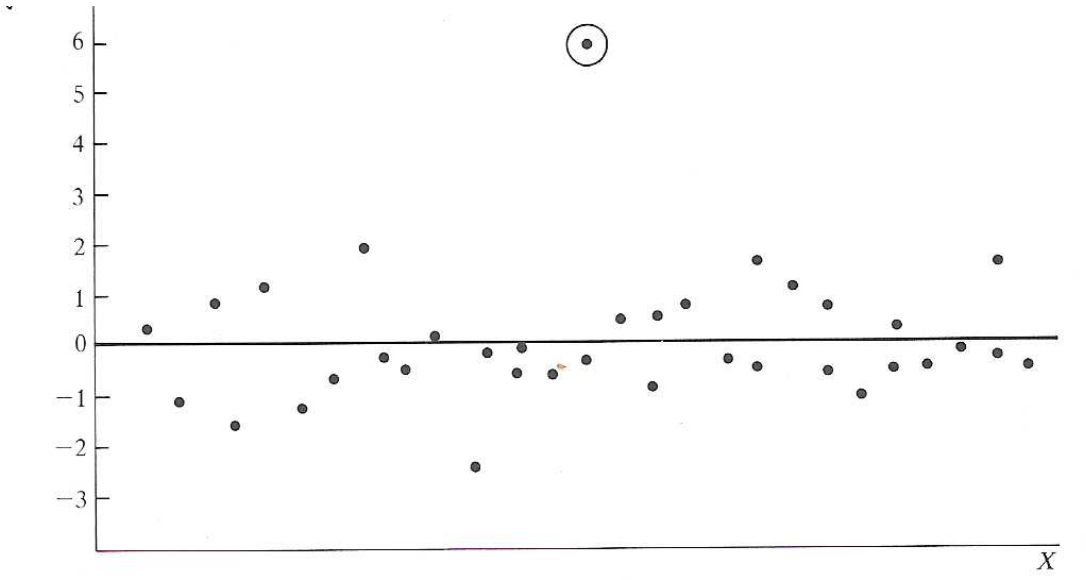
รูปที่ 7 แสดงรูปแบบการแจกแจงของค่าความผิดพลาดกับเวลา[8]

จากรูปที่ 7 ถ้าเราคิดว่าตัวแปรที่สำคัญที่ถูกละเลยไปคือ เวลา การตรวจสอบเบื้องต้นคือ การสร้างกราฟแสดงความสัมพันธ์ระหว่างค่าความผิดพลาดกับเวลา ในกรณีนี้จะเห็นได้ว่ามีความสัมพันธ์กันอย่างชัดเจน ดังนั้นถ้าเราใส่ตัวแปร เวลา เข้าไปในสมการถดถอย จะทำให้สมการถดถอยนี้อธิบายการเปลี่ยนแปลงของตัวแปรตาม y ได้ดีขึ้น

