ข้อตกลงการถดถอยและกระบวนการวิเคราะห์การถดถอย

(Regression assumption and Regression process)

มนตรี พิริยะกุล[1]

บทคัดย่อ

การวิเคราะห์การถดถอยคือวิทยาสถิติที่ใช้ศึกษาความเชื่อมโยงระหว่างตัวแปรตามซึ่งเป็นเป้าหมายของการศึกษากับตัวแปรอิสระซึ่งเป็นปัจจัยสาเหตุที่สามารถควบคุมหรืออธิบายความผันแปรของตัวแปรตามโดยมุ่งประโยชน์ 2 ประการคือเพื่อการพยากรณ์และ/หรือศึกษาความสัมพันธ์เชิงสาเหตุ การศึกษาด้วยวิธีวิทยานี้จำเป็นต้องกำหนดข้อตกลงหลายประการเพื่อควบคุมความเคลื่อนไหวของตัวแปรส่วนเหลือรวมทั้งควบคุมความเกี่ยวข้องกันระหว่างตัวแปรอิสระ ด้วยเหตุนี้การศึกษาตามวิธีวิทยานี้จึงจำเป็นต้องตรวจสอบความเป็นจริงของข้อตกลงว่าข้อมูลเชิงประจักษ์สอดคล้องกับข้อตกลงหรือไม่ หากตรวจพบว่าไม่สอดคล้องเราจำเป็นจะต้องแก้ไขให้สอดคล้องมิเช่นนั้นผลการวิจัยจะผิดพลาดและนำสู่การนำไปใช้กำหนดนโยบายที่ผิดพลาดด้วย

Abstract

Regression analysis is a powerful statistical methodology aims at linking between explained variables and their target variable in order to predict it or to explain its causes of variation. There are several assumptions for fulfilling this kind of analysis and needed for detections and also corrections if empirical analysis revealed that data and assumptions were not congruent, else the finding would miss leading the implications followed.

คำสำคัญ: สมการถดถอย ข้อตกลงการถดถอย ภาวะความเป็นสมการเชิงเส้น ภาวะร่วมเส้นตรงพหุ autocorrelation heteroscedasticity

Key word: Regression analysis, regression assumption, linearity, multicollinearity, autocorrelation, heteroscedasticity

[1] ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยรามคำแหง

ความนำ

การวิเคราะห์การถดถอยคือกระบวนการทางสถิติเพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปร โดยมีเป้าหมาย 2 ประการคือเพื่อพยากรณ์ค่าของตัวแปรตามและเพื่อศึกษาปัจจัยสาเหตุของตัวแปรตาม ผลการศึกษานำไปสู่การใช้ประโยชน์ที่ต่างกันคือ ในกรณีศึกษาเพื่อการพยากรณ์เราจะนำผลการพยากรณ์ไปใช้กำหนดค่าของตัวแปรตามเมื่อตัวแปรอิสระมีค่าอื่นตามกำหนด หรือเพื่อกำหนดค่าอนาคตของตัวแปรเมื่อตัวแปรอิสระมีค่าอื่นที่คาดว่าจะเป็นไปได้ในภายหน้า เช่น การพยากรณ์ยอดขายซึ่งจะถูกนำไปใช้สนับสนุนการตัดสินใจด้านการผลิต ด้านการขาย ด้านการส่งเสริมการขาย ค่าพยากรณ์จำนวนนักท่องเที่ยวนำสู่การบริหารการท่องเที่ยวและบริหารทรัพยากรการท่องเที่ยว ในกรณีศึกษาตามแนวทางความสัมพันธ์เชิงสาเหตุนำไปสู่การกำหนดนโยบายลงไปที่ตัวแปรอิสระเพื่อควบคุมให้ค่าตัวแปรตามเป็นไปตามความต้องการ

สมการถดถอยและข้อตกลงการถดถอย

เนื่องจากในการวิเคราะห์การถดถอยนั้นมีตัวแปรอิสระที่ใช้ควบคุมหรืออธิบายตัวแปรตามได้มากมายหลายตัว ทั้งที่เป็นไปตามทฤษฎีและวรรณกรรมและที่เป็นไปตามเหตุผลและสถานการณ์เชิงประจักษ์ ในจำนวนนี้จะมีตัวแปรจำนวนมากที่เราไม่รู้จักหรือรู้จักแต่ไม่มีข้อมูลหรือหาข้อมูลไม่ได้ หรือไม่อาจหา proxy ที่เหมาะสมมาใช้แทน หรือตัวแปรมีความเป็นนามธรรมหรือเป็นข้อมูลอนุกรมเวลาแต่บันทึกไว้ด้วยนิยามเวลาต่างกัน หรือไม่บันทึกข้อมูลอย่างต่อเนื่อง ตัวแปรเหล่านี้แม้ไม่ปรากฏในสมการถดถอยแต่ก็ยังคงแอบส่งอิทธิพลต่อตัวแปรตาม นอกนี้ยังมีความคลาดเคลื่อนจากการวัดอีกด้วย ด้วยเหตุนี้เราจึงรวมตัวแปรดังกล่าวที่นอกเหนือจากที่กำหนดไว้ในสมการรวมทั้งความคลาดเคลื่อนจากการวัดไว้ในที่เดี่ยวกัน เรียกว่า residual หรือ disturbance term ใช้สัญลักษณ์ ε หรือ u ตัวแปร u จึงมีส่วนประกอบที่หลากหลายมากมายยากแก่การควบคุมให้อยู่นิ่งๆ ทั้งไม่ทราบทิศทางหรือปริมาณที่จะพึงมีในแต่ละวาระในข้อมูลอนุกรมเวลา (time series) หรือในแต่ละหน่วยวิเคราะห์ในข้อมูลภาคตัดขวาง (cross section)

กำหนดให้ Y = ตัวแปรตาม (dependent variable หรือ target variable หรือ endogenous variable)

กำหนดให้ X_j ; j =1 ,2 ,..., k คือตัวแปรอิสระที่ทำหน้าที่อธิบายหรือคาดคะเนความผันแปรของตัวแปรตาม (เรียกว่า explained variable หรือ predictor หรือ independent variable หรือ exogenous variable)

