บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

การบริหารงานในองค์กรจำเป็นต้องมีการวางแผนงานและการตัดสินใจ และวิธีที่นำมาใช้ได้อย่างมีประสิทธิภาพสำหรับการวางแผนงานและการตัดสินใจคือ วิธีการพยากรณ์ การพยากรณ์นั้นจะมีส่วนประกอบที่สำคัญอยู่ 2 ส่วน คือ การสร้างตัวแบบเพื่อใช้ในการพยากรณ์ และการนำตัวแบบที่ได้ไปใช้พยากรณ์ วิธีการพยากรณ์มีอยู่หลายวิธี การวิเคราะห์การถดถอยเชิงเส้นเป็นวิธีการหนึ่งที่สามารถนำไปใช้ในการพยากรณ์เกี่ยวกับตัวแปรที่ต้องการศึกษา ซึ่งเรียกว่า ตัวแปรตาม โดยอาศัยความรู้เกี่ยวกับค่าของตัวแปรอื่น ๆ ที่เกี่ยวข้อง ซึ่งเรียกว่า ตัวแปรอิสระ ถ้ามีตัวแปรอิสระเพียงตัวเดียว จะเรียกการวิเคราะห์ในลักษณะนี้ว่า การวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย (Simple Linear Regression) และสำหรับกรณีที่มีตัวแปรอิสระหลายตัว จะเรียกว่า การวิเคราะห์การถดถอยเชิงเส้นแบบพหุ (Multiple Linear Regression) ซึ่งเป็นวิธีการพยากรณ์ที่สนใจในการศึกษาครั้งนี้

ในการวิเคราะห์การถดถอยเชิงเส้นแบบพหุ ส่วนที่สำคัญคือการสร้างตัวแบบการถดถอยเชิงเส้นที่แสดงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม โดยสามารถเขียนอยู่ในรูปของเมทริกซ์ได้ดังนี้

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{e} \tag{1.1}$$

โดยที่

$\mathbf{Y}$ คือเวกเตอร์ของตัวแปรตามที่ทราบค่า มีมิติ $n \times 1$

$\mathbf{X}$ คือเมทริกซ์ของตัวแปรอิสระที่ทราบค่า มีมิติ $n \times p$ และมี rank = p

$\mathbf{b}$ คือเวกเตอร์ของสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า $b_1, b_2, ..., b_p$

$\mathbf{e}$ คือเวกเตอร์ของความคลาดเคลื่อน มีมิติ $n \times 1$

n คือจำนวนค่าสังเกต

และ p คือจำนวนตัวแปรอิสระ

โดยทั่วไปแล้วในการประมาณค่าพารามิเตอร์นั้น ผู้วิจัยจะเลือกใช้วิธีกำลังสองน้อยที่สุดแบบสามัญ (Ordinary Least Squares : OLS) ซึ่งเป็นวิธีที่ได้ตัวประมาณที่มีคุณสมบัติไม่เอนเอียงและให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดในบรรดาตัวประมาณเชิงเส้นที่ไม่เอนเอียงทั้งหมด (Best Linear Unbiased Estimator : BLUE) โดยมีข้อสมมติเกี่ยวกับความคลาดเคลื่อน

มหาวิทยาลัยบูรพา Burapha University

คือ ค่าเฉลี่ยของความคลาดเคลื่อนมีค่าเป็นศูนย์ และความแปรปรวนของความคลาดเคลื่อนมีค่าคงที่ และเป็นอิสระต่อกัน นั่นคือ

$$E(\mathbf{e}) = \mathbf{0} \text{ และ } Cov(\mathbf{e}) = E(\mathbf{ee}') = \sigma^2\mathbf{I}$$

โดยที่

$\mathbf{I}$ คือเมทริกซ์เอกลักษณ์ มิติ $n \times n$

σ^2 คือค่าคงตัวที่ไม่ทราบค่า

$Cov(\mathbf{e})$ คือเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมของความคลาดเคลื่อน ซึ่งเรียกข้อสมมติดังกล่าวนี้ว่า ข้อสมมติของ OLS