รูปแบบที่ 5 การมีค่าที่ผิดปกติ (Outliers) ในข้อมูลที่เก็บมา

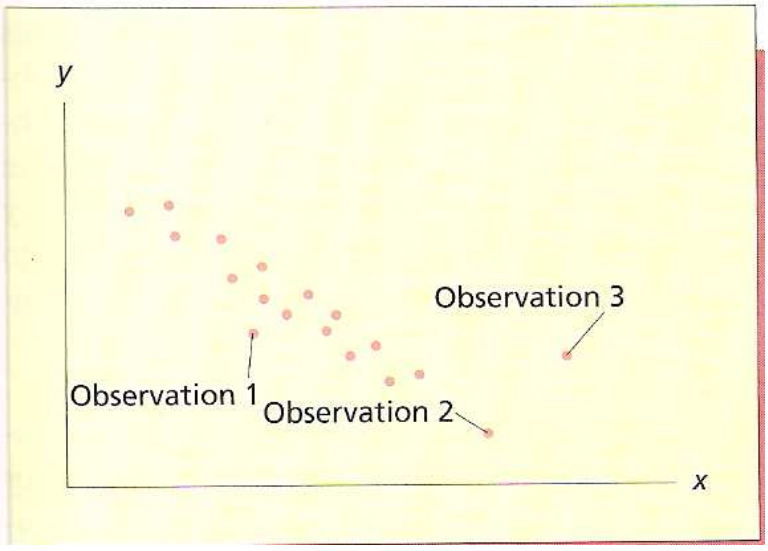
ค่าที่ผิดปกติ คือ ข้อมูลที่มีค่ามากหรือน้อยผิดปกติ การใช้เทคนิค OLS ในการสมการถดถอยนั้น จะพิจารณาค่าผิดปกตินี้ออกเป็นสองประเภทคือ ค่าผิดปกติที่ไม่มีผลต่อการคำนวณค่า b_i (Outliers Observations) กล่าวคือ ไม่ว่าจะมีค่านี้หรือไม่ ค่า b_i ไม่ต่างกันมากนัก ส่วนค่าผิดปกติอีกประเภทจะมีอิทธิพลต่อการคำนวณค่า b_i ในสมการถดถอย (Influential Observations) หรืออาจกล่าวได้ว่าการตัดข้อมูลนี้จะมีผลต่อค่า b_i ของสมการถดถอย การตรวจสอบอาจทำได้โดยการใช้ Residual Plot ดังรูปที่ 8 และการใช้ Scatter Plot ดังรูปที่ 9

[8] ที่มา: Aczel (2002), **Complete Business Statistics,** 5th Edition, McGraw-Hill.



รูปที่ 8 แสดง Residual Plot กรณีมีค่าผิดปกติ (Outlier)[9]

จากรูปที่ 8 เมื่อพิจารณาการแจกแจงของค่าความผิดพลาดใน Residual Plot สามารถตรวจสอบค่าผิดปกติได้ คือ ค่าที่ต่างจากกลุ่มมากๆ แต่ถ้าต้องการรู้ถึงประเภทของค่าผิดปกติว่ามีอิทธิพลต่อสมการถดถอยหรือไม่ สามารถตรวจสอบเบื้องต้นโดยการใช้ Scatter Plot ดังรูปที่ 9



รูปที่ 9 แสดง Scatter Plot กรณีมีค่าผิดปกติ (Outlier)[10]

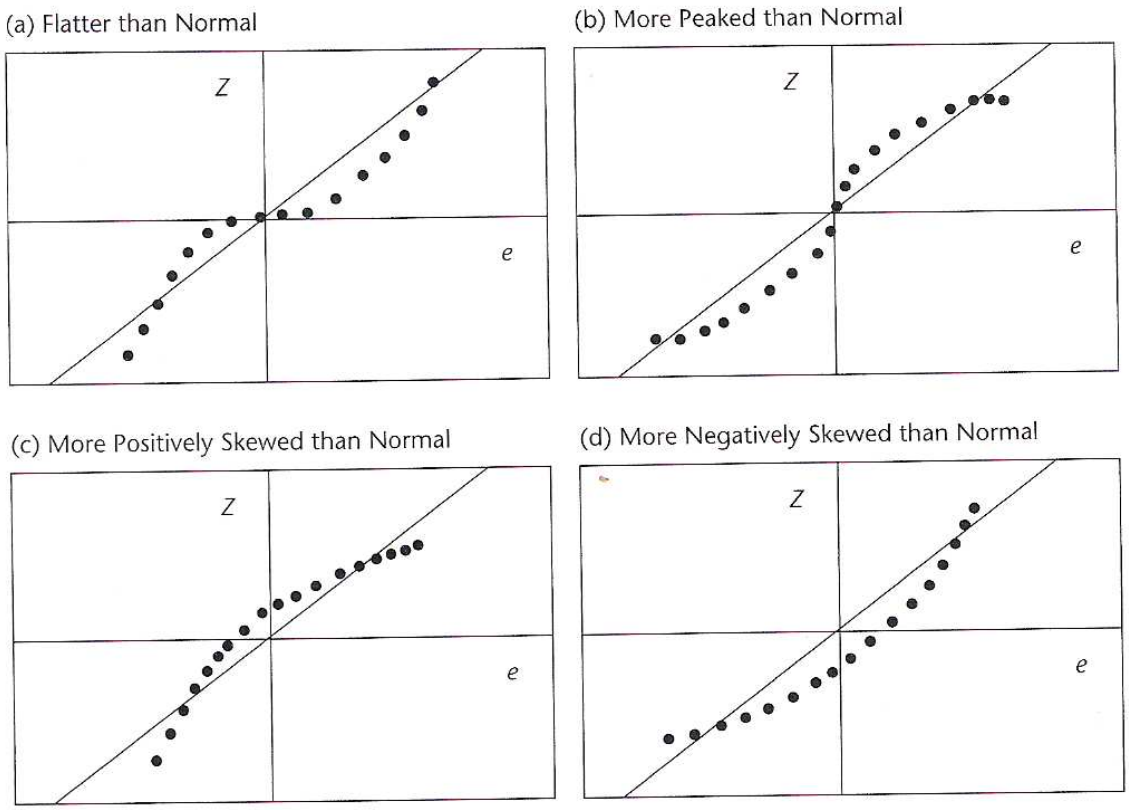
[9] ที่มา: Bowerman, O'Connell, and Hand (2001), **Business Statistics in Practice**, 2nd Edition, McGraw-Hill.