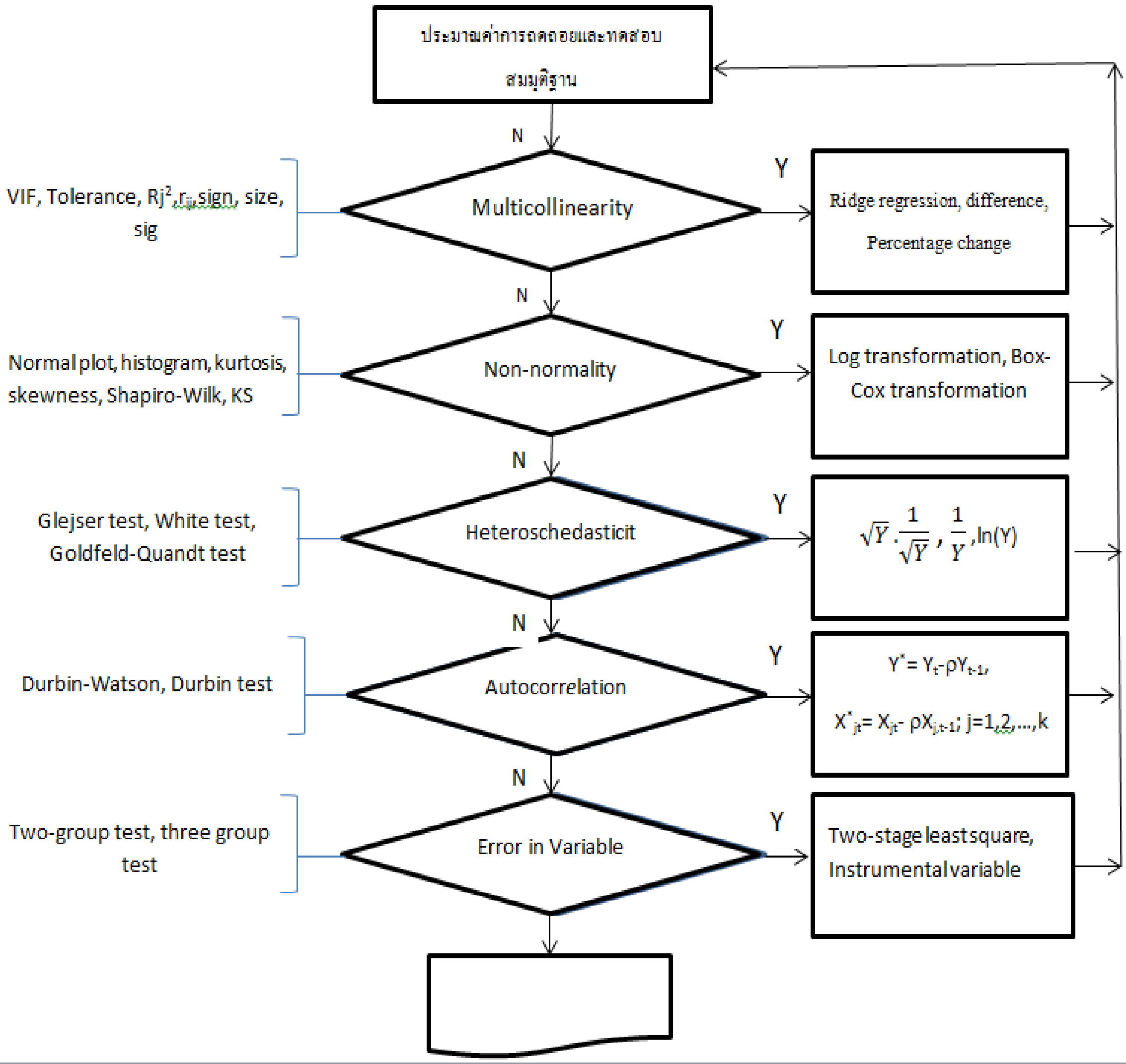
กำหนดให้ β_j ; j =1,2 ,…, k คือน้ำหนักของ X_j ที่มีต่อ Y คือ $\beta_j = \frac{\Delta Y}{\Delta X_j}$ เรียกว่า partial regression weight หรือ partial regression coefficient

กำหนดให้ u คือส่วนเหลือหรือตัวก่อกวน

ดังนั้นสมการถดถอยเชิงเส้นคือสมการความสัมพันธ์เชิงสาเหตุระหว่าง X_j's กับ Y ดังนี้คือ $Y = \beta_o + \sum_{j=1}^{k} \beta_j X_j + u$ โดยที่ $\beta_j = \frac{\Delta Y}{\Delta X_j}$ หมายถึงอัตราการเปลี่ยนแปลงค่าของตัวแปร Y เมื่อตัวแปร X_j's เปลี่ยนค่าไป 1 หน่วย

กระบวนการวิเคราะห์การถดถอย

เนื่องจากการพัฒนาตัวประมาณค่าในสมการถดถอยดำเนินการตามข้อตกลงการถดถอย ทั้งนี้การตกลงก็มิใช่ทำได้ตามใจ แต่เป็นไปตามเหตุผลดังได้กล่าวมาแล้ว ดังนั้นผลการวิเคราะห์การถดถอยในขั้นปฏิบัติจึงต้องสอดคล้องกับข้อตกลงมิเช่นนั้นผลการศึกษาจะคลาดเคลื่อนและนำสู่การสรุปผลผิดพลาด (miss-leading) ซึ่งหากนำผลการศึกษาไปใช้ในระดับนโยบายหรือในระดับปฏิบัติการก็อาจก่อให้เกิดความเสียหายได้ กระบวนการวิเคราะห์การถดถอยดำเนินการดังภาพ ในภาพจะแสดงขั้นการวิเคราะห์ที่เริ่มจากการประมาณค่าสมการถดถอยและทดสอบสมมุติฐาน ซึ่งเป็นพื้นฐานหรือก้าวแรกของการวิเคราะห์ด้วยวิธีวิทยาสถิตินี้ ห้ามหยุดที่ก้าวนี้ ให้ดำเนินการต่อเนื่องจนจบเพื่อตรวจสอบและแก้ปัญหาข้อตกลงการถดถอยถ้าตรวจพบทุกข้อตกลง ในภาพจะแสดงรายละเอียดเบื้องต้น ข้อความต่อจากนี้ไปจะเป็นรายละเอียดเรียงตามข้อตกลง ทุกอย่างที่จะอ่านพบต่อไปอาจไม่ได้แสดงไว้ในภาพ



1. Linearity

ข้อตกลงนี้คือตกลงว่า $E(Y|X\text{'s}) = \beta_o + \beta_1X_1 + \beta_2X_2 + \cdots + \beta_kX_k$ หรือสมการถดถอยคือสมการที่เป็นทางเดินจุดของค่าเฉลี่ยของตัวแปรตาม (Y) ณ ค่าเฉพาะของตัวแปรอิสระ X's หากข้อตกลงไม่เป็นจริงค่าพยากรณ์ของ Y (ที่จริงคือค่าพยากรณ์ของ $E(Y|X)$ จะคลาดเคลื่อน

