บทที่ 1

บทนำ

1.1 ที่มาและความสำคัญของปัญหา

การวิเคราะห์การถดถอย (Regression Analysis) เป็นวิธีการทางสถิติวิธีหนึ่งที่มีผู้นิยมใช้อย่างมากในงานหลายสาขา เช่น เศรษฐศาสตร์ บริหารธุรกิจ วิศวกรรมศาสตร์ เป็นต้น เพื่ออธิบายความสัมพันธ์ระหว่างตัวแปรต่างๆ เริ่มตั้งแต่ตัวแปรสองตัวว่ามีความสัมพันธ์กันหรือไม่ และถ้ามีความสัมพันธ์กันจะมีลักษณะความสัมพันธ์เป็นแบบใด นอกจากนี้การวิเคราะห์การถดถอยสามารถนำไปใช้ในการประมาณและการพยากรณ์เกี่ยวกับตัวแปร โดยกำหนดให้ตัวแปรที่ต้องการจะศึกษาเป็นตัวแปรตาม (Dependent Variable : Y) และอาศัยความรู้เกี่ยวกับค่าตัวแปรอื่นๆ ที่แทนอิทธิพลหรือปัจจัยที่มีผลต่อตัวแปรตามที่ต้องการจะศึกษาเป็นตัวแปรอิสระ (Independent Variable : X) เพื่อศึกษาว่าตัวแปรอิสระมีอิทธิพลอย่างไรต่อตัวแปรตาม หรือตัวแปรอิสระมีส่วนในการอธิบายความผันแปรของตัวแปรตามอย่างไร ซึ่งในการวิเคราะห์การถดถอยนั้นจะต้องสร้างตัวแบบการถดถอยแสดงความสัมพันธ์ของตัวแปรขึ้นมาก่อน จากนั้นจะทำการวิเคราะห์การถดถอยตามตัวแบบและข้อสมมติของตัวแบบที่กำหนดไว้ล่วงหน้า ข้อสมมติหนึ่งของตัวแบบซึ่งเกี่ยวข้องกับความคลาดเคลื่อนมีความสำคัญยิ่งต่อการทดสอบสมมติฐานและการประมาณค่าพารามิเตอร์ในตัวแบบ หากข้อสมมติฐานข้อใดข้อหนึ่งไม่เป็นจริง จะส่งผลให้ตัวประมาณที่ได้ไม่มีคุณสมบัติของตัวประมาณที่ดี (Best Linear Unbiased Estimator : BLUE) และการทดสอบสมมติฐานเกี่ยวกับข้อมูลต่างๆ ที่นำมาใช้วิเคราะห์ผิดพลาด ในการวิจัยครั้งนี้จะกล่าวถึงข้อตกลงที่กำหนดว่าค่าคลาดเคลื่อนมีการแจกแจงที่เป็นอิสระกัน (ไม่มีความสัมพันธ์กัน) ซึ่งในทางปฏิบัติมักพบเสมอว่าค่าคลาดเคลื่อนมีการแจกแจงที่ไม่เป็นอิสระกัน (ทรงศิริ แต้สมบัติ, 2541) การที่ค่าคลาดเคลื่อนไม่เป็นอิสระกัน หรือมีความสัมพันธ์ซึ่งกันและกันในแต่ละช่วงเวลาเราเรียกเหตุการณ์เช่นนี้ว่า **การมีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อน** (Autocorrelation)

การมีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อนมักจะเกิดกับกรณีที่ค่าของตัวแปรตามและตัวแปรอิสระมีการเก็บต่อเนื่องกันตามเวลา ที่เรียกว่า อนุกรมเวลา (Box และคณะ, 1994) ข้อมูลอนุกรมเวลา (Time Series Data) ส่วนใหญ่ที่นำมาใช้ในการวิเคราะห์เป็นข้อมูลทางด้านเศรษฐศาสตร์และด้านธุรกิจ เช่น จากข้อมูลที่บันทึกพบว่าฟังก์ชันการผลิตขึ้นอยู่กับแรงงานและทุน ถ้าไตรมาสแรกของการผลิตแรงงานเกิดนัดหยุดงาน ผลผลิตที่ลดลงในไตรมาสแรกอัน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

