

การประยุกต์ใช้ตัวแบบถดถอยโทบิต-พีชไวส์ กับข้อมูลด้านเศรษฐศาสตร์ ที่มีปัญหาค่านอกเกณฑ์

Application of Tobit-piecewise Regression in
Economics Data Consisting of Outliers

ธิวัรัตน์ ทิพรส*

Titirut Thipbharos*

* ผู้ช่วยศาสตราจารย์ กลุ่มวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ประยุกต์ มหาวิทยาลัยบูรพา

* Assistant Professor, Section of Mathematics and Statistics, Faculty of Applied Science,
Dhurakij Pundit University

* Email: titirut.thi@dpu.ac.th

บทคัดย่อ

บทความวิชาการฉบับนี้กล่าวถึงการประมาณค่าพารามิเตอร์ในตัวแบบถดถอยโทบิต-พีชไวลส์ ที่พัฒนาขึ้นโดย Mekbunditkul (2010) เพื่อจัดการกับปัญหาค่านอกเกณฑ์ที่เกิดขึ้นในการวิเคราะห์การถดถอย การประมาณค่าพารามิเตอร์ในตัวแบบถดถอยนี้ใช้วิธีการประมาณแบบภาวะน่าจะเป็นสูงสุด ดังนั้นตัวประมาณที่สร้างได้จะมีคุณสมบัติคงเส้นคงวาและปกติเมื่อใกล้อนันต์ (Consistency and Best asymptotically normal) ในบทความนี้ได้มีการกล่าวถึงการแปลผลค่าสัมประสิทธิ์การถดถอยในตัวแบบถดถอยโทบิต-พีชไวลส์ ที่สามารถแปลความหมายได้สองประเด็น ได้แก่ การเปลี่ยนแปลงค่าของตัวแปรตามและการเปลี่ยนแปลงความน่าจะเป็นของตัวแปรตาม เมื่อตัวแปรอิสระเปลี่ยนแปลงไปหนึ่งหน่วย อันจะก่อให้เกิดประโยชน์ต่อการประยุกต์ใช้มากยิ่งขึ้น อีกทั้งได้มีการกล่าวถึงตัวอย่างงานวิจัยทางด้านเศรษฐศาสตร์ที่ข้อมูลมีขีดจำกัดบน และ/หรือขีดจำกัดล่าง หรือข้อมูลมีค่านอกเกณฑ์ และแนวทางการนำตัวแบบถดถอยโทบิต-พีชไวลส์ไปประยุกต์ใช้กับข้อมูลจริง เพื่อให้เห็นประโยชน์ของตัวแบบถดถอยโทบิต-พีชไวลส์ได้อย่างชัดเจน

คำสำคัญ : การวิเคราะห์การถดถอยโทบิต-พีชไวลส์ การวิเคราะห์การถดถอยโทบิต ตัวแบบถดถอยพีชไวลส์ ขีดจำกัด ของตัวแปรตามค่านอกเกณฑ์ในการวิเคราะห์การถดถอย

Abstract

This article was presented the estimation of Tobit-piecewise regression model. This model was developed by Mekbunditkul (2010) in order to cope with regression outliers problem. The maximum likelihood estimation (MLE) was employed thus the parameters estimate keep all properties of MLE such as consistency and best asymptotically normal. The interpretation of regression coefficients in Tobit-piecewise regression for application was explained in the sense of unconditional marginal effect and conditional marginal effect. Moreover, literatures related to the economics data, which dependent variable has been limited or the data consist of outliers, was demonstrated. The application of Tobit-piecewise regression was suggested in this paper in order to point out the contribution of this model.

Keywords: Tobit-Piecewise Regression, Tobit Regression, Piecewise Regression, Limitation in Dependent Variable, Regression Outliers

บทนำ

เป็นที่ทราบกันดีว่าเมื่อผู้วิจัยต้องการวิเคราะห์อิทธิพลเชิงเส้นของตัวแปรอิสระ (Independent variables) ที่มีต่อตัวแปรตาม (Dependent variable) โดยที่ตัวแปรอิสระและตัวแปรตามเป็นตัวแปรเชิงปริมาณชนิดต่อเนื่อง (Continuous variable) ผู้วิจัยจะทำการวิเคราะห์การถดถอยเชิงเส้นด้วยวิธีกำลังสองน้อยที่สุด (Ordinary least square (OLS) linear regression analysis) เพื่อใช้ในการทำนายค่าของตัวแปรตามโดยอาศัยลักษณะความสัมพันธ์เชิงเส้นที่สร้างขึ้นได้ ซึ่งตัวแบบการถดถอย (Regression model) สามารถเขียนได้ดังนี้

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i, i=1, \dots, n$$

เมื่อ Y_i คือตัวแปรตาม $x_{1i}, x_{2i}, \dots, x_{ki}$ เป็นตัวแปรอิสระตัวที่ 1 จนถึงตัวที่ k และ $\beta_0 + \beta_1 + \beta_2, \dots, \beta_k$ เป็นสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า สำหรับ ε_i แทนความคลาดเคลื่อนสุ่ม โดยมีข้อสมมติ (Assumption) ว่าความคลาดเคลื่อนจะต้องมีการแจกแจงความน่าจะเป็นปกติ มีค่าเฉลี่ยเป็น 0 และความแปรปรวนคงที่ หรือกล่าวได้ว่า $\varepsilon_i \sim \text{i.i.d.N}(0, \sigma^2)$ ทุกค่าของ $i = 1, \dots, n$ และจะได้ว่าลักษณะการแจกแจงความน่าจะเป็นของตัวแปรตาม เป็นแบบปกติที่มีค่าเฉลี่ยเป็น $\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$ และความแปรปรวนคงที่หรือกล่าวได้ว่า $Y_i \sim \text{i.i.d.N}[(\beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki}), \sigma^2]$ ทุกค่าของ $i = 1, \dots, n$ โดยที่ตัวแปรตาม Y_i เป็นสมาชิกของจำนวนจริง ($Y_i \in \mathbb{R}$) ซึ่งภายใต้ข้อสมมติดังกล่าว ตัวประมาณกำลังสองน้อยที่สุด (Ordinary least squares estimator) จะมีคุณสมบัติเป็นตัวประมาณเชิงเส้นไม่เอนเอียงที่ดีที่สุด (Best linear unbiased estimator) อย่างไรก็ตาม ในกรณีที่ผู้วิจัยมีการละเมิดข้อสมมติ ดังกล่าวไม่ว่าจะเป็น ความคลาดเคลื่อนไม่มีการแจกแจงแบบปกติ หรือความแปรปรวนไม่คงที่นั้น จะทำให้ตัวประมาณกำลังสองน้อยที่สุดนั้นเป็นตัวประมาณที่ไม่ดีพอทำให้สมการถดถอยที่สร้างได้นั้นไม่เหมาะสมในการทำนายค่าของตัวแปรตามที่ผู้วิจัยต้องการศึกษา

นอกจากนั้นยังมีปัญหาที่สำคัญอีกประการหนึ่งของการวิเคราะห์การถดถอยคือ การเกิดปัญหาค่านอกเกณฑ์ (Outliers) ซึ่ง Rousseeuw and Leroy (1987) ได้กล่าวว่า ค่านอกเกณฑ์ในตัวแปรตาม (*y-direction outliers*) ค่านอกเกณฑ์ในตัวแปรอิสระ (*x-direction outliers*) จะทำให้เกิดปัญหาความแปรปรวนไม่คงที่ได้ นอกจากนั้น ซึ่งเมื่อเกิดปัญหาเช่นนี้แล้วผลที่ตามมาหากผู้วิจัยใช้ตัวประมาณกำลังสองน้อยที่สุดคือ สมการถดถอยที่ได้ไม่แสดงความสัมพันธ์เชิงเส้นที่ดี บ่อยครั้งที่เมื่อเกิดปัญหาค่านอกเกณฑ์ขึ้นนั้นผู้วิจัยส่วนใหญ่ใช้วิธีการตัดค่านอกเกณฑ์เหล่านั้นทิ้ง ซึ่งวิธีการดังกล่าวจะทำให้ผู้วิจัยได้ผลลัพธ์ที่คลาดเคลื่อนอันเนื่องมาจากค่านอกเกณฑ์ที่ตัดทิ้งนั้นอาจเป็นส่วนหนึ่งของประชากร ไม่ใช่ค่านอกเกณฑ์ที่เป็นค่าที่เกิดจากความผิดพลาด (Bad outliers)