วิธีตรวจสอบข้อตกลงความเป็นสมการเชิงเส้น สามารถกระทำได้ดังนี้

1) วิธี RESET (Ramsey Regression Equation Specification Error Test) วิธีนี้มุ่งตรวจสอบว่า หากเพิ่มเทอมต่อไปนี้คือ $\Upsilon_1\widehat{Y}^2,\Upsilon_2\widehat{Y}^2,\dots,\Upsilon_h\widehat{Y}^h$ ลงในสมการ

$$Y = \beta_o + \sum_j^k \beta_j X_j + u$$

จะช่วยอธิบาย Y ได้ดีขึ้นหรือไม่ ถ้าช่วยให้ดีขึ้นก็แสดงว่าสมการเดิมมีปัญหา Miss-specification

วิธีดำเนินการเป็นดังนี้

(1) วิเคราะห์สมการถดถอย $Y = \beta_o + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า $\widehat{Y}$ และ R^2 เรียกว่า R_1^2 มี df = n-(k+1)

(2) วิเคราะห์สมการถดถอย $Y = \beta_o + \sum_j^k \beta_j X_j + \sum_j^h \Upsilon_j \widehat{Y}^j + u$ แล้วบันทึกค่า R^2 เรียกว่า R_2^2 มี df = n-(k+h+1)

(3) สมมุติฐานที่ต้องการทดสอบคือ $\Upsilon_1 = \Upsilon_2 = \cdots = \Upsilon_h = 0$ สถิติทดสอบคือ

$$F_c = \frac{\frac{(R_2^2 - R_1^2)}{h}}{\frac{1-R_2^2}{n-(k+h+1)}} \sim F_{n-(k+h+1)}$$

ถ้าพบว่า p-value ≤ .05 แสดงว่ามีปัญหา nonlinearity

2) กราฟความสัมพันธ์ระหว่าง standardized residual กับ standardized predicted value ถ้าพบว่ากราฟ (scatter plot) มีแนวเป็นเส้นตรงขนานแกนนอนแสดงว่าไม่มีปัญหา nonlinearity ถ้ามีรูปแบบ เช่น เส้นโค้งหรือเส้นแนวโน้มแสดงว่ามีปัญหา nonlinearity

3) การใช้กฎตายตัว (rule of thumb) วิธีนี้ให้ดูจากส่วนเบี่ยงเบนมาตรฐานของตัวแปรตามกับของส่วนเหลือ (e) ถ้าพบว่า SD ของ e มีค่ามากกว่า SD ของ Y แสดงว่ามีปัญหา nonlinearity

4) ความเปลี่ยนแปลงคุณภาพของตัวแบบ วิธีให้เพิ่มตัวแปรอิสระที่ยกกำลังสูงเข้าไปในตัวแบบคือเพิ่ม X^2, X^3 ลงในตัวแบบแล้ววิเคราะห์สมการถดถอย ถ้าพบว่าการเพิ่มตัวแปรดังกล่าวลงในตัวแบบมีผลให้ R^2 สูงขึ้นอย่างเด่นชัดแสดงว่าข้อตกลง linearity ไม่เป็นจริง

2. ภาวะร่วมเส้นตรงพหุ (Multicollinearity)

ปัญหาภาวะร่วมเส้นตรงพหุเกิดจากสถานการณ์ที่ตัวแปรอิสระมีความเกี่ยวข้องกันมากเกินไป ผลที่ตามมาคือทำให้เราไม่อาจแยกได้ว่าตัวแปรอิสระตัวใดกันแน่ที่มีอิทธิพลต่อตัวแปรตามและเนื่องจากตัวแปรอิสระมีความเกี่ยวข้องกันมากจึงมีผลให้ Det $(X^tX) \to 0$ ซึ่งมีผลให้สมาชิกของเมทริกซ์ $(X^tX)^{-1}$มีค่าสูง ผลก็คือ $t_j = \frac{\widehat{\beta}_j}{\sqrt{s^2 (X^tX)^{-1}_{j+1,j+1}}} \to 0$ ตัวแปรอิสระจึงไม่มีนัยสำคัญกันเป็นส่วนมากทั้ง ๆ ที่ความจริงไม่เป็นเช่นนั้น

วิธีตรวจสอบภาวะร่วมเส้นตรงพหุ ทำได้หลายวิธีดังนี้

1) พิจารณาจากนัยสำคัญของตัวแปรอิสระ วิธีนี้พิจารณาง่าย ๆ คือถ้ามีปัญหาภาวะร่วมเส้นตรงพหุตัวแปรอิสระจำนวนมากหรือเกือบทั้งหมดจะไม่มีนัยสำคัญทั้งนี้เพราะค่า t มีแนวโน้มที่จะมีค่าสูง เรื่องนี้ห้ามปล่อยให้ผ่านไปเฉย ๆ เพราะขัดแย้งกับทฤษฎีและวรรณกรรมของเรื่องนั้น

2) พิจารณาจากเครื่องหมายของสัมประสิทธิ์การถดถอย ถ้าพบว่ามีเครื่องหมายผิดจากทฤษฎีหรือวรรณกรรมแสดงว่าอาจเกิดปัญหาภาวะร่วมเส้นตรงพหุ เช่น พบว่าค่าสัมประสิทธ์ของความเครียดในงานซึ่งเป็นตัวแปรอิสระมีเครื่องหมายบวก ตัวแปรตามคือความพึงพอใจในงาน จะเห็นว่าไม่น่าจะเป็นเช่นนั้นเพราะควรจะมีเครื่องหมายลบซึ่งหมายความว่ายิ่งพนักงานมีความเครียดในงานสูงมากเพียงใดความพึงพอใจในงานก็จะน้อยลงมากเพียงนั้น

3) พิจารณาจากความอ่อนไหวของค่าสัมประสิทธิ์การถดถอย ถ้าเพิ่มตัวแปรอิสระขึ้นหรือลดตัวแปรอิสระลงกลับมีผลให้สัมประสิทธ์การถดถอยเปลี่ยนแปลงไปมากแสดงว่าอาจมีปัญหาภาวะร่วมเส้นตรงพหุ

4) พิจารณาจากความสอดคล้องระหว่าง $\text{partial t} - \text{test}$ กับ $\text{overall F} - \text{test}$ ผลการทดสอบทั้ง 2 ต้องสอดคล้องกัน คือถ้า overall F-test คือ Ho: $\beta_1 = \beta_2 = \cdots = \beta_k = 0$ สรุปว่าตัวแปรอิสระทุกตัวมีอิทธิพลต่อตัวแปรตาม หากผลจากการทดสอบสมมุติฐานแบบ partial t-test คือ $H_0: \beta_j = 0 \text{ vs } H_1: \beta_j \neq 0; j = 1,2, \dots, k$ สรุปว่าทุกตัวแปรหรือเกือบทุกตัวแปรไม่มีอิทธิพลต่อตัวแปรตามก็แสดงว่าอาจมีปัญหาภาวะร่วมเส้นตรงพหุ