[10] ที่มา: Neter, Wasserman, and Kutner (1990), **Applied Linear Statistics Models**, 3rd Edition, Irwin.

จากรูปที่ 9 ข้อสรุปเบื้องต้น คือ ค่า Observation 1 และ 2 ไม่มีผลทำให้การประมาณค่าสมการถดถอยที่แท้จริงคลาดเคลื่อน ในขณะที่ Observation 3 นั้นค่อนข้างเป็นที่น่าสงสัยว่าจะมีอิทธิพลต่อการประมาณค่าดังกล่าว แต่อย่างไรก็ตามการตัดสินใจที่เชื่อถือได้ในกรณีนี้ต้องมีการคำนวณค่าสถิติเพื่อประกอบการทดสอบสมมติฐาน

รูปแบบที่ 6 การแจกแจงของค่าความผิดพลาดไม่เป็นการแจกแจงแบบปกติ

การแจกแจงแบบปกติของค่าความผิดพลาดไม่ใช้สมมติฐานที่ส่งผลให้ค่าที่ได้ในสมการถดถอย โดยเทคนิค OLS เป็น BLUE แต่การแจกแจงแบบปกติของค่าความผิดพลาดก็สำคัญสำหรับการทดสอบสมมติฐานต่าง ๆของสมการ ซึ่งการตรวจสอบดูว่าการแจกแจงของค่าความผิดพลาดเป็นการแจกแจงแบบปกติหรือไม่ สามารถตรวจสอบโดยใช้ Normal Probability Plot โดยที่ค่าความผิดพลาดจะอยู่บนแกนนอน ถ้าการแจกแจงของค่าความผิดพลาดเป็นการแจกแจงแบบปกติ ค่าความผิดพลาดจะอยู่เป็นเส้นทะแยงมุม ถ้าการแจกแจงไม่เป็นเส้นตรงทะแยงมุม แสดงว่าเป็นรูปแบบการแจกแจงแบบอื่น ดังรูปที่ 10



รูปที่ 10 แสดง Normality Plot กรณีการแจกแจงของค่าความผิดพลาดไม่เป็นการแจกแจงแบบปกติ[11]

จากรูปที่ 10 (a) แสดงถึงการแจกแจงที่ค่อนข้างแบนกว่าการแจกแจงแบบปกติ รูป 10 (b) แสดงถึงการแจกแจงที่ค่อนข้างโด่งกว่าการแจกแจงแบบปกติ รูป 10 (c) แสดงถึงการแจกแจงที่มีลักษณะเบ้ขวา ส่วนรูป 10 (d) แสดงถึงการแจกแจงที่มีลักษณะเบ้ซ้าย

[11] ที่มา: Aczel (2002), **Complete Business Statistics**, 5th Edition, McGraw-Hill.