ดังนั้น การจัดการกับค่านอกเกณฑ์นั้นมักมีกลวิธีพัฒนาขึ้นหลายวิธีซึ่งแต่ละวิธีนั้นใช้หลักการเดียวกันคือ ไม่ตัดค่านอกเกณฑ์นั้นทิ้ง แต่พยายามลดอิทธิพลของค่านอกเกณฑ์นั้นอย่างสมเหตุสมผล เพื่อให้สมการถดถอยนั้นมีความเหมาะสมกับข้อมูลมากที่สุด ซึ่งหนึ่งในนั้นคืองานวิจัยของ Mekbunditkut (2010) ที่ได้พัฒนาตัวแบบถดถอยโทบิต-พีซไวส์ (Tobit-piecewise regression) เพื่อลดอิทธิพลของค่านอกเกณฑ์โดยวิธีดังกล่าวมีความเข้าใจได้ง่ายกว่าตัวแบบถดถอยแกร่ง (Robust regression) แบบอื่น ตัวแบบถดถอยโทบิต-พีซไวส์นี้ถูกพัฒนาขึ้นภายใต้แนวคิด 2 แนวคิด

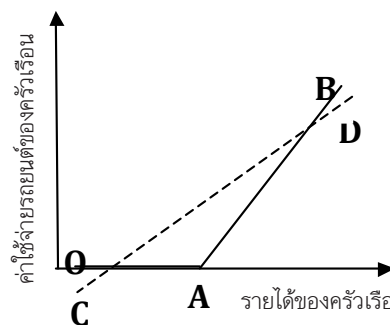
คือ แนวคิดของตัวแบบถดถอยโทบิต (Tobit regression) และแนวคิดของตัวแบบถดถอยพีซไวส์ (Piecewise regression) ซึ่งได้อธิบายไว้ในหัวข้อที่ 2 สำหรับหัวข้อที่ 3 นั้นผู้เขียนได้กล่าวถึงการแปลงผลค่าพารามิเตอร์ในตัวแบบถดถอยโทบิต-พีซไวส์ และหัวข้อที่ 4 เป็นการยกตัวอย่างการประยุกต์ใช้ตัวแบบถดถอยโทบิต-พีซไวส์ ในเชิงเศรษฐศาสตร์ และด้านอื่นๆ สุดท้ายคือหัวข้อที่ 5 นั้นเป็นแนวทางการนำตัวแบบถดถอยโทบิต-พีซไวส์ไปประยุกต์ใช้

แนวคิดพื้นฐาน

ตัวแบบถดถอยโทบิต-พีซไวส์ได้พัฒนาขึ้นจากตัวแบบ 2 ตัวแบบ คือตัวแบบถดถอยโทบิต และตัวแบบถดถอยพีซไวส์ เพื่อใช้ในการผลกระทบของค่านอกเกณฑ์ที่เกิดขึ้นในการวิเคราะห์การถดถอย ผู้ที่พัฒนาขึ้นคนแรกคือ Mekbunditkul (2010) โดยผู้พัฒนาตัวแบบมีแนวคิดในการจำกัดค่านอกเกณฑ์เพื่อลดผลกระทบของค่านอกเกณฑ์ด้วยการกำหนดให้ค่านอกเกณฑ์มีค่าเท่ากับค่าขีดจำกัดบน และ/หรือขีดจำกัดล่าง ด้วยตัวแบบถดถอยโทบิต สำหรับการพัฒนาตัวแบบร่วมกับตัวแบบถดถอยพีซไวส์นั้นก็เพื่อให้ขีดจำกัดบน และ/หรือขีดจำกัดล่างนั้นไม่เป็นค่าคงที่ตลอดช่วงของข้อมูล และตัวแบบถดถอยพีซไวส์นี้สามารถลดอิทธิพลของค่านอกเกณฑ์ในตัวแปรอิสระในบางกรณีได้ด้วย ดังนั้นผู้เขียนจึงได้อธิบายแนวคิดพื้นฐานของแต่ละตัวแบบไว้ดังต่อไปนี้

1. ตัวแบบถดถอยโทบิต (Tobit regression)

ตัวแบบถดถอยโทบิตนั้นเป็นตัวแบบที่นักเศรษฐศาสตร์ชื่อ Tobin (1958) พัฒนาขึ้นมาเพื่อใช้กับข้อมูลที่ตัวแปรตามมีลักษณะที่มีขีดจำกัดบน และ/หรือขีดจำกัดล่าง ซึ่งพบบ่อยในข้อมูลที่แสดงคุณลักษณะทางเศรษฐศาสตร์ เช่น ในกรณีที่ผู้วิจัยต้องการศึกษาปัจจัยที่มีอิทธิพลต่อค่าใช้จ่ายในการซื้อรถยนต์ในช่วงปี 2557 ซึ่งผู้วิจัยจะพบว่ามียกเว้นค่าสังเกตจำนวนมากที่มีค่าเป็น 0 บาท เนื่องจากไม่ได้ซื้อรถยนต์ในปีดังกล่าว ซึ่งผู้วิจัยไม่สามารถตัดค่าสังเกตเหล่านี้ทิ้งได้ ทั้งนี้เนื่องจากแม้ว่าในช่วงปีดังกล่าวตัวอย่างที่ทำการสำรวจไม่ได้มีการซื้อรถยนต์ (ค่าใช้จ่ายในการซื้อรถยนต์เป็น 0 บาท) แต่ไม่ได้หมายความว่าไม่เคยซื้อรถยนต์เลย ดังนั้น การตัดค่าสังเกตเหล่านั้นทิ้งจึงไม่ถูกต้อง ซึ่งเมื่อเกิดเหตุการณ์ดังกล่าวนี้หากผู้วิจัยใช้การวิเคราะห์การถดถอยเชิงเส้นเชิงเส้นด้วยวิธีกำลังสองน้อยที่สุดจะทำให้ได้ตัวประมาณที่มีเอนเอียง (Biased estimator) พิจารณาภาพที่ 1 จากจุด O ถึงจุด A เป็นข้อมูลของครัวเรือนที่ไม่มีค่าใช้จ่ายรถยนต์ในปีที่ทำการศึกษา



ภาพที่ 1 ความสัมพันธ์ของรายได้กับค่าใช้จ่ายรถยนต์ของครัวเรือน

จะเห็นว่าเส้นแองเจิล (Engel curve) จะถูกแบ่งออกเป็น 2 ส่วน จากเส้น OA และ AB ซึ่งเห็นได้ชัดจนกว่าค่าใช้จ่ายรถยนต์ของครัวเรือนไม่สามารถอธิบายได้ด้วยสมการถดถอยเชิงเส้นที่สร้างด้วยวิธีกำลังสองน้อยที่สุด (แทนด้วยเส้นปะ CD) เนื่องจากสมการถดถอยเชิงเส้น CD มีค่าประมาณความชันต่ำเกินไป (Underestimate) ทั้งที่แท้จริงแล้วค่าความชันเป็นดังเส้นตรง AB เมื่อข้อมูลมีลักษณะดังที่กล่าวมาในปี ค.ศ.1958 Tobin จึงได้เสนอวิธีการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามเมื่อตัวแปรตามมีขีดจำกัด (Limited dependent variable) ต่อมาวิธีดังกล่าวถูกเรียกว่า การวิเคราะห์การถดถอยโทบิต (Tobit regression analysis) โดยมีตัวแบบถดถอยโทบิต เป็นตัวแบบแสดงความสัมพันธ์เชิงเส้น

ตัวแบบถดถอยโทบิต (Tobit regression model) เขียนได้ดังนี้

$$Y_i = \begin{cases} L & ; Y_i^* \leq L, \\ Y_i^* & ; Y_i^* > L, \end{cases} \quad (1)$$

เมื่อ Y_i^* , $i=1, \dots, n$ แทนฟังก์ชันเชื่อมโยง (link function) ที่สร้างขึ้นโดยสมการถดถอยเชิงเส้น

$$Y_i^* = \beta_0 + \beta_1 x_i + \dots + \beta_k x_{ki} + \varepsilon_i \quad (2)$$