5) ถ้า det $(X^tX)^{-1} \to 0$ แสดงว่าอาจมีปัญหาภาวะร่วมเส้นตรงพหุ

6) ในสมการ $X_j = f(X_1, X_2, \dots, X_{j-1}, X_{j+}1, \dots, X_k) + u\ ; j = 1, 2, \dots, k$ นั้นถ้า R_j^2 มีค่าสูงมากก็แสดงว่า X_j ถูกสร้างขึ้นมาจากตัวแปรอิสระตัวอื่น โดยทั่วไปเรายอมให้ R_j^2 มีค่าไม่เกิน 0.90 แต่ก็อาจผ่อนปรนได้ถึงไม่เกิน 0.95

7) จากข้อ 6) เราสามารถพัฒนาตัวตรวจสอบเพิ่มขึ้นมาได้อีก 2 ตัวคือ

(1) $\text{Tolerance}_j = 1 - R_j^2$ ซึ่งพบว่าหาก R_j^2 มีค่าเท่ากับ 0.90 จะมีผลให้ Tolerance_j มีค่าเท่ากับ 0.10 ถ้า R_j^2 มีค่าเท่ากับ 0.95 จะมีผลให้ Tolerance_j มีค่าเท่ากับ 0.05 แปลว่าถ้า Tolerance_j มีค่าน้อยกว่า 0.05 แสดงว่าอาจมีปัญหาภาวะร่วมเส้นตรงพหุ

(2) $VIF_j = \frac{1}{1-R_j^2}$ ซึ่งพบว่าหาก R_j^2 มีค่าเท่ากับ 0.90 จะมีผลให้ VIF_j มีค่าเท่ากับ 10 ถ้า R_j^2 มีค่าเท่ากับ 0.95 จะมีผลให้ VIF_j มีค่าเท่ากับ 20 (VIF หมายถึง Variance Inflation factor) แปลว่า ถ้า VIF_j มีค่ามากกว่า 20 แสดงว่าอาจมีปัญหาภาวะร่วมเส้นตรงพหุ

8) Eigen value and condition index โดยปกติ Eigen value ควรมีค่าสูง แต่ถ้ามีค่าต่ำก็ไม่ควรมีค่าใกล้ 0 เนื่องจาก $(X^tX) \sim \ldots \sim \Lambda$ = diag. $[\lambda_1, \lambda_2, \ldots, \lambda_{k+1}]$ ดังนั้นถ้า $\lambda j \rightarrow 0$; some j จะมีผลให้ det $(X^tX) \rightarrow 0$ เรียกว่า near singularity มีผลให้สมาชิกของ $(X^tX)^{-1}$มีค่าสูงและ t_j มีค่าใกล้ 0

จากค่า Eigen เราสามารถตรวจสอบภาวะร่วมเส้นตรงพหุได้จาก condition index ต่อไปนี้คือ

$(\text{condition index})_j = \sqrt{\frac{\max\lambda}{\lambda_j}}$; j=1, 2, …, k+1 โดยที่ถ้า $(\text{condition index})_j > 0.30$ แสดงว่ามีปัญหาภาวะร่วมเส้นตรงพหุที่ตัวแปรอิสระ X_j โดยทั่วไปเราจะพิจารณาร่วมกับ variance proportion (VP) คือถ้า $(\text{condition index})_j > 0.30$ และ $VP_j > 0.90$ แสดงว่ามีปัญหาภาวะร่วมเส้นตรงพหุที่ตัวแปรอิสระ X_j

9) พิจารณาจากค่าสหสัมพันธ์เชิงเส้นระหว่างตัวแปรอิสระคือถ้า $r_{ij} > 0.80\text{-}0.85$ แสดงว่าอาจมีปัญหาภาวะร่วมเส้นตรงพหุที่ตัวแปรอิสระคู่นั้น โดยที่ $r_{ij} = \frac{\sum x_i x_j}{\sqrt{\sum x_i^2 \sum x_j^2}}$

วิธีแก้ปัญหาภาวะร่วมส้นตรงพหุ สามารถกระทำได้ดังนี้

1) ระมัดระวังเรื่องการใช้ตัวแปรหุ่น (dummy variable) คืออย่ากำหนดให้ใช้ตัวแปรหุ่นทุกตัวในขณะที่ในตัวแบบมีเทอมคงที่อยู่ด้วย

2) เพิ่มข้อมูล ถ้าสามารถกระทำได้

3) แปลงค่าตัวแปรอิสระเป็นค่าที่ปรับลบด้วยค่าเฉลี่ย เรียกว่า mean-centered variable หรือ demean หรือแปลงเป็นค่ามาตรฐาน (standardized variable) คือแปลง X_{ji} เป็น $X_{ji} - \bar{X}_j$ หรือแปลง X_{ji} เป็น $\frac{X_{ji}-\bar{X}_j}{s_j}$; j= 1, 2,…, k

4) วิเคราะห์สมการถดถอยด้วย Ridge Regression (RR) วิธี RR เป็นวิธีวิเคราะห์การถดถอยด้วยการทำให้ det (X^tX) มีค่าสูงขึ้นโดยการปรับค่าของเมทริกซ์ Λ = diag.$[\lambda_1, \lambda_2, \ldots, \lambda_{k+1}]$ ให้เป็น Λ = diag. $[\lambda_1+c , \lambda_2+c, \ldots, \lambda_{k+1} +c]$ ค่าคงที่ c เป็นค่าที่ผ่านการวนเวียนคำนวณจนกระทั่งมีผลให้ s^2 มีค่าต่ำที่สุด

5) Principal Component Regression (PCR) วิธีนี้ใช้วิธีสร้างตัวแปรใหม่ที่เกิดจาก linear combination ของตัวแปรอิสระบางส่วน และสร้างตัวแปรใหม่เพิ่มขึ้นอีกจากตัวแปรอิสระที่เหลือ และสร้างเพิ่มขึ้นเรื่อย ๆ ตราบเท่าที่ตัวแปรใหม่เป็นอิสระกับตัวแปรที่ได้สร้างไว้ก่อนหน้านี้

6) Partial Least Square Regression (PLSR) วิธีนี้คล้ายวิธี PCR แตกต่างกันที่วิธีพัฒนา ทั้งวิธี PCR และวิธี PLSR จะสร้างตัวแปรใหม่ขึ้นมาจากกลุ่มตัวแปรอิสระเดิมเรียกตัวแปรใหม่ว่าตัวแปรแฝง (latent variable) ปัญหาที่เกิดก็คือปัญหาการแปลผลเพราะตัวแปรแฝงประกอบขึ้นจากกลุ่มตัวแปรอิสระทำให้ไม่อาจกล่าวได้ว่าตัวแปรตามได้รับผลกระทบจากตัวแปรอะไร

7) ใช้วิธี Differencing หรือ Percentage change วิธีนี้ใช้ได้กับข้อมูลอนุกรมเวลาและมักเกิดปัญหา autocorrelation ตามมาด้วยเสมอ