เนื่องมาจากการนัดหยุดงาน (ซึ่งสะท้อนอยู่ในค่าคลาดเคลื่อนที่เกิดในไตรมาสนี้) ส่งผลต่อเนื่องไปยังผลผลิตต่ำกว่าที่ควรจะเป็นในไตรมาสที่ 2 หรือในกรณีที่ผู้วิจัยทำการวิเคราะห์ปัจจัยที่มีต่อการเช่าห้องพักของนักท่องเที่ยว ลักษณะความสัมพันธ์ของจำนวนห้องพักที่นักท่องเที่ยวจองกับปัจจัยที่มีผลต่อการจองห้องพักไม่ได้เป็นเชิงเส้น แต่ผู้วิจัยกำหนดตัวแบบการถดถอยให้มีความสัมพันธ์กันเชิงเส้นตรง แล้วค่าคลาดเคลื่อนจะเกิดขึ้นจากการกำหนดตัวแบบที่ไม่ถูกต้อง หรือกรณีที่ขาดตัวแปรอิสระที่สำคัญบางตัวในตัวแบบการถดถอย ผลกระทบของตัวแปรที่ขาดไปนี้จะพบในเทอมของค่าคลาดเคลื่อน ลักษณะที่กล่าวมาข้างต้นล้วนส่งผลให้มีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อนเกิดขึ้นทั้งสิ้น

ผลของการมีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อนจะทำให้ตัวประมาณของพารามิเตอร์ β_i ไม่เป็นตัวประมาณที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เนื่องจากมีค่าความแปรปรวนไม่ต่ำที่สุด ถึงแม้จะยังคงเป็นตัวประมาณที่ไม่เอนเอียงก็ตาม ซึ่งจะมีผลให้การอนุมานเกิดความผิดพลาด

ตัวสถิติทดสอบที่ใช้ในการตรวจสอบอัตตสหสัมพันธ์สำหรับข้อมูลทั่วไปมีหลายวิธี เช่น การตรวจสอบด้วยการสร้างกราฟค่าความคลาดเคลื่อน (Graphical Method), ตัวสถิติทดสอบเดอร์บินและวัตสัน (Durbin-Watson test statistic) ตัวสถิติทดสอบเดอร์บิน H (Durbin's H test statistic) การทดสอบรัน (Run test) ตัวสถิติทดสอบตัวคูณลากรานจ์ (Lagrange Multiplier test), ตัวสถิติทดสอบบ็อก-เพียซ (Box-Pierce test) และสถิติทดสอบจุง-บ็อก (Ljung-Box test) เป็นต้น ซึ่งสถิติทดสอบแต่ละตัวมีข้อจำกัดและเหมาะสมกับข้อมูลที่แตกต่างกันออกไป และสำหรับการแก้ปัญหาการมีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อนสามารถทำได้หลายวิธีขึ้นอยู่กับสาเหตุของการมีอัตตสหสัมพันธ์ เช่นถ้าเกิดจากการขาดตัวแปรอิสระที่สำคัญบางตัวในตัวแบบการถดถอย จะแก้ไขปัญหาโดยการเพิ่มตัวแปรอิสระที่สำคัญบางตัวเข้ามาในรูปแบบการถดถอย หรือหากปัญหาการมีอัตตสหสัมพันธ์เกิดจากการสร้างตัวแบบการถดถอยไม่ถูกต้อง จะแก้ไขโดยการเปลี่ยนตัวแบบการถดถอย แต่ถ้าทดสอบแล้วว่าการมีอัตตสหสัมพันธ์เกิดจากเทอมของค่าคลาดเคลื่อนเองในช่วงเวลาหนึ่งส่งผลกระทบไปยังช่วงเวลาอื่นๆ เราจะต้องหาค่าพารามิเตอร์ของตัวแบบ ซึ่งวิธีการแก้ไขนี้กระทำได้หลายวิธี ในแต่ละวิธีจะมีหลักการเหมือนกันประการหนึ่งที่สำคัญ คือ ต้องคำนวณหาค่าประมาณของพารามิเตอร์อัตตสหสัมพันธ์ (r) แล้วนำไปแปลงตัวแปรทุกตัวเพื่อขจัดปัญหาการมีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อน