โดยที่ $x_{1i}, x_{2i}, \dots, x_{ki}$ เป็นตัวแปรอิสระตัวที่ 1 จนถึงตัวที่ k พารามิเตอร์ $\beta_0 + \beta_1 + \beta_2, \dots, \beta_k$ เป็นสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า และ ε_i แทนความคลาดเคลื่อนสุ่ม โดยมีข้อสมมติ (assumption) เช่นเดียวกับการวิเคราะห์การถดถอย นั่นคือ $\varepsilon_i \sim i.i.d.N(0, \sigma^2)$ ทุกค่าของ $i=1, \dots, n$

2. ตัวแบบถดถอยพีซไวส์ (Piecewise regression)

Quandt (1958) ได้พัฒนาตัวแบบถดถอยพีซไวส์ขึ้นเพื่อใช้ในกรณีที่เส้นความสัมพันธ์เชิงเส้นเพียงเส้นเดียว (Single regression line) นั้นไม่เหมาะสมกับข้อมูล ทั้งนี้เนื่องจากความคลาดเคลื่อนมีการแจกแจงที่แตกต่างกัน ตัวแบบถดถอยพีซไวส์ในกรณีที่ประกอบไปด้วยเส้นถดถอย 2 เส้น เขียนได้ดังนี้

$$Y_i = \begin{cases} \alpha_1 + \beta_1 x_i + \varepsilon_i & ; x_i \leq \delta \\ \alpha_2 + \beta_2 x_i + \varepsilon_i & ; x_i > \delta \end{cases} \quad (3)$$

เมื่อ Y_i เป็นตัวแปรตาม X_i เป็นตัวแปรอิสระ $\alpha_1, \alpha_2, \beta_1, \beta_2$ เป็นสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า และ ε_i แทนความคลาดเคลื่อนสุ่มที่มีคุณสมบัติ $\varepsilon_i \sim i.i.d.N(0, \sigma^2)$ ทุกค่าของ $i = 1, \dots, n$

สมการที่ (3) สามารถเขียนอยู่ในรูปตัวแบบถดถอยเชิงเส้นที่มีตัวแปรหุ่น (Dummy variable)

ได้ตั้งสมการที่ (4)

$$Y_i = \alpha_1 + \beta_0 + \beta_1 x_i + \beta_2 (x_i - \delta) D_i + \varepsilon_i \quad (4)$$

$$\text{เมื่อ } x_i^* = (x_i - \delta) D_i \text{ และ } D_i = \begin{cases} 0 & ; x_i \leq \delta \\ 1 & ; x_i > \delta \end{cases}$$

3. ตัวแบบถดถอยโทบิต-พีซไวส์ (Tobit-piecewise regression)

ตัวแบบถดถอยโทบิต-พีซไวส์ถูกพัฒนาขึ้นโดย Mekbunditkul (2010) เพื่อใช้ในการวิเคราะห์การถดถอยกรณีที่มีปัญหาค่านอกเกณฑ์ในการวิเคราะห์การถดถอย (Regression outliers) ซึ่งตัวแบบถดถอยโทบิต-พีซไวส์ เขียนได้ตั้งสมการที่ (5)¹

$$Y_i = \begin{cases} L_0 & ; Y_i^* \leq L_0 \\ Y_i^* & ; L_0 < Y_i^* < U_0 \\ U_0 & ; Y_i^* \geq U_0 \end{cases} \quad (5)$$

$$\text{เมื่อ } Y_i^* = \alpha_1 + \beta_1 x_i + \beta_2 x_i^* + \varepsilon_i \quad L_0 = \begin{cases} L_a & ; x_i \leq \delta \\ L_b & ; x_i > \delta \end{cases} \text{ และ } U_0 = \begin{cases} U_a & ; x_i \leq \delta \\ U_b & ; x_i > \delta \end{cases}$$

$$\text{และ } x_i^* = (x_i - \delta) D_i \text{ โดยที่ } D_i \text{ เป็นตัวแปรหุ่นที่มีค่าเป็นดังนี้ } D_i = \begin{cases} 0 & ; x_i \leq \delta \\ 1 & ; x_i > \delta \end{cases}$$

การประมาณค่าพารามิเตอร์ในตัวแบบถดถอยโทบิต-พีซไวส์นี้ Mekbunditkul (2010) ได้ใช้วิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด (Maximum likelihood estimation) ดังนั้นตัวประมาณค่าประสิทธิภาพการถดถอยจึงมีคุณสมบัติที่ดีตามคุณสมบัติของตัวประมาณภาวะน่าจะเป็นสูงสุด ได้แก่ ความคงเส้นคงวา และมีการแจกแจงปกติเมื่อเข้าใกล้อนันต์ (Best asymptotically normal: B.A.N) ซึ่งผู้พัฒนาตัวแบบได้นำตัวแบบไปประยุกต์ใช้ในกรณีที่มีการวิเคราะห์การถดถอยมีปัญหาค่านอกเกณฑ์หลายๆ กรณี ดังแสดงในหัวข้อที่ 5

การแปลผลค่าพารามิเตอร์ในตัวแบบถดถอยโทบิต-พีซไวส์

เนื่องจากตัวแบบถดถอยโทบิต-พีซไวส์ พัฒนาขึ้นจากหลักการ 2 หลักการดังกล่าวข้างต้น ดังนั้นการแปลผลค่าพารามิเตอร์ในตัวแบบจึงมีความแตกต่างไปจากการแปลผลค่าพารามิเตอร์ในตัวแบบถดถอยการถดถอยเชิงเส้นด้วยวิธีกำลังสองน้อยที่สุด ซึ่งผู้เขียนได้อธิบายการแปลผลค่าพารามิเตอร์ไว้ในหัวข้อนี้ โดยใช้การอธิบายด้วยตัวแบบถดถอยโทบิตที่มีค่าขีดจำกัดล่างเป็น 0 และไม่มีขีดจำกัดบน เพื่อให้ง่ายต่อความเข้าใจ ซึ่งกรณีดังกล่าวถือเป็นกรณีเฉพาะของตัวแบบถดถอยโทบิต-พีซไวส์ ในปี ค.ศ. 1980 McDonald and Moffitt ได้แสดงให้เห็นว่าสัมประสิทธิ์การถดถอยในตัวแบบถดถอยโทบิตนั้นไม่ได้มีความหมายเช่นเดียวกับสัมประสิทธิ์การถดถอยในตัวแบบถดถอย

¹ผู้เขียนแสดงเฉพาะตัวแบบโทบิต-พีซไวส์อย่างง่ายเพื่อให้ง่ายต่อความเข้าใจ

แบบดั้งเดิม เนื่องจากเมื่อค่าของตัวแปรอิสระเปลี่ยนแปลงไปหนึ่งหน่วยนอกจากจะมีอิทธิพลให้ค่าของตัวแปรตามเปลี่ยนไปตามค่าของพารามิเตอร์ที่เกี่ยวข้องกับตัวแปรอิสระนั้นแล้วยังมีผลต่อการเปลี่ยนแปลงความน่าจะเป็นของตัวแปรตามด้วย แสดงรายละเอียดดังต่อไปนี้ จากสมการที่ (1) เมื่อ $L = 0$ จะได้ค่าคาดหวังของ Y เป็นดังนี้

$$E(Y) = x\beta\Phi(z) +, \quad (6)$$

เมื่อค่าปกติมาตรฐานของ Y หรือกล่าวได้ว่า $z = x\beta / \sigma$ และค่าคาดหวังแบบมีเงื่อนไขของ Y^* เป็นดังนี้

$$E(Y^* | Y^* > 0) = x\beta + \frac{\Phi(z)}{\sigma\phi(z)} \quad (7)$$

ดังนั้น จากสมการที่ (6) และ (7) จะได้ว่า

$$E(Y) = E(Y^* | Y^* > 0)\Phi(z)$$

ซึ่งเราสามารถหาอิทธิพลการเปลี่ยนแปลงของตัวแปรอิสระที่ หรือ ที่มีต่อตัวแปรตาม ได้ดังนี้

$$\frac{\partial E(Y)}{\partial X_j} = \Phi(z) \frac{\partial E(Y^* | Y^* > 0)}{\partial X_j} + E(Y^* | Y^* > 0) \frac{\partial \Phi(z)}{\partial X_j} \quad (8)$$