วิธี differencing ให้ดำเนินการดังนี้คือวิเคราะห์สมการถดถอย

$$Y_t - Y_{t-1} = \beta_1(X_{1t} - X_{1,t-1}) + \beta_2(X_{2t} - X_{2,t-1}) + \cdots + \beta_k(X_{kt} - X_{k,t-1}) + (u_t - u_{t-1})$$

วิธี Percentage change ให้ดำเนินการดังนี้คือวิเคราะห์สมการถดถอย

$$\frac{Y_t - Y_{t-1}}{Y_t} = \beta_1\frac{X_{1t} - X_{1,t-1}}{X_{1t}} + \beta_2\frac{X_{2t} - X_{2,t-1}}{X_{2t}} + \cdots + \beta_k\frac{X_{kt} - X_{k,t-1}}{X_{kt}} + \frac{u_t - u_{t-1}}{u_t}$$

8) ทิ้งตัวแปรอิสระที่ตรวจพบว่าก่อให้เกิดปัญหาภาวะร่วมเส้นตรงพหุ วิธีนี้โดยปกติจะไม่ค่อยนิยมใช้เพราะตัวแปรที่ถูกตัดทิ้งอาจเป็นตัวแปรสำคัญ ที่จริงแล้วตัวแปรอิสระทุกตัวมีความ สำคัญด้วยกันทั้งหมดเพราะมาตามวรรณกรรมเพียงแต่ตัวใดมีอิทธิพลมากก็ถือว่าสำคัญมาก การที่ตัวแปรบางตัวไม่มีนัยสำคัญก็เพราะข้อมูลเชิงประจักษ์ยังไม่สนับสนุนทฤษฎีหรือวรรณกรรมดังกล่าว การตัดสินใจทิ้งตัวแปรไปเองจึงไม่ควรปฏิบัติ

5. Stability of variance (homoscedasticity)

ข้อตกลงนี้คือตกลงว่า V(u) (ซึ่งก็คือ V(Y) นั่นเอง) ต้องคงที่เสมอไม่ว่าเวลาจะเปลี่ยนไปอย่างไรหรือไม่ว่าตัวอิสระจะมีค่าเพิ่มขึ้นหรือลดลงอย่างไร

เราสามารถพิสูจน์ได้ว่า $E(\sum_i^n e_i^2)$ = (n-k-1) σ^2 หรือ $\hat{\sigma}^2$ คือ $s^2 = \frac{\sum_{i=1}^n e_i^2}{n-k-1}$ คือค่าประมาณของ σ^2 กรณีนี้เป็นสถานการณ์ที่ σ^2 คงที่ ซึ่งหาก σ^2 คงที่เราย่อมหาค่าประมาณของ σ^2 ได้ ในทางกลับกันหาก σ^2 ไม่คงที่แต่กลับแปรค่าไปตามเวลาหรือแปรไปตามค่าตัวแปรอิสระเป็น $\sigma_1^2, \sigma_2^2, \ldots$

เช่นในสมการ HB =a + b INC + u โดยที่ HB = household budget INC = household income พบว่า u แปรค่าไปตาม INC ทั้งนี้เพราะยังมีตัวแปรอีกมากที่อยู่ภายนอกตัวแบบ เช่น จำนวนสมาชิกครอบครัว จำนวนผู้มีเงินได้ จำนวนเด็ก อาชีพ จำนวนรถ ลักษณะการครอบครองที่อยู่อาศัย ภูมิภาค และอื่น ๆ รวมกันอยู่ในตัวแปร u และสัมพันธ์กับรายได้ครอบครัว ความผันแปรของ u จึงแปรไปตามตัวแปรเหล่านี้ไม่คงที่ตามข้อตกลง เรียกว่า heteroscedasticity ผลกระทบจากเหตุการณ์นี้คือ s^2 มีค่าเอนเอียงคือต่ำกว่าความเป็นจริง ทำให้ t มีค่าสูงกว่าความเป็นจริงและส่งผลให้ปฏิเสธสมมุติฐานหลักทั้ง ๆ ที่ไม่ควรปฏิเสธ (คือรับตัวแปรอิสระเอาไว้ทั้งที่อาจไม่มีประโยชน์)

การตรวจสอบปัญหา heteroscedasticity อาจกระทำได้หลายวิธีดังนี้

1) ***Levene's test*** ใช้ทดสอบสมมุติฐาน Ho: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_p^2$ โดยที่ σ_i^2 คือความผันแปรของ Y ในกลุ่มตัวอย่างย่อยที่ i ที่เกิดจากการแบ่งกลุ่มค่าสังเกตเป็น p กลุ่ม และ

$$W = \frac{n-p}{p-1}\frac{\sum_i^p n_i(Z_{i.}-Z_{..})^2}{\sum\sum(Z_{ij}-\bar{Z}_{i.})^2} \sim F_{\alpha,\,p-1,\,n-p} \text{ โดยที่ } \bar{Z}_{i.} = \frac{1}{n_i}\sum(e_{ij}-\bar{e}_{i.})^2,\ \bar{Z}_{..} = \frac{1}{n}\sum\sum(e_{ij}-\bar{e}_{..})^2$$

และ $Z_{ij} = |e_{ij} - e_{i.}|$ และในที่ที่เป็น $\bar{e}_{i.}$ และ $\bar{e}_{..}$ เราสามารถใช้เป็น median และ trimmed mean ได้ถ้าข้อมูลแจกแจงแบบ chi-square และแบบ Cauchy ตามลำดับ ที่จริงค่า F ดังกล่าวก็คือค่า F จากตาราง one-way ANOVA

2) ***Goldfeld-Quandt test*** วิธีทดสอบนี้มีขั้นตอนดังนี้

(1) วิเคราะห์สมการถดถอย Y = $\beta 0 + \sum_j^k \beta_j X_j$+u แล้วบันทึกค่า predict ($\hat{Y}$)

(2) เรียงลำดับค่าสังเกตตามค่าของ $\hat{Y}$ (ถือว่า $\hat{Y}$ คือตัวแทนของตัวแปรอิสระทั้งหมดเพราะ $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \cdots + \hat{\beta}_k X_k$) และให้ตัด central observation ทิ้งไปประมาณ 25 % ทำให้แยกค่าสังเกตเป็น 2 กลุ่มคือกลุ่มที่สอดคล้องกับ $\hat{Y}$ ค่ามากกับกลุ่มที่สอดคล้องกับ $\hat{Y}$ ค่าน้อย

(3) วิเคราะห์สมการถดถอย Y = $\beta 0 + \sum_j^k \beta_j X_j$+u กับค่าสังเกตทั้ง 2 ชุด ได้ค่า MSE_1 (คือs_1^2) และ MSE_2 (คือ s_2^2)