การศึกษาวิจัยครั้งนี้สนใจที่จะศึกษาเปรียบเทียบประสิทธิภาพวิธีการแก้ปัญหาการมีอัตตสหสัมพันธ์โดยการหาค่าพารามิเตอร์ด้วยวิธีการ 3 วิธี ได้แก่ วิธีคอคแครนและออร์คัต (Cochrane-Orcutt Method), วิธีฮิลเดรธและลู (Hildreth-Lu Method) และวิธีผลต่างอันดับที่หนึ่ง (First Differences Method) เปรียบเทียบกับวิธีกำลังสองน้อยที่สุด (OLS) เมื่อค่าคลาดเคลื่อนเกิด

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

อัตตสหสัมพันธ์อันดับที่ 1 และเปรียบเทียบประสิทธิภาพของวิธีคอคแครนและออร์คัต, วิธีฮิลเดรธและลู กับวิธีกำลังสองน้อยที่สุด เมื่อค่าคลาดเคลื่อนเกิดอัตตสหสัมพันธ์อันดับที่ 2 และ 3 เนื่องจากวิธีผลต่างอันดับที่หนึ่งเป็นวิธีการประมาณค่าพารามิเตอร์ที่สมมติให้ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าเท่ากับหนึ่ง เพราะค่าสัมประสิทธิ์สหสัมพันธ์มักมีค่ามาก และผลรวมกำลังสองของความคลาดเคลื่อน (error sum of squares : SSE) เป็นฟังก์ชันของค่าสัมประสิทธิ์สหสัมพันธ์ที่มากขึ้นจนเกือบจะเท่ากับหนึ่ง ซึ่งจะเห็นว่ามีค่ามากกว่าค่าสัมประสิทธิ์สหสัมพันธ์ที่แท้จริง (บางครั้งอาจมากกว่าค่าสัมประสิทธิ์ที่แท้จริงอย่างมาก) ดังนั้นการกำหนดให้ค่าสัมประสิทธิ์สหสัมพันธ์อันดับที่ 2 และ 3 มีค่าใกล้เคียงหรือ เท่ากับหนึ่งด้วยนั้นอาจทำให้การประมาณค่าพารามิเตอร์ผิดพลาดได้ โดยการทดลองจะสร้างข้อมูลที่มีปัญหาอัตตสหสัมพันธ์ของค่าคลาดเคลื่อนด้วยเทคนิคการจำลองมอนติคาร์โล (Monte Carlo Simulation Technique) แล้วใช้วิธีการทั้ง 3 วิธีดังกล่าวแก้ปัญหาและตรวจสอบค่าคลาดเคลื่อนที่ได้จากแต่ละวิธีว่ามีอัตตสหสัมพันธ์กันหรือไม่ ด้วยสถิติทดสอบตัวคูณลากรางจ์ แล้วเปรียบเทียบประสิทธิภาพของวิธีการดังกล่าวเพื่อที่จะได้เสนอแนะวิธีการแก้ปัญหาที่เหมาะสมสำหรับข้อมูลที่มีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อน

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาวิธีการแก้ปัญหาการมีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อน ในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ โดยการหาค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) ด้วยวิธีการ 3 วิธี ได้แก่ วิธีคอคแครนและออร์คัต วิธีฮิลเดรธและลู และวิธีผลต่างอันดับที่หนึ่ง ซึ่งทำให้ได้ตัวแบบการถดถอยและค่าประมาณสัมประสิทธิ์การถดถอยที่เหมาะสม