$j = 1, 2, \dots, k$

จากสมการที่ (8) จึงเห็นได้ชัดแจ้งว่าการเปลี่ยนแปลงของตัวแปรตาม Y ประกอบไปด้วย 2 องค์ประกอบหลัก คือ 1) การเปลี่ยนแปลงของตัวแปรตามเมื่อค่าสังเกตมีค่ามากกว่าขีดจำกัดล่าง (Y_2) ที่ถ่วงน้ำหนักด้วยฟังก์ชันการแจกแจงความน่าจะเป็นสะสมของ ดังแสดงในพจน์แรกของสมการที่ (8) และ 2) การเปลี่ยนแปลงความน่าจะเป็นของตัวแปรตามเมื่อค่าสังเกตมีค่ามากกว่าขีดจำกัดล่าง (Y_2) ที่ถ่วงน้ำหนักด้วยค่าคาดหวังแบบมีเงื่อนไขของ Y^* ดังแสดงในพจน์ที่สองของสมการที่ (8)

อัตราส่วน $\frac{\partial E(Y)}{\partial X_j}$ และ $\frac{\partial E(Y^* | Y^* > 0)}{\partial X_j}$ เรียกว่า อิทธิพลการเปลี่ยนแปลงขอบแบบไม่มีเงื่อนไข

(Unconditional marginal effect) และอิทธิพลการเปลี่ยนแปลงขอบแบบมีเงื่อนไข (Conditional marginal effect) ตามลำดับ ค่าของอัตราส่วนสองจำนวนดังกล่าวแสดงวิธีการหาดังต่อไปนี้

ค่าคาดหวังแบบมีเงื่อนไขของตัวแปรตาม เมื่อค่าสังเกตมีค่ามากกว่าขีดจำกัดล่าง เป็นดังนี้

$$E(Y^* | Y^* > 0) = \underline{x}\beta + E(\varepsilon | \varepsilon > -\underline{x}\beta)$$

โดยที่

$$E(\varepsilon | \varepsilon > -\underline{x}\beta) = \frac{\sigma\phi(-\underline{x}\beta/\sigma)}{1-\Phi(-\underline{x}\beta/\sigma)}$$

ดังนั้น

$$E(Y^* | Y^* > 0) = \underline{x}\beta + \frac{\Phi(z)}{\sigma\phi(z)}$$

เมื่อ $z = \underline{x}\beta / \sigma$

ดังนั้น

$$\frac{\partial E(Y^* | Y^* > 0)}{\partial X_j} = \beta_j \left(1 - \frac{z\phi(z)}{\Phi(z)} - \frac{(\phi(z))^2}{(\Phi(z))^2} \right) \quad (9)$$

แทนค่าสมการที่ (9) ลงในสมการที่ (8) จะได้ว่า

$$\frac{\partial E(Y)}{\partial X_j} = \Phi(z)\beta_j \left(1 - \frac{z\phi(z)}{\Phi(z)} - \frac{(\phi(z))^2}{(\Phi(z))^2} \right) + \phi(z) \frac{\beta_j}{\sigma} \left(\underline{x}\beta + \frac{\sigma\phi(z)}{\Phi(z)} \right) \quad (10)$$

และจากสมการที่ (10) จะได้ว่า $\frac{\partial E(Y)}{\partial X_j} = \beta_j \Phi(z)$ ซึ่งเห็นได้ชัดเลยว่าอิทธิพลการเปลี่ยนแปลงของ

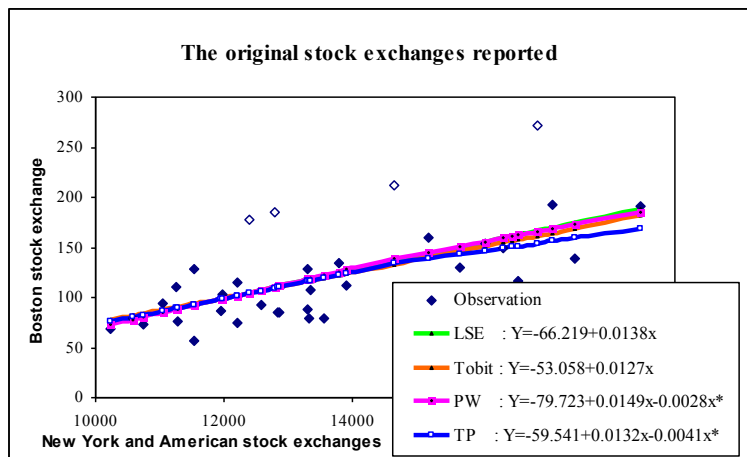
ตัวแปรอิสระ X_j ที่มีต่อการเปลี่ยนแปลงของ Y ขึ้นอยู่กับค่าสัมประสิทธิ์การถดถอยโทบิตคูณกับความน่าจะเป็นสะสมปกติมาตรฐาน ของค่าสังเกตมีค่ามากกว่าขีดจำกัดล่าง (Y_2) ดังนั้นหากผู้วิจัยประยุกต์ใช้ตัวแบบถดถอยโทบิต หรือตัวแบบถดถอยโทบิต-พีชไพล์ในงานวิจัยควรแปลผลค่าสัมประสิทธิ์ถดถอยตามที่คุณเขียนได้กล่าวไว้

การประยุกต์ใช้ตัวแบบถดถอยโทบิต-พีชไพล์

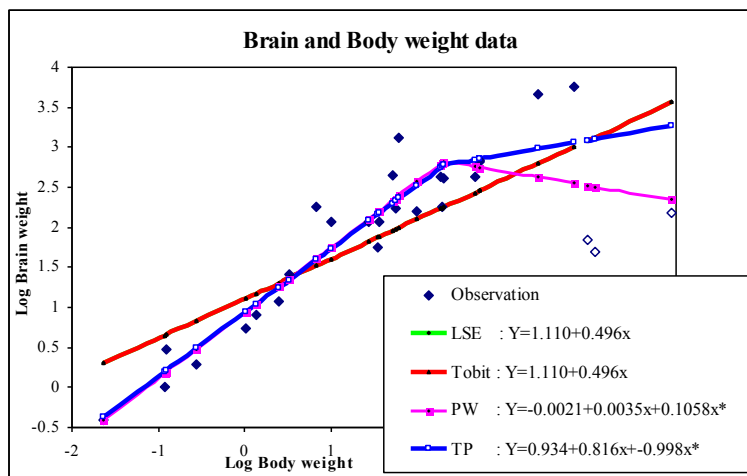
1. การเปรียบเทียบเส้นถดถอย

ในปี ค.ศ. 2010 Mekbunditkut ได้ทำการเปรียบเทียบเส้นถดถอย 4 แบบ คือ เส้นถดถอยกำลังสองน้อยที่สุด เส้นถดถอยโทบิต เส้นถดถอยพีชไพล์ และเส้นถดถอยโทบิต-พีชไพล์ กับข้อมูล 3 ชุด ดังนี้ ข้อมูลชุดที่ 1 คือ ข้อมูลตลาดหลักทรัพย์ (Stock exchanges) ของอเมริกา และบอสตัน ข้อมูลชุดที่ 2 คือ ข้อมูลน้ำหนักสมองและน้ำหนักร่างกาย และข้อมูลชุดที่ 3 คือ ข้อมูลการขนส่งสินค้า ซึ่งผลการวิเคราะห์แสดงดังภาพที่ 2(a) – 2(c)

จากผลการวิเคราะห์ที่แสดงดังภาพที่ 2(a) – 2(c) เห็นได้ว่า ในกรณีที่ข้อมูลมีค่านอกเกณฑ์ในตัวแปรตาม (*y-direction outliers*) แต่เป็นค่านอกเกณฑ์ที่ไม่ได้มีผลกระทบต่อเส้นถดถอยมากนัก ข้อมูลในภาพที่ 2(a) นั้น เส้นถดถอยทั้ง 4 แบบ ไม่ได้ให้ผลลัพธ์ที่แตกต่างกันมาก ในขณะที่หากข้อมูลที่ผู้วิจัยนำมาวิเคราะห์นั้นมีค่านอกเกณฑ์ที่ค่อนข้างมีอิทธิพลต่อเส้นถดถอย จะทำให้เส้นถดถอยกำลังสองน้อยที่สุด เป็นเส้นถดถอยที่มีความเหมาะสมกับข้อมูลน้อยที่สุด ดังแสดงในภาพที่ 2(b) – 2(c) ที่เห็นได้ชัดเลยว่าเส้นถดถอยกำลังสองน้อยที่สุดจะอยู่ห่างไปจากข้อมูลส่วนใหญ่มากที่สุด ในขณะที่เส้นถดถอยโทบิต-พีชโวลส์จะมีความเหมาะสมกับข้อมูลมากที่สุดและให้ค่า R.E.²(Relative efficiency) ต่ำที่สุดเมื่อเทียบกับอีก 3 ตัวแบบ