อย่างไรก็ตามในทางปฏิบัติอาจพบว่าข้อมูลที่นำมาใช้ในการวิเคราะห์นั้นไม่เป็นไปตามข้อสมมติดังกล่าว ซึ่งจะพบมากในการวิเคราะห์ข้อมูลอนุกรมเวลาทางด้านเศรษฐศาสตร์ ปัญหาที่พบมากปัญหาหนึ่ง คือ ความคลาดเคลื่อนมีความสัมพันธ์ต่อกัน หรือความคลาดเคลื่อนมีอัตตสหสัมพันธ์ (Autocorrelation) และรูปแบบที่พบโดยทั่วไปของอัตตสหสัมพันธ์ของความคลาดเคลื่อนสำหรับข้อมูลอนุกรมเวลา คือ อัตตสหสัมพันธ์อันดับที่ 1 (First Order Autocorrelation) ในกรณีเช่นนี้ถ้าทำการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบสามัญ จะทำให้เกิดผลกระทบต่อคุณภาพของตัวประมาณของพารามิเตอร์ ดังนั้นวิธีการที่จะนำมาใช้ในการแก้ปัญหาดังกล่าวคือ ใช้วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดแบบทั่วไป (Generalized Least Squares : GLS) แทนวิธีกำลังสองน้อยที่สุดแบบสามัญ ซึ่งมีข้อสมมติ ดังนี้

$$E(\mathbf{e}) = \mathbf{0} \text{ และ } Cov(\mathbf{e}) = E(\mathbf{ee}') = \sigma^2\mathbf{V}$$

โดยที่ $\mathbf{V}$ คือเมทริกซ์อัตตสหสัมพันธ์ของ e_i มิติ $n \times n$ ที่มี rank = n

และจะเรียกข้อสมมติดังกล่าวนี้ว่า ข้อสมมติของ GLS

ในการประมาณค่าสัมประสิทธิ์การถดถอยเชิงเส้นแบบพหุด้วยวิธีกำลังสองน้อยที่สุดแบบสามัญ มีข้อสมมติที่จำเป็นอีกข้อหนึ่ง คือ ตัวแปรอิสระจะต้องไม่มีความสัมพันธ์กัน ซึ่งในทางปฏิบัติเป็นไปได้น้อยมากที่ข้อสมมตินี้จะเป็นจริง เนื่องจากตัวแปรต่าง ๆ ที่นำมาศึกษาอาจมีความสัมพันธ์กัน หรือมีพหุสัมพันธ์ระหว่างตัวแปรอิสระ (Multicollinearity) ทำให้ความแปรปรวนและความแปรปรวนร่วมของตัวประมาณค่าสัมประสิทธิ์การถดถอยมีค่าสูง

วิธีการแก้ปัญหาที่นิยมใช้คือ การตัดตัวแปรบางตัวที่มีความสัมพันธ์กับตัวแปรอื่นออกไปจากตัวแบบ เช่น วิธีเพิ่มตัวแปรอิสระ (Forward Selection) วิธีลดตัวแปรอิสระ(Backward Elimination) วิธีเพิ่มตัวแปรอิสระแบบขั้นตอน (Stepwise regression) และการวิเคราะห์

มหาวิทยาลัยบูรพา Burapha University

ส่วนประกอบหลัก (Principal Component Analysis) เป็นต้น แต่ในกรณีที่ไม่ต้องการตัดตัวแปรอิสระตัวใดตัวหนึ่งออกไปจากตัวแบบเนื่องจากตัวแปรอิสระเหล่านั้นอาจมีความสำคัญ จึงมีผู้ศึกษาวิธีการที่ใช้แก้ปัญหาพหุสัมพันธ์เชิงเส้นดังกล่าวไว้หลายวิธี ซึ่งวิธีหนึ่งที่น่าสนใจคือ วิธีวิเคราะห์การถดถอยแบบริดจ์ (Ridge Regression) ที่นำเสนอโดย Hoerl and Kennard (1970)

ทั้งปัญหาการมีอัตตสหสัมพันธ์ และการมีพหุสัมพันธ์ระหว่างตัวแปรอิสระ เป็นปัญหาที่สำคัญยิ่งของการวิเคราะห์การถดถอยเชิงเส้นแบบพหุ ซึ่งส่วนใหญ่แล้วการทดสอบและการแก้ปัญหาเหล่านี้จะกระทำแยกส่วนกัน แต่ในการศึกษาครั้งนี้ผู้วิจัยต้องการที่จะแก้ปัญหาทั้งสองอย่างนี้พร้อม ๆ กัน

วัตถุประสงค์ของการวิจัย

1. เพื่อเสนอตัวประมาณค่าสัมประสิทธิ์การถดถอยเมื่อความคลาดเคลื่อนมีอัตตสหสัมพันธ์ และมีพหุสัมพันธ์ระหว่างตัวแปรอิสระด้วยวิธีใหม่

2. เพื่อศึกษาเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยจากการศึกษาครั้งนี้สำหรับข้อมูลทางด้านเศรษฐศาสตร์ และข้อมูลทางด้านธุรกิจประกันชีวิตโดยเปรียบเทียบกับตัวประมาณค่าสัมประสิทธิ์การถดถอยที่ได้จากวิธีดังต่อไปนี้