(4) คำนวณหาค่า $F_c = \frac{s_1^2}{s_2^2}$ ถ้าพบว่า $F_c > F_{tab}$ ให้สรุปว่ามีปัญหา heteroscedasticity

3) *Glejser test* วิธีนี้มีข้อดีคือใช้รูปแบบ heteroscedasticity เป็นตัวถ่วงน้ำหนักของ WLS ได้ด้วย ขั้นดำเนินการมีดังนี้

(1) วิเคราะห์สมการถดถอย Y = $\beta 0 + \sum_j^k \beta_j X_j$+u แล้วบันทึกค่า residual (e)

(2) วิเคราะห์สมการถดถอย $|e| = m_o + m_1 X_j^c$; c= -1,-½, ½, 1 โดยที่ X_j คือตัวแปรที่มีผลให้ความผันแปรไม่คงที่ (เรียกว่า inflator)

(3) ถ้า m_1 มีนัยสำคัญคือ $m_1 \neq 0$ แสดงว่ามีปัญหา heteroscedasticity ตามรูปแบบนั้นเช่น ถ้าพบว่า m_1 ในสมการ $|e| = m_o + m_1 X_3^{-1}$ มีนัยสำคัญก็ให้สรุปว่ามีปัญหา heteroscedasticity ตามรูปแบบนี้คือ $|e| = m_o + m_1 X_3^{-1}$ และให้ใช้รูปแบบนี้ใน WLS

4) *Park test* มีขั้นตอนดังนี้

(1) วิเคราะห์สมการถดถอย $Y = \beta 0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

(2) วิเคราะห์สมการถดถอย $\ln(e_i^2) = a + b \ln(X_j)$ โดยที่ X_j คือตัวแปรที่มีผลให้ความผันแปรไม่คงที่ (inflator)

(3) ทดสอบสมมุติฐาน $H_o: b = 0$ vs $H_1: b \neq 0$ ถ้า b มีนัยสำคัญก็แสดงว่ามีปัญหา heteroscedasticity

5) ***White test*** มีขั้นตอนดังนี้

(1) วิเคราะห์สมการถดถอย $Y = \beta 0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

สมมุติว่า k = 2

(2) วิเคราะห์สมการถดถอย

$$e_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i} + v_i$$

แล้วคำนวณหาค่าสถิติ $\chi_c^2 = nR^2$ สมมุติฐานหลักคือ

$$Ho: \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = 0$$

(3) ถ้า $\chi_c^2 > \chi_{5,\alpha}^2$ ให้สรุปว่ามีปัญหา heteroscedasticity

วิธีนี้จะประสบข้อยุ่งยากบ้างถ้ามีตัวแปรอิสระมากรวมทั้งมีตัวแปรหุ่น (dummy variable) กรณีมีตัวแปรหุ่น เช่น $D_i = 0, 1$ ก็ไม่ต้องกำหนดให้มี D_i^2 ร่วมเป็นตัวแปรอิสระเพราะมีค่าเท่ากัน หากมี D_i^2 ด้วย จะเป็นสาเหตุให้เกิดปัญหาภาวะร่วมเส้นตรงพหุ

6) *Breusch-Pagan-Godfrey test* วิธีนี้เหมาะสำหรับตัวอย่างขนาดใหญ่ มีขั้นตอนดังนี้

(1) วิเคราะห์สมการถดถอย $Y = \beta 0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

(2) วิเคราะห์สมการถดถอย

$$\frac{e_i^2}{s^2} = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_3 + \cdots + \alpha_m X_m + v_i$$

โดยที่ $X_1, X_2, X_3, \ldots, X_m$ คือตัวแปรอิสระที่อาจเป็นสาเหตุของปัญหา heteroscedasticity

(3) คำนวณหาค่า $\chi^2_c = \frac{\text{sum of square regression}}{2}$ ถ้าพบว่า $\chi^2_c > \chi^2_{m,\alpha}$ แสดงว่ามีปัญหา heteroscedasticity

7) *Ramsey test* มีขั้นตอนดังนี้

(1) วิเคราะห์สมการถดถอย Y = $\beta_0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

(2) วิเคราะห์สมการถดถอย

$$e_i^2 = \alpha_0 + \alpha_1 \widehat{Y}^2 + \alpha_2 \widehat{Y}^3 + \cdots + \alpha_m \widehat{Y}^{m+1} + v_i$$

(3) คำนวณหาค่าสถิติ $\chi^2_c = nR^2$ และปฏิเสธสมมุติฐานหลักคือ

Ho: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \ldots = \alpha_m = 0$ ถ้าพบว่า $\chi^2_c > \chi^2_{m,\alpha}$

8) ***Keonker test*** วิธีนี้เหมือน Ramsey test เพียงแต่ใช้ตัวแปรอิสระอื่นนอกตัวแบบ คือใช้ $Z_1, Z_2, Z_3, \ldots, Z_m$ มาเป็นเครื่องมือช่วยทดสอบ ขั้นตอนมีดังนี้

(1) วิเคราะห์สมการถดถอย Y = $\beta_0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

(2) วิเคราะห์สมการถดถอย

$$e_i^2 = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \cdots + \alpha_m Z_m + v_i$$

(3) คำนวณหาค่าสถิติ $\chi^2_c = nR^2$ และปฏิเสธสมมุติฐานหลักคือ

Ho: $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \ldots = \alpha_m = 0$ ถ้าพบว่า $\chi^2_c > \chi^2_{m,\alpha}$

9) *Autoregressive Condition Heteroscedastic* (ARCH) วิธีนี้ตรวจสอบทั้งปัญหา heteroscedasticity และ autocorrelation ในขณะเดียวกัน มีขั้นตอนดังนี้

(1) วิเคราะห์สมการถดถอย Y = $\beta_0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

(2) วิเคราะห์สมการถดถอย

$$e_t = \alpha_0 + \alpha_1 e_t^2 + \alpha_2 e_{t-1}^2 + \alpha_3 e_{t-3}^2 + \cdots + \alpha_m e_{t-m}^2 + v_i$$

แล้วคำนวณหาค่าสถิติ $\chi^2_c = nR^2$ ถ้า $\chi^2_c > \chi^2_{m,\alpha}$ ให้ปฏิเสธสมมุติฐานหลักคือ

Ho: $\alpha_1 = \alpha_2 = \alpha_3 = \cdots = \alpha_m = 0$ สรุปว่ามีปัญหา heteroscedasticity และ autocorrelation