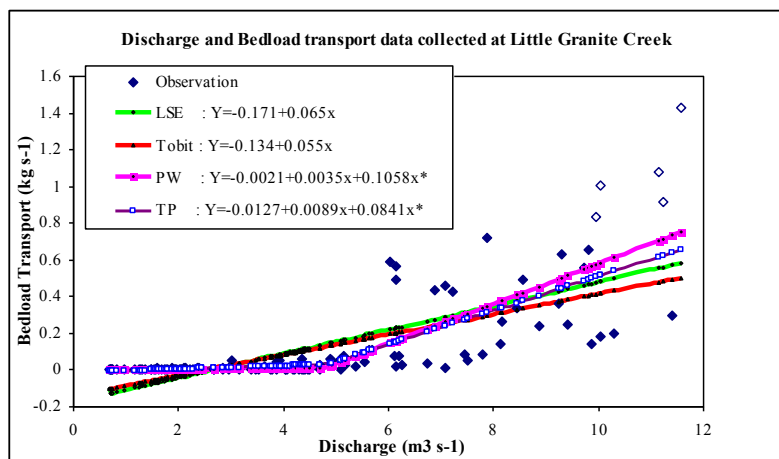


ภาพที่ 2(a) การเปรียบเทียบเส้นถดถอย 4 แบบ สำหรับข้อมูลชุดที่ 1
แหล่งที่มา: Mekbunditkul (2010)



ภาพที่ 2(b) การเปรียบเทียบเส้นถดถอย 4 แบบ สำหรับข้อมูลชุดที่ 2
แหล่งที่มา: Mekbunditkul (2010)

² Relative efficiency คำนวณจากอัตราส่วนของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Mean square error) ของตัวแบบถดถอยโทบิต ตัวแบบถดถอยพีชโวลส์ และตัวแบบถดถอยโทบิต-พีชโวลส์ เทียบกับตัวแบบถดถอยกำลังสองน้อยที่สุด



ภาพที่ 2(c) การเปรียบเทียบเส้นถดถอย 4 แบบ สำหรับข้อมูลชุดที่ 3
แหล่งที่มา: Mekbunditkul (2010)

2. ผลการวิเคราะห์ข้อมูลเชิงเศรษฐศาสตร์

ในหัวข้อนี้ผู้เขียนได้ให้ความสำคัญกับการประยุกต์ใช้ เพื่อให้เกิดประโยชน์ต่อนักวิจัยที่จะนำตัวแบบถดถอยโทบิต (กรณีเฉพาะของตัวแบบถดถอยโทบิต-พีชไวส์) และตัวแบบถดถอยโทบิต-พีชไวส์ไปใช้ในการวิเคราะห์ข้อมูลจริง

1) การวิเคราะห์ข้อมูลเชิงเศรษฐศาสตร์ด้วยตัวแบบถดถอยโทบิต

ในปัจจุบันนี้มีงานวิจัยจำนวนมากที่นำการวิเคราะห์การถดถอยโทบิตไปใช้ในงานวิจัยเนื่องจากข้อมูลในปัจจุบันที่มีขีดจำกัดล่าง และ/หรือขีดจำกัดบน โดยเฉพาะอย่างยิ่งข้อมูลที่เกี่ยวข้องกับเศรษฐศาสตร์การศึกษา เช่น ค่าใช้จ่ายทางการศึกษา รายได้ค่าธรรมเนียมการศึกษา งบประมาณ เป็นต้น สำหรับบทความนี้จะกล่าวถึงงานวิจัยที่ได้ประยุกต์ใช้ตัวแบบถดถอยโทบิตในงานวิจัยทางด้านเศรษฐศาสตร์การศึกษา 2 เรื่อง ซึ่งทั้งสองเรื่องที่เกี่ยวข้องข้อมูลตัวแปรตามเป็นข้อมูลเกี่ยวกับค่าใช้จ่ายซึ่งแน่นอนว่ามีขีดจำกัดล่างเป็น 0 (หรืออาจมากกว่านั้น) ผู้เขียนได้สรุปผลการศึกษาเพื่อให้เห็นการประยุกต์ใช้และการแปลผลการวิเคราะห์ ดังต่อไปนี้

ในปี ค.ศ. 2006 Tansel and Bircan ได้ศึกษาเรื่อง “ความต้องการทางการศึกษาในประเทศตุรกี : การวิเคราะห์การถดถอยโทบิตในข้อมูลค่าใช้จ่ายด้านการกวดวิชา” ซึ่งผู้วิจัยเห็นว่าปัจจุบันการศึกษาในสถาบันกวดวิชาเป็นที่นิยมอย่างมากโดยเฉพาะในภูมิภาคเอเชียตะวันออก เป็นต้นว่าในประเทศอียิปต์และไต้หวัน ที่มีการตั้งสถาบันกวดวิชาเพิ่มขึ้นอย่างแพร่หลาย (Bray and Kwok, 2003; Bray, 1999) เนื่องจากนักเรียนส่วนใหญ่ที่เรียนในสถาบันกวดวิชาเชื่อว่าการเข้าเรียนกวดวิชาจะช่วยให้มีโอกาสในการเข้าศึกษาต่อในระดับอุดมศึกษาได้ โดยเฉพาะในประเทศที่มีการแข่งขันสอบเข้าเรียนในระดับมหาวิทยาลัยที่มีชื่อเสียงที่สูงมาก ดังเช่นในประเทศเกาหลีใต้ กรีซ ญี่ปุ่น ตุรกี เป็นต้น สำหรับประเทศที่กำลังพัฒนาปัจจัยที่ทำให้มีการแข่งขันเข้ามหาวิทยาลัยของรัฐสูงนั้นเหตุผลที่สำคัญคือ จำนวนมหาวิทยาลัยของรัฐน้อยไม่เพียงพอต่อความต้องการ เนื่องจากมหาวิทยาลัยเหล่านี้มีค่าเล่าเรียนที่ไม่สูงมากนัก ทั้งนี้ไม่รวมถึง

หลักสูตรภาคพิเศษต่างๆ ซึ่งปัจจัยเหล่านี้ล้วนมีส่วนให้นักเรียนต้องการเรียนในสถาบันกวดวิชาเพื่อเพิ่มความสามารถในการแข่งขันของตนเอง ด้วยเหตุนี้ทำให้ Tansel and Birca ต้องการศึกษาค่าใช้จ่ายด้านการกวดวิชาของครัวเรือน

งานวิจัยนี้มีการศึกษาโดยใช้ข้อมูลค่าใช้จ่ายด้านการกวดวิชาของครัวเรือนจากข้อมูลการสำรวจค่าใช้จ่ายครัวเรือน ปีค.ศ.1994 ซึ่งสำรวจโดยสำนักงานสถิติแห่งรัฐตุรกี (State Institute of Statistics of Turkey) จากข้อมูลดังกล่าวรายงานว่ามี 4,279 ครัวเรือน ที่มีค่าใช้จ่ายค่าธรรมเนียมการศึกษา ทั้งที่ศึกษาในระบบโรงเรียนสังกัดเอกชน (โรงเรียนที่สังกัดรัฐบาลไม่มีค่าธรรมเนียมทางการศึกษา) และค่ากวดวิชา ซึ่งในบรรดาครัวเรือนดังกล่าวมี 646 ครัวเรือน ที่มีค่าใช้จ่ายด้านการกวดวิชา และ 3,252 ครัวเรือน ที่ไม่มีค่าใช้จ่ายด้านกวดวิชา มีเพียงแต่ค่าธรรมเนียมการศึกษาเท่านั้น และครัวเรือนส่วนที่เหลือไม่มีการรายงาน ผู้วิจัยได้ทำการศึกษาเส้นเองเงา (Engel curve) โดยใช้ตัวแบบโทบิตแสดงความสัมพันธ์ของตัวแปร และแน่นอนว่าข้อมูลค่าใช้จ่ายด้านการกวดวิชา มีค่าขีดจำกัดล่างเป็น 0