2. เปรียบเทียบประสิทธิภาพในการแก้ปัญหาการมีอัตตสหสัมพันธ์ 3 วิธีดังกล่าว โดยพิจารณาจากค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean Square Error : MSE) และค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination : R^2)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

1.3 ขอบเขตของการวิจัย

1. ตัวแบบการถดถอยที่ใช้ในการศึกษาเป็นตัวแบบการถดถอยเชิงเส้นพหุคูณ เมื่อข้อมูลเป็นอนุกรมเวลา ซึ่งมีตัวแบบการถดถอยเป็นดังนี้

$$Y_t = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + \beta_4 X_{t4} + \varepsilon_t$$

โดยที่ Y_t เป็นตัวแปรตาม

$X_{t1}, X_{t2}, X_{t3}, X_{t4}$ เป็นตัวแปรอิสระ

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ เป็นค่าสัมประสิทธิ์การถดถอย (Regression Coefficient)

และ ε_t เป็นตัวแปรสุ่มค่าคลาดเคลื่อน ซึ่ง $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

2. ศึกษาภายใต้ลักษณะการของค่าคลาดเคลื่อน (ε_t) ที่มีการแจกแจงแบบปกติมีค่าเฉลี่ยเป็น 0 และค่าความแปรปรวน σ_ε^2 หรือ $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

3. ค่าคลาดเคลื่อน (ε_t) มีความสัมพันธ์กันโดยกำหนดความสัมพันธ์ของตัวแบบเป็นกระบวนการถดถอยอันดับที่ p : AR(p) ดังนี้

$$Y_i = \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + ... + \beta_k X_{tk} + \varepsilon_t$$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \rho_3 \varepsilon_{t-3} + ... + \rho_p \varepsilon_{t-p} + u_t$$

เมื่อ ρ_j คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง ε_t กับ ε_{t-j} , $j = 1,2,....p$ และ $|\rho| < 1$

u_t คือ ค่าคลาดเคลื่อนที่ไม่มีอัตตสหสัมพันธ์

โดยมีข้อตกลงเบื้องต้นของ u_t คือ มีการแจกแจงแบบปกติ (Normal Distribution) ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็นค่าคงที่ σ_u^2 , $E(u_t, u_s) = 0$ สำหรับ $t \neq s$

ในการวิจัยครั้งนี้จะทำการศึกษาการมีอัตตสหสัมพันธ์อันดับ 1, 2 และ 3 กล่าวคือ กำหนดความสัมพันธ์เป็นกระบวนการถดถอยอันดับที่ p เมื่อ $p = 1, 2$ และ 3

4. ค่าสัมประสิทธิ์การถดถอยที่ใช้มีค่าดังนี้

$$\beta_0 = 1, \beta_1 = 1, \beta_2 = 1, \beta_3 = 1, \beta_4 = 1$$

เนื่องจากได้มีผู้ทำการศึกษาในเรื่องอัตตสหสัมพันธ์ในความคลาดเคลื่อน พบว่า β_i เมื่อเปลี่ยนค่าไป จะไม่มีผลกระทบต่อผลสรุปของการตรวจสอบอัตตสหสัมพันธ์ในความคลาดเคลื่อนแต่ละวิธี (สิริพรรณ เฉลิมงาม, 2536) ดังนั้นผู้วิจัยจึงกำหนดค่า $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$ เป็นค่าคงที่ เพื่อลดสถานการณ์ในการเปลี่ยนแปลงค่าพารามิเตอร์ต่างๆ

5. ขนาดตัวอย่าง (n) ที่นำมาศึกษามี 4 ขนาด คือ 15, 30, 40 และ 60

7. ระดับความรุนแรงของอัตตสหสัมพันธ์แบ่งเป็น 3 ระดับ ดังนี้

ระดับต่ำ คือ $\rho = 0.1, 0.3$

ระดับปานกลาง คือ $\rho = 0.5$

ระดับสูง คือ $\rho = 0.7, 0.9$

ตามที่ทราบโดยทั่วไปแล้วว่าค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) มีค่าอยู่ในช่วง $-1 \leq \rho \leq 1$ แต่เนื่องจากในทางปฏิบัติโดยเฉพาะข้อมูลทางธุรกิจและเศรษฐศาสตร์ ส่วนมากจะพบว่าค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวก และเนื่องจากความสมมาตรของค่าสัมประสิทธิ์สหสัมพันธ์