วิธีแก้ปัญหา Heteroscedasticity

การแก้ปัญหา Heteroscedasticity อาจกระทำได้หลายวิธีต่อไปนี้

1) วิธีแปลงข้อมูล วิธีที่นิยมใช้คือวิธี log-log model วิธีปฏิบัติคือแปลงข้อมูลทั้งตัวแปรอิสระและตัวแปรตามคือ แปลง *Y* เป็น ln Y และแปลง X_j เป็น $\ln(X_j)$; j = 1, 2, …, k แล้ววิเคราะห์สมการถดถอย lnY = $\beta_0 + \sum_j^k \beta_j \ln(X_j) + u$

***2) วิธี* WLS** วิธีนี้จะกระทำได้ต่อเมื่อได้ทราบรูปแบบของ heteroscedsticity เสียก่อนซึ่งวิธีตรวจสอบแบบ Glejser test จะช่วยให้สามารถระบุรูปแบบได้

สมมุติตรวจพบว่ารูปแบบของ heteroscedasticity คือ $|e| = 3 + 2X_3^{-1}$ ดังนั้น weight matrix คือ Diag.[$(3+2X_{31}^{-1})^2, (3+2X_{32}^{-1})^2, (3+2X_{33}^{-1})^2, \ldots, (3+2X_{3n}^{-1})^2$] และค่าประมาณของ β คือ $\hat{\beta} = (X'W^{-1}X)^{-1}X'W^{-1}Y$

3) วิธี *Box-Cox Transformation* วิธีนี้จะแปลง Y เป็น $g(Y,\lambda_1,\lambda_2)$ โดยที่

$$g(Y,\lambda_1,\lambda_2) = \frac{(Y+\lambda_2)^{\lambda_1}-1}{\lambda_1}$$

รูปแบบการแปลงนี้ย่อเป็นรูปแบบต่าง ๆ ได้หลายแบบดังนี้

(1) ถ้า $\lambda_1 \to 0, \lambda_2 = 1 \quad g(Y,\lambda_1,\lambda_2) = \frac{(Y+\lambda_2)^{\lambda_1}-1}{\lambda_1} \to \log(Y+1)$

(2) ถ้า $\lambda_1 \to 0 \quad g(Y,\lambda_1,\lambda_2) = \frac{(Y+\lambda_2)^{\lambda_1}-1}{\lambda_1} = \log(Y+\lambda_2)$

(3) ถ้า $\lambda_1 = ½, \lambda_2 = 0 \quad g(Y,\lambda_1,\lambda_2) = \frac{(Y+\lambda_2)^{\lambda_1}-1}{\lambda_1} = \frac{\sqrt{Y}-1}{½}$

วิธีที่ (1) และ (2) เหมาะสำหรับกรณีข้อมูลติดลบ กรณีที่ (3) เหมาะสำหรับกรณีมีค่าเป็น 0

4) วิธีถ่วงน้ำหนัก วิธีนี้จะใช้ X_j หรือ $\sqrt{X_j}$ ที่เป็น inflator หรือใช้ตัวแปรภายนอกคือ *Z* ที่เป็น inflator เป็นตัวถ่วงน้ำหนักแล้ววิเคราะห์สมการถดถอยตามปกติ เช่นจากสมการถดถอย $Y = \beta 0 + \sum_j^k \beta_j X_j + u$ ถ้าพบว่า inflator คือ X_2 และ Z เราอาจแก้ปัญหา heteroscedasticity โดยการวิเคราะห์การถดถอย

(1) $\frac{Y}{X_2} = \frac{\beta 0}{X_2} + \sum_j^k \beta_j \frac{X_j}{X_2} + u$ หรือ

(2) $\frac{Y}{\sqrt{X_2}} = \frac{\beta 0}{\sqrt{X_2}} + \sum_j^k \beta_j \frac{X_j}{\sqrt{X_2}} + u$ หรือ

(3) $\frac{Y}{Z} = \frac{\beta 0}{Z} + \sum_j^k \beta_j \frac{X_j}{Z} + u$

ปัญหา Autocorrelation

Autocorrelation คือสถานการณ์ที่ error term ไม่เป็นอิสระต่อกันในระหว่างเวลาคือ $E(u_i u_j) \neq 0$ โดยเราจะตรวจสอบกับข้อมูลที่เป็นอนุกรมเวลาเท่านั้น และในการตรวจสอบปัญหา autocorrelation เรานิยมตรวจสอบตามรูปแบบของ autoregressive process of order 1: AR(1) คือ $u_t = \rho_1 u_{t-1} + v_t$ หรือ autoregressive process of order 2: AR(2) คือ $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t$ เท่านั้น

ผลกระทบจากปัญหา autocorrelation คือ $V(\hat{\beta}) \neq \sigma^2(X^tX)^{-1}$ และ s^2 อาจมีค่าต่ำกว่าความเป็นจริงหรือสูงกว่าความเป็นจริงก็ได้ขึ้นอยู่กับว่าเป็นสหสัมพันธ์เชิงบวกหรือว่าเชิงลบ ผลกระทบทั้ง 2 ประการนี้ก่อให้เกิดความเสียหายมากเพราะทำให้เราไม่อาจเชื่อถือในผลทดสอบสมมุติฐาน กล่าวคือ จาก H_o:$\beta_j = 0$ vs H_1: $\beta j \neq 0$ ตัวทดสอบคือ

$$t_j = \frac{\hat{\beta}_j}{\sqrt{s^2(X^tX)^{-1}_{j+1,j+1}}} \;;\; j = 1, 2, \ldots, k$$

ปัญหา autocorrelation ทำให้โครงสร้างสูตร t เสียไปเพราะสูตร $V(\hat{\beta}) \neq \sigma^2(X^tX)^{-1}$คือมีรูปแบบอื่น ทั้งนี้เพราะ s^2 มีรูปแบบอื่น s^2 อาจมีค่าสูงกว่าที่ควรจะเป็นมีผลให้ t มีค่าต่ำซึ่งอาจต่ำมากจนถึงระดับทำให้ยอมรับ H_o:$\beta_j = 0$ คือปฏิเสธตัวแปรทิ้งๆ ที่เป็นตัวแปรสำคัญ หรือ s^2 อาจต่ำกว่าความเป็นจริง ซึ่งมีผลให้ t_j สูง อาจสูงมากจนถึงทำให้ปฏิเสธสมมุติฐานหลัก คือยอมรับเอาตัวแปรอิสระไว้ทิ้งๆ ที่อาจเป็นตัวแปรที่ไม่มีความสำคัญ สรุปว่ากระบวนการพิจารณานัยสำคัญของตัวแปรผิดพลาดรวนไปทั้งหมด เรื่องผิดปกตินี้นักวิจัยจะไม่รู้ตัวเลยเพราะก็ใช้สูตรตามปกติคือ $s^2 = \frac{1}{n-k-1}\sum_i^n e_i^2$ แต่ที่จริงแล้วเมื่อเกิดปัญหา autocorrelation สูตรของ s^2 จะเป็นรูปอื่น