สำหรับการศึกษานี้มีตัวแปรตามที่ศึกษาคือค่าใช้จ่ายด้านการกวดวิชาของครัวเรือน และมีตัวแปรอิสระที่ศึกษาทั้งสิ้น 10 ตัวแปร ได้แก่ อายุของหัวหน้าครัวเรือน อายุของหัวหน้าครัวเรือนยกกำลังสอง จำนวนปีที่เรียนของหัวหน้าครัวเรือน จำนวนปีที่เรียนของมารดา และตัวแปรอิสระที่เป็นตัวแปรหุ่น (Dummy variable) มีดังนี้ สถานภาพการทำงานและสถานภาพสมรสของมารดา เขตที่พักอาศัย สภาพถนน (เฉพาะที่อาศัยอยู่นอกเขตเทศบาล) สถานภาพการถือครองที่อยู่อาศัย และจำนวนเด็กในครัวเรือน จากผลการวิเคราะห์ข้อมูลด้วยการถดถอยโทบิต โดยวิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด พบว่า ตัวแปรอิสระที่มีอิทธิพลเชิงเส้นต่อค่ากวดวิชาของครัวเรือน อย่างมีนัยสำคัญ ได้แก่ ค่าใช้จ่ายครัวเรือน อายุของหัวหน้าครัวเรือน อายุของหัวหน้าครัวเรือนยกกำลังสอง จำนวนปีที่เรียนของหัวหน้าครัวเรือนและของมารดา สถานภาพการสมรสของมารดา เขตที่ตั้ง และจำนวนเด็กในครัวเรือน ซึ่งผลที่ได้ดังกล่าวมีความสมเหตุสมผลและมีหลักฐานยืนยันเพียงพอจากการศึกษาวิจัยของ Tansel and Bercan (2006)

จากการศึกษาพบว่าผู้วิจัยได้นำเสนอค่าสัมประสิทธิ์การถดถอย ค่าอิทธิพลการเปลี่ยนแปลงขอบแบบไม่มีเงื่อนไข และอิทธิพลการเปลี่ยนแปลงขอบแบบมีเงื่อนไข สำหรับตัวแปรอิสระที่มีอิทธิพลเชิงเส้นต่อค่าใช้จ่ายด้านการกวดวิชาอย่างมีนัยสำคัญยิ่ง ($p < 0.01$) ที่มีค่าสัมประสิทธิ์การถดถอยสูงที่สุดคือตัวแปรค่าใช้จ่ายครัวเรือน ($\beta = 6.33$) ค่าอิทธิพลการเปลี่ยนแปลงขอบแบบไม่มีเงื่อนไขเป็น 0.91 และมีค่าอิทธิพลการเปลี่ยนแปลงขอบแบบมีเงื่อนไขเป็น 1.21 หรือกล่าวได้ว่า $\frac{\partial E(Y)}{\partial X_i}$ ซึ่งอธิบายได้ว่าเมื่อนั้นคือเมื่อตัวแปรอื่น ๆ คงที่ ค่าลอกการิทึมของค่าใช้จ่ายครัวเรือนเพิ่มขึ้น 1 หน่วย จะทำให้ค่าลอกการิทึมของค่าใช้จ่ายด้านการกวดวิชาเพิ่มขึ้น 6.33 หน่วย และหากค่าใช้จ่ายครัวเรือนเพิ่มขึ้นร้อยละ 1 จะทำให้ค่าใช้จ่ายด้านการกวดวิชาเพิ่มขึ้นร้อยละ 0.91 หรือประมาณร้อยละ 1 ด้วย สำหรับตัวแปรอื่นๆ สามารถอธิบายได้ในลักษณะเดียวกัน ซึ่งสามารถศึกษาได้ในงานของ Tansel and Bercan

เมื่อพิจารณาผลการวิจัยของ Psacharopolus and Papakonstantinou (2005) ได้ทำการศึกษาในลักษณะคล้ายคลึงกันคือศึกษาความสัมพันธ์ระหว่างค่าใช้จ่ายด้านการกวดวิชากับ

รายได้ครัวเรือน แต่พบว่ารายได้ครัวเรือนไม่มีอิทธิพลเชิงเส้นต่อค่าใช้จ่ายด้านการกวดวิชาของครัวเรือน ซึ่งผลที่ได้ดังกล่าวไม่สอดคล้องกับความจริงอาจเนื่องมาจากการวิจัยใช้การวิเคราะห์การถดถอยแบบกำลังสองน้อยที่สุดนั่นเอง

นอกจากนี้ยังมีงานวิจัยของ Stair, Rephan and Heberling (2006) ได้ทำการศึกษาเรื่อง “ความต้องการการศึกษาในโรงเรียนสังกัดภาครัฐ : เฉพาะในเขตเทศบาล” ซึ่งได้มีการประยุกต์ใช้ตัวแบบถดถอยโทบิต และใช้วิธีการประมาณค่าแบบภาวะน่าจะเป็นสูงสุด โดยทำการศึกษาอิทธิพลเชิงเส้นของตัวแปรอิสระที่มีต่อตัวแปรตาม ซึ่งตัวแปรตามในการศึกษาค้างครั้งนี้คือ ค่าธรรมเนียมการศึกษาที่ยินดีจ่ายเพื่อให้สถาบันการศึกษาภาครัฐมีคุณภาพที่ดีขึ้น ซึ่งทำการศึกษาใน 2 พื้นที่ในเขตเทศบาลของรัฐเพนซิลวาเนีย สำหรับตัวแปรอิสระมีดังนี้ 1) รายได้รวมครัวเรือน 2) รายได้รวมครัวเรือนกำลังสอง 3) อาชีพหลัก 4) การศึกษาของหัวหน้าครัวเรือน 5)-7) การมีบุตรหลานในครัวเรือนที่อยู่ในระบบการศึกษา 8) ขนาดครัวเรือน 9) อายุของหัวหน้าครัวเรือน 10) การมีบุตรหลานในครัวเรือนที่ไม่ได้เรียนในโรงเรียนภาครัฐในเขตพื้นที่ที่อาศัยอยู่ 11) การมีบุตรหลานในครัวเรือนที่เรียนในโรงเรียนประจำอำเภอ 12) ความสัมพันธ์ระหว่างครัวเรือนกับผู้ที่ทำงานในสังกัดโรงเรียนของภาครัฐ 13) ความคาดหวังของผู้ปกครองที่มีต่อการบรรลุเป้าหมายของลูกหลานหลังจากที่โรงเรียนที่สังกัดมีการปรับปรุงแล้ว และ 14) รายได้หลักของครัวเรือน

ผลการศึกษาในงานวิจัยเมื่อใช้การวิเคราะห์การถดถอยโทบิต พบว่า ในย่าน Bedford ตัวแปรอิสระที่มีอิทธิพลเชิงเส้นต่อตัวแปรตามอย่างมีนัยสำคัญ มีทั้งสิ้น 5 ตัวแปร ได้แก่ รายได้รวมครัวเรือน รายได้รวมครัวเรือนกำลังสอง ความคาดหวังของผู้ปกครองที่มีต่อการบรรลุเป้าหมายของลูกหลานหลังจากที่โรงเรียนที่สังกัดมีการปรับปรุงแล้ว ความสัมพันธ์ระหว่างครัวเรือนกับผู้ที่ทำงานในสังกัดโรงเรียนของภาครัฐ และการมีบุตรหลานในครัวเรือนที่ไม่ได้เรียนในโรงเรียนภาครัฐในเขตพื้นที่ที่อาศัยอยู่

ในขณะที่ย่าน Hyndman ตัวแปรอิสระที่มีอิทธิพลเชิงเส้นต่อตัวแปรตามอย่างมีนัยสำคัญ มีทั้งสิ้น 3 ตัวแปร ได้แก่ รายได้รวมครัวเรือน การมีบุตรหลานในครัวเรือนที่อยู่ในระบบการศึกษา และการมีบุตรหลานในครัวเรือนที่เรียนในโรงเรียนประจำอำเภอ