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

(วิชัย สุรเชิดเกียรติ, 2535) ดังนั้นผู้วิจัยจึงขอเลือกกำหนดว่า $\rho \geq 0$

6. เปรียบเทียบประสิทธิภาพระหว่างวิธีการแก้ปัญหาอัตตสหสัมพันธ์ทั้ง 3 วิธีดังกล่าวกับวิธีกำลังสองน้อยที่สุด (OLS) โดยใช้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (Mean Square Error : MSE) และค่าสัมประสิทธิ์การตัดสินใจ (R^2)

8. ระดับนัยสำคัญของการทดสอบ $\alpha = 0.01$ และ 0.05

9. สำหรับวิธีการประมาณค่าพารามิเตอร์การถดถอยในกรณีเกิดอัตตสหสัมพันธ์อันดับที่ 1 จะทำการเปรียบเทียบ 3 วิธีดังที่ได้กล่าวมาแล้ว ส่วนกรณีเกิดอัตตสหสัมพันธ์อันดับที่ 2 และ 3 จะใช้วิธีคอคแครนและออร์คัต เปรียบเทียบกับวิธีฮิลเดรธและลู

1.4 นิยามศัพท์

1. ตัวแปรตาม (Dependent Variable) หมายถึง ตัวแปรที่เราต้องการจะศึกษา

2. ตัวแปรอิสระ (Independent Variable) หมายถึง ตัวแปรที่แทนอิทธิพลหรือปัจจัยที่มีผลต่อตัวแปรตามที่ต้องการจะศึกษา

3. ตัวแบบการถดถอย (Regression Model) หมายถึง สมการหรือตัวแบบทางสถิติที่สร้างขึ้นเพื่อใช้แทนลักษณะความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ

4. ข้อมูลอนุกรมเวลา (Time Series Data) หมายถึง ข้อมูลตัวเลขที่ถูกจัดเรียงตามเวลาที่บันทึกข้อมูลนั้น

5. การมีอัตตสหสัมพันธ์ (Autocorrelation) คือ ปัญหาที่เกิดขึ้นเนื่องจากค่าคลาดเคลื่อนมีการแจกแจงที่ไม่เป็นอิสระกัน หรือกล่าวได้ว่า ค่าคลาดเคลื่อนมีความสัมพันธ์ซึ่งกันและกันในแต่ละช่วงเวลา

6. ค่าเฉลี่ยกำลังสองของค่าคลาดเคลื่อน (Mean Square Error) เป็นค่าวัดความถูกต้องของการพยากรณ์ที่วัดจากค่าคลาดเคลื่อน

7. สัมประสิทธิ์การตัดสินใจ (Coefficient of Determination) เป็นค่าที่วัดว่าตัวแปรอิสระที่กำหนดในตัวแบบมีส่วนในการอธิบายความผันแปรของตัวแปรตามมากน้อยเพียงใด

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

1.5 ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบแนวคิด ทฤษฎี และวิธีการแก้ปัญหาอัตตสหสัมพันธ์ได้อย่างมีประสิทธิภาพ

2. สามารถเลือกใช้วิธีการแก้ปัญหาอัตตสหสัมพันธ์ได้อย่างเหมาะสมกับลักษณะข้อมูล และสามารถนำข้อมูลที่ได้รับการแก้ปัญหาอัตตสหสัมพันธ์แล้วไปใช้ในการอนุมานได้อย่างถูกต้อง

3. เพื่อใช้เป็นเอกสารอ้างอิงสำหรับนักศึกษาและผู้สนใจที่จะศึกษาค้นคว้าเกี่ยวกับการมีอัตตสหสัมพันธ์ของค่าคลาดเคลื่อน