บทสรุป

จะเห็นได้ว่าการตรวจสอบความเหมาะสมของสมการถดถอยโดยการวิเคราะห์ค่าความผิดพลาดจากกราฟ (Residual Plot) นั้นมีประโยชน์ ทำได้ไม่ยากและไม่เสียเวลามากนัก และจะทำให้ผลจากการประมาณค่าโดยสมการถดถอยซึ่งเป็นปัจจัยนำเข้า (Input) ของกระบวนการวางแผนมีความถูกต้อง แม่นยำมากขึ้น ดังนั้นเมื่อปัจจัยนำเข้าดีและกระบวนการวางแผนมีระบบ ก็สามารถรับประกันได้ว่าผลที่ออกมาน่าจะมีความเหมาะสม ซึ่งจะเป็นประโยชน์อย่างยิ่งในเชิงเศรษฐกิจ เนื่องจากถ้าข้อมูลที่พยากรณ์ซึ่งเป็นปัจจัยนำเข้าไม่ถูกต้อง ไม่ว่าวิธีการวางแผนจะดีเพียงใดแผนที่ได้ก็ผิดพลาดได้ ส่งผลให้การตัดสินใจของผู้บริหารก็จะผิดพลาด ทำให้เกิดการสูญเสียเชิงเศรษฐกิจ ไม่ว่าจะเป็นการเตรียมทรัพยากรไว้ไม่เพียงพอ หรือการเตรียมทรัพยากรไว้มากเกินไป การตรวจสอบความเหมาะสมของสมการถดถอยก่อนการนำไปใช้ซึ่งถือได้ว่าเป็นการป้องกันความผิดพลาดในการวางแผนระดับหนึ่ง คือ กันไว้ดีกว่าแก้นั่นเอง

บรรณานุกรม

กัลยาณี คูณมี (2541), **สถิติสำหรับเศรษฐศาสตร์และธุรกิจ**, พิมพ์ครั้งที่ 2, โรงพิมพ์แห่งจุฬาลงกรณ์.

ฉันทลักษณ์ ณ ป้อมเพชร์, เยาวมาลย์ เมธาภิรักษ์, และศรีเพ็ญ ทรัพย์มนชัย (2537), **สถิติเพื่อการวางแผน**, โรงพิมพ์มหาวิทยาลัยธรรมศาสตร์.

ไพฑูรย์ ไกรพรศักดิ์ (2546), **เศรษฐมิติเบื้องต้น**, พิมพ์ครั้งที่ 1, โรงพิมพ์แห่งจุฬาลงกรณ์.

ศรีเพ็ญ ทรัพย์มนชัย (2545), **การวิเคราะห์การถดถอยเชิงธุรกิจ**, พิมพ์ครั้งที่ 1, ปิ่นเกล้าการพิมพ์.

Aczel (2002), **Complete Business Statistics**, 5th Edition, McGraw-Hill.

Berenson, Levine, and Krehbiel (2004), **Basic Business Statistics: Concept and Applications**, 9th Edition, Prentice Hall.

Bowerman, O'Connell, and Hand (2001), **Business Statistics in Practice**, 2nd Edition, McGraw-Hill.

Breiman and Freedman (1983), "How Many Variables Should Be Entered in a Regression Equation?", **Journal of the American Statistical Association**, March, 131 - 136.

McCloskey and Zillak (1996), "The Standard Error of Regressions", **Journal of Economic Literature**, March, 97 - 114.

Neter, Wasserman, and Kutner (1990), **Applied Linear Statistics Models**, 3rd Edition, Irwin.

Tryfos (1998), **Methods For Business Analysis and Forecasting: Text & Case**, Wiley.

Studendmund (2001), **Using Econometrics: A Practical Guide**, 4th Edition, Addison Wesley Longman.

... (2004), Signifying Nothing?, **The Economist**, January.