จากงานวิจัยดังกล่าวทำให้ทราบว่าในระยะหลังการวิเคราะห์การถดถอยโทบิต ถูกนำมาใช้ในการวิเคราะห์ความสัมพันธ์เชิงเส้น ในหลากหลายสาขารวมไปถึงข้อมูลทางด้านเศรษฐศาสตร์ การศึกษาด้วย อาทิ ค่าใช้จ่ายในการกวดวิชา ค่าธรรมเนียมการศึกษาที่ยินดีจ่ายเพื่อปรับปรุงคุณภาพของสถาบันการศึกษาภาครัฐ งบประมาณ จำนวนคน รายได้จากค่าธรรมเนียมการศึกษา เงินเดือนบุคลากร เป็นต้น ซึ่งข้อมูลเหล่านี้เป็นข้อมูลที่มีขีดจำกัดล่าง อาทิ เงินเดือนบุคลากรจะมีขีดจำกัดล่างเป็นรายได้ขั้นต่ำต่อวัน ค่าใช้จ่าย งบประมาณ หรือจำนวนคน จะมีขีดจำกัดล่างอยู่ที่ 0 ดังเช่นงานวิจัย 2 เรื่องที่ผู้วิจัยได้กล่าวถึงข้างต้น และเมื่อข้อมูลมีลักษณะเป็นเช่นนี้และผู้วิจัยยังคงใช้การวิเคราะห์การถดถอยด้วยวิธีกำลังสองน้อยที่สุด ย่อมทำให้เส้นถดถอยที่สร้างได้มีลักษณะที่ค่าประมาณความชันต่ำเกินไป (Underestimate) อีกทั้งยังเป็นค่าประมาณที่เอนเอียงเมื่อเข้าใกล้อนันต์อีกด้วย

นอกจากนี้ยังมีงานวิจัยเชิงทฤษฎีที่กล่าวถึงการประยุกต์ใช้ตัวแบบโทบิตกับกรณีที่เกิดปัญหาค่านอกเกณฑ์ในการถดถอย (Regression outliers) ดังแสดงในงานวิจัยของ Mekbunditkul (2010) ซึ่งผลการวิจัยพบว่าตัวแบบถดถอยโทบิตสามารถลดอิทธิพลของค่าออกเกณฑ์ในตัวแปรตาม (*y-direction outliers*) บางรูปแบบได้

2) การวิเคราะห์ข้อมูลเชิงเศรษฐศาสตร์ด้วยตัวแบบถดถอยโทบิต-พีชไวลส์

มีผู้วิจัยได้นำตัวแบบถดถอยโทบิต-พีชไวลส์ ใช้ในการวิเคราะห์การถดถอยที่มีปัญหาค่านอกเกณฑ์ และพบว่าตัวแบบถดถอยโทบิต-พีชไวลส์ให้ผลลัพธ์แตกต่างไปจากตัวแบบการถดถอยกำลังสองน้อยสุด ตัวอย่างเช่น งานวิจัยของ Thipbharos (2014) ที่ได้รับปรุงตัวแบบถดถอยโทบิต-พีชไวลส์ โดยใช้หลักการของการแบ่งกลุ่มข้อมูลด้วยการวิเคราะห์จำแนกกลุ่ม (Cluster analysis) และนำตัวแบบถดถอยที่ได้ไปประยุกต์ใช้กับการวิเคราะห์หาปัจจัยที่มีอิทธิพลต่อหนี้สินสำหรับการศึกษาของบุตรหลานในครัวเรือน โดยปัจจัยที่นำมาศึกษา ได้แก่ ค่าใช้จ่ายครัวเรือน ค่าใช้จ่ายทางการศึกษา ระดับการศึกษาของหัวหน้าครัวเรือน รายได้ครัวเรือน อายุของหัวหน้าครัวเรือน มูลค่าทรัพย์สินของครัวเรือน เงินทุนการศึกษาที่บุตรหลานได้รับ เขตการปกครอง ขนาดครัวเรือน หนี้สินครัวเรือน และจำนวนบุตรหลานที่กำลังศึกษาอยู่ สำหรับข้อมูลที่ผู้วิจัยนำมาใช้นั้นเป็นข้อมูลสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ปีพ.ศ. 2554 ที่สำรวจโดยสำนักงานสถิติแห่งชาติ ซึ่งผู้วิจัยเลือกเฉพาะครัวเรือนที่มีบุตรหลานที่กำลังเรียนอยู่ตั้งแต่ระดับอนุบาลจนกระทั่งปริญญาเอก มีจำนวนทั้งสิ้น 10,261 ครัวเรือน

ผลการวิเคราะห์ด้วยตัวแบบถดถอยโทบิต-พีชไวลส์พบว่า ได้ตัวแบบถดถอยสำหรับทำนายหนี้สินครัวเรือนสำหรับการศึกษาทั้งสิ้น 3 ตัวแบบ ซึ่งเป็นตัวแบบสำหรับครัวเรือนที่มีฐานะยากจนปานกลาง และดี จึงทำให้ผลการวิเคราะห์แสดงให้เห็นว่าปัจจัยที่มีอิทธิพลต่อหนี้สินครัวเรือนสำหรับการศึกษานั้นแตกต่างกัน สำหรับครัวเรือนที่มีฐานะยากจน พบว่าปัจจัยที่มีอิทธิพลอย่างมีนัยสำคัญเมื่อวิเคราะห์ด้วยตัวแบบถดถอยโทบิต-พีชไวลส์ ได้แก่ ค่าใช้จ่ายทางการศึกษา ขนาดครัวเรือน และหนี้สินครัวเรือน ในขณะที่ครัวเรือนที่มีฐานะปานกลาง มีปัจจัยที่มีนัยสำคัญ ได้แก่ รายได้ครัวเรือน หนี้สินครัวเรือน และจำนวนบุตรหลานที่กำลังศึกษาอยู่ ส่วนครัวเรือนที่มีฐานะดีนั้น พบว่าปัจจัยที่มีอิทธิพลต่อหนี้สินสำหรับการศึกษาอย่างมีนัยสำคัญ ได้แก่ อายุของหัวหน้าครัวเรือน ขนาดครัวเรือน และหนี้สินครัวเรือน และนอกจากนั้นผลการวิเคราะห์แสดงให้เห็นว่าหากผู้วิจัยใช้ตัวแบบการถดถอยกำลังสองน้อยสุดจะทำให้ได้ปัจจัยที่มีนัยสำคัญแตกต่างไปจากที่กล่าวไว้ ซึ่งผลการวิเคราะห์ทำให้ทราบว่าเมื่อสถานะทางเศรษฐกิจของครัวเรือนแตกต่างกันย่อมทำให้ปัจจัยที่มีผลต่อจำนวนหนี้สินการศึกษาของครัวเรือนแตกต่างกัน

จากนั้น Thipbharos (2015) ได้พัฒนาตัวแบบถดถอยที่ชื่อว่าตัวแบบถดถอยปรับปรุง (Adjusted regression) ซึ่งเป็นตัวแบบที่พัฒนาจากตัวแบบถดถอยโทบิต-พีชไวลส์ เพื่อใช้ในการจัดการกับปัญหาค่านอกเกณฑ์ที่เกิดขึ้นในการวิเคราะห์การถดถอย ผู้วิจัยกำหนดให้ค่านอกเกณฑ์มีค่าอยู่ที่ขีดจำกัดบน และ/หรือล่างของช่วงความเชื่อมั่นของค่าทำนายแต่ละค่า (Confidence band of the individual prediction) ผู้วิจัยได้ประยุกต์ใช้ตัวแบบถดถอยที่พัฒนาขึ้นในการหาอิทธิพลของปัจจัยต่างๆ ที่มีต่อรายจ่ายด้านการประกันภัยของเกษตรกรไทย โดยงานวิจัยของ Thipbharos (2015) นี้ได้มีการเปรียบเทียบความเหมาะสมของตัวแบบด้วยค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Average sum square residual (ASSR)) ที่พัฒนาขึ้นโดย Mekbunditkul (2010) และพบว่าตัวแบบถดถอยปรับปรุงให้ค่า ASSR ที่ต่ำกว่าผลที่ได้จากตัวแบบถดถอยพีชไวลส์ ข้อมูลที่ใช้ในการวิเคราะห์ได้จากการสำรวจภาวะเศรษฐกิจและสังคมของครัวเรือน ปีพ.ศ. 2556 ที่สำรวจโดยสำนักงานสถิติแห่งชาติ จำนวนครัวเรือนเกษตรกรทั้งสิ้น 5,324 ครัวเรือน³