***วิธีตรวจสอบ* autocorrelation**

วิธีตรวจสอบ autocorrelation กระทำได้หลายวิธีดังนี้

1) ***Durbin-Watson test (DW)***

วิธี DW ใช้ทดสอบเฉพาะกรณีที่เป็น AR(1) คือ $u_t = \rho_1 u_{t-1} + v_t$ วิธีดำเนินการมีดังนี้

(1) วิเคราะห์สมการถดถอย $Y = \beta 0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

(2) คำนวณหาค่าสถิติ $DW = \frac{\sum_t^n (e_t - e_{t-1})^2}{\sum_t^n e_t^2}$ ทั้งนี้ DW สามารถลดรูปเป็น $DW = 2 - 2\rho$ ได้และเมื่อแทนค่า ρ ด้วยค่าที่กำหนดจะปรากฏกฎง่าย ๆ ดังนี้พบว่า $DW \cong 2$ (เมื่อแทนที่ ρ ด้วย 0 จะทำให้ DW = 2) แสดงว่าไม่มีปัญหา autocorrelation แต่ถ้า $DW \cong 0$ (เมื่อแทนที่ ρ ด้วย 1จะทำให้ DW = 0) แสดงว่ามี perfect positive autocorrelation ถ้า $DW \cong 4$ (เมื่อแทนที่ ρ ด้วย -1จะทำให้ DW = 4) แสดงว่ามีปัญหา perfect negative autocorrelation ถ้า DW มีค่าอื่นให้เปิดตารางเพื่อหาค่าวิกฤติ

หรืออาจวิเคราะห์ดังต่อไปนี้คือ

(3) วิเคราะห์สมการถดถอย $e_t = \rho e_{t-1} + v_t$ แล้วทดสอบ H_o: $\rho = 0$ vs H_1: $\rho \neq 0$ ถ้าปฏิเสธ Ho ก็แสดงว่ามีปัญหา autocorrelation ใน AR(1)

2) *Breusch-Godfrey test (BG)* วิธีนี้ใช้ตรวจสอบ autocorrelationในกรณี AR (p) โดยถือว่าเป็นกรณีทั่วไปที่ u_t อาจมีความ สัมพันธ์กันเองในหลายช่วงเวลา ในกรณีเช่นว่านี้ DW จะไม่อาจรองรับได้ วิธี BG ดำเนินการดังนี้

(1) วิเคราะห์สมการถดถอย $Y = \beta 0 + \sum_j^k \beta_j X_j + u$ แล้วบันทึกค่า residual (e)

(2) วิเคราะห์สมการถดถอย

$$e_t = \alpha_1 X_{1t} + \alpha_2 X_{2t} + \cdots + \alpha_k X_h + \rho_1 e_{t-1} + \rho_2 e_{t-2} + \cdots + \rho_p e_{t-p} + v_t$$

แล้วคำนวณหาค่า $\chi_c^2 = nR^2$ ถ้าพบว่า $\chi_c^2 > \chi_{p,\alpha}^2$ ให้ปฏิเสธ Ho: $\rho_1 = \rho_2 = \cdots = \rho_p = 0$ แสดงว่ามีปัญหา autocorrelation ใน AR (p)

หรือทดสอบสมมุติฐานจากวิธีการในข้อ (3) คือ

(3) วิเคราะห์สมการถดถอย $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \cdots + \rho_p e_{t-p} + v_t$ แล้วทดสอบสมมุติฐาน Ho: $\rho_1 = \rho_2 = \cdots = \rho_p = 0$ ด้วยตาราง ANOVA

วิธีแก้ปัญหา autocorrelation

การแก้ปัญหา autocorrelation ให้แก้ตามรูปแบบที่ตรวจพบกล่าวคือ

(1) ถ้าใช้ DW ซึ่งเป็นเพียงการตรวจสอบในรูปแบบ AR(1) ให้แปลงข้อมูลดังนี้คือแปลง Y_t เป็น $Y_t - \hat{\rho}_1 Y_{t-1}$ และแปลง X_{jt} เป็น $X_{jt} - \hat{\rho}_1 X_{j,t-1}$; j = 1, 2, …, k แล้ววิเคราะห์สมการถดถอย

$$Y_t - \hat{\rho}_1 Y_{t-1} = \sum_j^k \beta_j (X_{jt} - \hat{\rho}_1 X_{j,t-1}) + v_t$$

โดยที่ $\hat{\rho}_1$; j= 1, 2…, p ได้มาจากการวิเคราะห์การถดถอย $e_t = \rho_1 e_{t-1} + v_t$

(2) ถ้าพบว่าเป็น AR(2) คือ $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t$ ให้แปลงค่าตัวแปร Y_t เป็น $Y_t - \hat{\rho}_1 Y_{t-1} - \hat{\rho}_2 Y_{t-2}$ และแปลง X_{jt} เป็น $X_{jt} - \hat{\rho}_1 X_{j,t-1} - \hat{\rho}_2 X_{j,t-2}$; j = 1, 2, …, k แล้ววิเคราะห์สมการถดถอย

$$Y_t - \hat{\rho}_1 Y_{t-1} = \sum_j^k \beta_j (X_{jt} - \hat{\rho}_1 X_{j,t-1} - \hat{\rho}_2 X_{j,t-2}) + v_t$$

โดยที่ $\hat{\rho}_1$และ$\hat{\rho}_2$ได้มาจากการวิเคราะห์การถดถอย $e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + v_t$

เอกสารอ้างอิง

มนตรี พิริยะกุล. ***เทคนิคการวิเคราะห์สมการถดถอย***. กรุงเทพมหานคร :มหาวิทยาลัยรามคำแหง, 2544.

CHAPTER 3 MISSPECIFICATION TESTS *ค้นเมื่อ 15 มกราคม 2557จาก* http://www.tnet.teagasc.ie/fapri/downloads/rochemanual/mrochech3.pdf.

Choiv, Syngjoo (2008), *Environmental Econometrics*, *ค้นเมื่อ 10 มกราคม 2557 จาก* http://www.homepages.ucl.ac.uk/~uctpsc0/Teaching/GR03/SLRM.pdf.

Garson, D. G. (2012), ***Testing Statistical Assumption***, North Carolina State University.

Heteroscedasticity, *ค้นเมื่อ 14 มกราคม 2557* จาก http://www.acc.ncku.edu.tw/chinese/faculty/mingyuan/myweb11/EnClass/Slides/Chapter5-20060928.ppt.

Hoffmann, John P.(2010) ***Linear Regression Analysis: Applications and Assumptions***, Second Edition, Electronic Production: Adobe Acrobat Professional.

Schwarz, Carl James (2013)**, *Sampling, Regression, Experimental Design and Analysis for Environmental Scientists, Biologists, and Resource Managers***, Department of Statistics and Actuarial Science, Simon Fraser University.

.