   2.1 วิธีกำลังสองน้อยที่สุดแบบสามัญ

   2.2 วิธีกำลังสองน้อยที่สุดแบบทั่วไป

   2.3 วิธีวิเคราะห์การถดถอยแบบริดจ์

   2.4 วิธีของ Bayhan

สัญลักษณ์ที่ใช้ในการวิจัย

1. $\hat{\mathbf{b}}_{OLS}$ หมายถึงตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดแบบสามัญ

2. $\hat{\mathbf{b}}_{GLS}$ หมายถึงตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยที่สุดแบบทั่วไป

3. $\hat{\mathbf{b}}_{RR}$ หมายถึงตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีวิเคราะห์การถดถอยแบบริดจ์

4. $\hat{\mathbf{b}}_{BH}$ หมายถึงตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีของ Bayhan

5. $\hat{\mathbf{b}}_{NEW}$ หมายถึงตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีใหม่

มหาวิทยาลัยบูรพา Burapha University

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ได้ตัวประมาณค่าสัมประสิทธิ์การถดถอยเมื่อเทอมความคลาดเคลื่อนมีอัตตสหสัมพันธ์ และมีพหุสัมพันธ์ระหว่างตัวแปรอิสระ

2. ให้ข้อสรุปในการเลือกวิธีการประมาณค่าสัมประสิทธิ์การถดถอยเพื่อการพยากรณ์ได้อย่างเหมาะสม กรณีที่เทอมความคลาดเคลื่อนมีอัตตสหสัมพันธ์ และมีพหุสัมพันธ์ระหว่างตัวแปรอิสระ

ขอบเขตของการวิจัย

1. ศึกษาลักษณะการแจกแจงของเทอมความคลาดเคลื่อนภายใต้ความคลาดเคลื่อนที่มีการแจกแจงปรกติ (Normal Distribution)

2. ศึกษาเมื่อมีพหุสัมพันธ์ระหว่างตัวแปรอิสระ และเทอมความคลาดเคลื่อนมีอัตตสหสัมพันธ์อันดับที่ 1 นั่นคือ ความคลาดเคลื่อนในช่วงเวลาที่ t มีความสัมพันธ์กับความคลาดเคลื่อนในช่วงเวลาที่ $t-1$

3. การวิจัยครั้งนี้ ได้ศึกษารูปแบบของสมการถดถอยเชิงเส้นแบบพหุ โดยประยุกต์ใช้กับข้อมูลทางด้านเศรษฐศาสตร์และข้อมูลทางด้านธุรกิจประกันชีวิตที่ประกอบด้วยตัวแปรอิสระ 2 ตัว ตัวแปรตาม 1 ตัว และทำการวิเคราะห์ข้อมูลโดยใช้โปรแกรมสำเร็จรูปมินิแทบ (Minitab)

เกณฑ์การตัดสินใจ

ในการวิจัยครั้งนี้ต้องการศึกษาเปรียบเทียบตัวประมาณค่าสัมประสิทธิ์การถดถอยด้วยวิธีใหม่ กับตัวประมาณค่าจากวิธีอื่น ๆ โดยประยุกต์ใช้กับข้อมูลทางด้านเศรษฐศาสตร์ และข้อมูลทางด้านธุรกิจประกันชีวิต ผู้วิจัยจึงใช้เมทริกซ์ความแปรปรวนและความแปรปรวนร่วมของค่าประมาณสัมประสิทธิ์การถดถอย $\left(Cov\left(\hat{\mathbf{b}}\right)\right)$ เป็นเกณฑ์ในการตัดสิน โดยมีหลักการดังนี้

1. เปรียบเทียบค่าความแปรปรวนของสัมประสิทธิ์การถดถอย ถ้าค่าที่ได้จากการประมาณค่าด้วยวิธีใหม่มีค่าต่ำกว่าวิธีอื่น แสดงว่าตัวประมาณค่าด้วยวิธีใหม่มีคุณสมบัติของตัวประมาณที่ดีกว่า

2. เปรียบเทียบเครื่องหมายของค่าความแปรปรวนร่วมของสัมประสิทธิ์การถดถอย ถ้าเครื่องหมายมีค่าเป็นบวก แสดงว่าเครื่องหมายของค่าความแปรปรวนร่วมของสัมประสิทธิ์การถดถอยถูกต้อง

มหาวิทยาลัยบูรพา Burapha University

นิยามศัพท์เฉพาะ

1. อัตตสหสัมพันธ์ คือ เหตุการณ์การที่เทอมความคลาดเคลื่อนมีความสัมพันธ์กัน กล่าวคือ $Cov(e_t, e_{t-1}) \neq 0$

2. การมีพหุสัมพันธ์ระหว่างตัวแปรอิสระ คือ เหตุการณ์ที่ตัวแปรอิสระมีความสัมพันธ์กัน

มหาวิทยาลัยบูรพา Burapha University