ปัจจัยที่ใช้ในการศึกษาอิทธิพลที่มีต่อค่าใช้จ่ายด้านการประกันภัยของเกษตรกรไทย ได้แก่ พื้นที่เพาะปลูก ระดับการศึกษาของหัวหน้าครัวเรือน อายุของหัวหน้าครัวเรือน ขนาดครัวเรือน และรายได้ครัวเรือน ผลการวิจัยนี้พบว่า ผลการวิเคราะห์ด้วยตัวแบบถดถอยปรับปรุง ทำให้ได้ตัวแบบถดถอยสำหรับการทำนายค่าใช้จ่ายด้านการประกันภัยของเกษตรกร 2 ตัวแบบ สำหรับครัวเรือนยากจน (Poor household) และครัวเรือนไม่ยากจน (Non-poor household) ซึ่งชัดเจนว่าปัจจัยที่มีอิทธิพลต่อตัวแปรตามนั้นแตกต่างกัน โดยที่ครัวเรือนยากจน ปัจจัยที่มีอิทธิพล ได้แก่ ระดับการศึกษาของหัวหน้าครัวเรือน และรายได้ครัวเรือน โดยปัจจัยที่มีขนาดอิทธิพลมากกว่าคือ ระดับการศึกษาของหัวหน้าครัวเรือน ในขณะที่ครัวเรือนที่ไม่ยากจน มีปัจจัยที่มีนัยสำคัญ คือ อายุ และรายได้ของหัวหน้าครัวเรือน และพบว่าปัจจัยที่มีขนาดอิทธิพลมากกว่าคือ อายุของหัวหน้าครัวเรือน นอกจากนี้ผลการวิจัยยังพบว่า ตัวแบบถดถอยปรับปรุงนี้มีค่า ASSR ที่ต่ำกว่าตัวแบบถดถอยพีชโวลส์

แนวทางการนำตัวแบบถดถอยโทบิต-พีชโวลส์ไปประยุกต์ใช้

การวิเคราะห์การถดถอยโทบิตถูกนำมาใช้ในการวิเคราะห์อิทธิพลเชิงเส้นระหว่างตัวแปรอย่างแพร่หลาย ในกรณีที่ตัวแปรตามมีขีดจำกัดบน และ/หรือขีดจำกัดล่าง ซึ่งตัวแบบถดถอยโทบิตนี้ถูกนำมาใช้ในหลายสาขาวิชา โดยเฉพาะอย่างยิ่งในกรณีที่ข้อมูลเกี่ยวข้องกับด้านเศรษฐศาสตร์ เช่น รายได้ ค่าใช้จ่าย งบประมาณ เป็นต้น

สำหรับตัวแบบถดถอยโทบิต-พีชโวลส์ถูกพัฒนาขึ้นเพื่อจัดการกับปัญหาข้อมูลที่มีค่านอกเกณฑ์ และใช้วิธีการลดอิทธิพลของค่านอกเกณฑ์โดยการถ่วงน้ำหนักให้เท่ากับค่าที่อยู่บริเวณขีดจำกัดบนหรือขีดจำกัดล่าง จะทำให้ตัวแบบถดถอยที่มีความเหมาะสมมากกว่าตัวแบบถดถอยกำลังสองน้อยที่สุด ซึ่งการจัดการกับค่านอกเกณฑ์ในลักษณะนี้ถือเป็นวิธีแกร่ง (Robust method) วิธีหนึ่งซึ่งมีความยุ่งยากซับซ้อนน้อยกว่าการถดถอยแบบแกร่ง (Robust regression) วิธีอื่น

ในกรณีที่ข้อมูลมีขีดจำกัดบน และ/หรือขีดจำกัดล่าง ได้มีการพิสูจน์ทางทฤษฎีให้เห็นอย่างชัดเจนจากงานวิจัยของ Greene (1981) ที่แสดงว่าการวิเคราะห์การถดถอยกำลังสองน้อยที่สุด ทำให้ได้สมการถดถอยที่สร้างได้มีคุณสมบัติเอนเอียงในลักษณะค่าประมาณของความชันต่ำเกินไป (Underestimate) อีกทั้งยังเป็นค่าประมาณที่เอนเอียงเมื่อเข้าใกล้อนันต์ ดังนั้นผู้วิจัยควรพึงระวังการวิเคราะห์ข้อมูลที่มีลักษณะดังกล่าว นอกจากนั้นการแปลผลค่าสัมประสิทธิ์การถดถอยในตัวแบบถดถอยโทบิต และตัวแบบถดถอยโทบิต-พีชโวลส์ ยังสามารถแปลความหมายได้สองประเด็น ได้แก่ การเปลี่ยนแปลงของตัวแปรตามเมื่อตัวแปรอิสระเปลี่ยนแปลงไปหนึ่งหน่วย และการเปลี่ยนแปลงที่ความน่าจะเป็นของตัวแปรตาม ซึ่งการแปลผลดังกล่าวนี้จะมีประโยชน์ต่อผู้วิจัยมากยิ่งขึ้น อย่างไรก็ตามก็ตามตัวแบบถดถอยโทบิต และตัวแบบถดถอยโทบิต-พีชโวลส์ มีข้อสมมติเช่นเดียวกับตัวแบบถดถอยด้วยวิธีกำลังสองน้อยที่สุด ดังนั้นผู้วิจัยจะต้องตรวจสอบหรือทดสอบข้อสมมติด้วยเช่นเดียวกัน

จากผลการวิจัยของ Mekbunditkul (2010) และ Thipbharos (2014, 2015) นั้นทำให้ทราบว่า ตัวแบบถดถอยโทบิต-พีชโวลส์ และตัวแบบถดถอยปรับปรุง สามารถลดอิทธิพลของค่านอกเกณฑ์ในตัวแปรตาม (*y-direction outliers*) ได้ดี และลดอิทธิพลของค่านอกเกณฑ์ใน

³ครัวเรือนที่ประกอบอาชีพการเกษตรเป็นอาชีพหลัก

ตัวแปรอิสระ (*x-direction outliers*) และทั้งในตัวแปรอิสระและตัวแปรตาม ได้ในบางกรณี อีกประเด็นหนึ่งที่น่าสนใจที่เป็นผลที่ได้จากการสร้างตัวแบบโดยแนวคิดของตัวแบบถดถอยพีชโวลส์ คือ ตัวแบบที่ได้จะเหมาะสมกับกรณีที่ข้อมูลที่น่ามาวิเคราะห์ไม่ได้มีความกลมกลืนกันทั้งเซตของข้อมูล ซึ่งอาจเกิดจากการที่ประชากรมีการแจกแจงที่แตกต่างกันหรือมีความเปลี่ยนแปลงทางโครงสร้างของข้อมูล ดังนั้นการวิเคราะห์การถดถอยสำหรับข้อมูล 1 ชุด ด้วยเส้นถดถอยเพียง 1 เส้น (Single regression line) นั้นอาจไม่เหมาะสมกับข้อมูลในปัจจุบันอีกต่อไป

เอกสารอ้างอิง

- Bray, M. (1999). The shadow education system: Private tutoring and its implications for planners. *Fundamentals in Educational Planning*. Paris: UNESCO International Institute for Educational Planning (IIEP).
- Bray, M. & Kwok, P. (2003). Demand for private supplementary tutoring: Conceptual considerations and socio-economic patterns in Hong Kong. *Economics of Education Review*, 22, 611-620.
- Greene, W.H. (1981). On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model. *The Econometric Society*, 42(March), 505-513.
- McDonald, J.F. & Moffitt, R.A. (1980). The Uses of Tobit Analysis. *The Review of Economics and Statistics*, 62(May), 318-321.
- Mekbunditkul, T. (2010). *An Alternative Estimation for Regression Coefficients with Outliers* (Dissertation). Bangkok: National Institute of Development Administration.
- Psacharopoulos, G. & Papakonstantinou, G. (2005). The real university cost in a “free” higher education country. *Economics of Education Review*, 24(1), 103-108.
- Quandt, R.E. (1958). The estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes. *Journal of the American Statistical Association*, 53, 873-880.
- Rousseeuw, P.J. & Leroy, A. M. (1987). *Robust Regression and Outlier detection*. New York : John Wiley & Sons.
- Stair, A., Rephann, T. & Heberling, M. (2006). Demand for public education: Evidence from a rural school district. *Economics of Education Review*, 25, 521-531.
- Tansel, A. & Bircan, F. (2006). Demand for education in Turkey: A tobit analysis of private tutoring expenditures. *Economics of Education Review*, 25, 303-313.
- Thipbharos, T. (2014). What can Tobit-piecewise regression tell us about the determinants of household educational debt?. *Journal of International Education Studies*, 7(4), 1-13.
- Thipbharos, T. (2015). The Derivation of Maximum Likelihood Estimation for Linear Relation with Outliers. *International Journal of Applied Mathematics and Statistics*, 53(3).
- Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, 26, 24-36.