บทที่ 1

บทนำ

ที่มาและความสำคัญของปัญหา

การวิเคราะห์ความถดถอยพหุคูณ (Multiple Regression Analysis) เป็นวิธีการหารูปแบบความสัมพันธ์ของตัวแปรตาม (dependent variable) กับตัวแปรอิสระ (independent variables) ชุดหนึ่ง มีหลักเกณฑ์ว่าการใช้ตัวแปรอิสระที่เหมาะสมมากกว่าหนึ่งตัวโดยทั่วไปย่อมทำให้ผลการประมาณค่าตัวแปรตามที่ถูกต้องมากกว่าการใช้ตัวแปรอิสระเพียงตัวเดียว ถ้าตัวแปรอิสระนั้นไม่มีความสัมพันธ์กันและมีอิทธิพลต่อตัวแปรตามมากพอสมควร เราสามารถเขียนตัวแบบทั่วไป (general model) ที่แสดงความสัมพันธ์เชิงเส้น (linear relationship) ระหว่างตัวแปรอิสระและตัวแปรตามได้ดังนี้

$$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ คือ เวกเตอร์ของตัวแปรตามขนาด n x 1

X คือ เมทริกซ์ของตัวแปรอิสระขนาด n x (p+1)

$\underset{\sim}{\beta}$ คือ เวกเตอร์ของสัมประสิทธิ์การถดถอยพหุคูณขนาด (p+1) x 1

$\underset{\sim}{\varepsilon}$ คือ เวกเตอร์ค่าความคลาดเคลื่อนที่เกิดขึ้นขนาด n x 1

โดยที่ $E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0}$, $\text{cov}(\underset{\sim}{\varepsilon}) = \sigma^2 I_n$

n คือ ขนาดตัวอย่าง

และ p คือ จำนวนตัวแปรอิสระ

ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณจากตัวแบบดังกล่าวนี้ วิธีที่นิยมใช้มากที่สุดคือวิธีกำลังสองน้อยสุด (least square method) โดยจะได้รูปแบบตัวประมาณเป็น $\hat{\underset{\sim}{\beta}} = (X'X)^{-1}X'\underset{\sim}{y}$ ซึ่งเป็นตัวประมาณที่ไม่เอนเอียงและให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด แต่ในการประมาณค่าสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีกำลังสองน้อยสุด มีสมมติฐานที่จำเป็นข้อหนึ่ง คือ ตัวแปรอิสระต้องไม่มีความสัมพันธ์กันในลักษณะเชิงเส้น ซึ่งในทางปฏิบัติเป็นไปได้น้อยมาก เพราะตัวแปรบางตัวที่นำมาศึกษาอาจเป็นฟังก์ชันของตัวแปรอิสระตัวอื่น นั่นคือ ตัวแปรอิสระมีพหุสัมพันธ์ (multicollinearity) ทำให้การประมาณค่าตัวแปรตามอาจไม่เหมาะสมและมีผลทำให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยของค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณมีค่ามากขึ้น กล่าวคือ ค่าประมาณ

สัมประสิทธิ์การถดถอยพหุคูณที่ได้ขาดความเที่ยงตรง (accuracy) ซึ่งอาจแก้ไขโดยการตัดตัวแปรบางตัวที่มีความสัมพันธ์กับตัวแปรอื่นออกจากตัวแบบ ถ้าหากความสัมพันธ์ระหว่างตัวแปรอิสระนั้นชัดเจนพอ แต่ในบางครั้งการตัดตัวแปรอิสระตัวใดตัวหนึ่งออกจากตัวแบบทำได้ยาก เนื่องจากลักษณะความสัมพันธ์ระหว่างตัวแปรอิสระที่เกิดขึ้นไม่ชัดเจน และถือว่าตัวแปรอิสระทุกตัวมีผลต่อการเปลี่ยนแปลงของตัวแปรตามบ้างพอสมควร [1]

"ในปี ค.ศ. 1970 โฮเอริ่น (Hoerl) และเคนนาร์ด (Kennard) ได้เสนอวิธีริดจ์รีเกรสชัน (Ridge Regression Method) ซึ่งเป็นวิธีการที่ไม่ต้องตัดตัวแปรอิสระออกจากตัวแบบ ถึงแม้จะเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ โดยวิธีนี้จะให้ค่าประมาณสัมประสิทธิ์การถดถอยพหุคูณที่ให้ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยต่ำกว่าวิธีกำลังสองน้อยสุด การที่จะลดค่าความคลาดเคลื่อนกำลังสองเฉลี่ยให้ต่ำลงจึงต้องพยายามลดค่า $(X'X)^{-1}$ โดยการบวกค่าคงที่ที่มากกว่าศูนย์กับสมาชิกทุกตัวในแนวทแยงมุมของเมทริกซ์ $(X'X)$ เนื่องจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ยเป็นฟังก์ชันของ $(X'X)^{-1}$ [2] โดยจะได้รูปแบบการประมาณสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีริดจ์รีเกรสชัน คือ

$$\hat{\beta}_R = (X'X + k\,I)^{-1}X'\,y \qquad ;\ k > 0$$

ตัวประมาณสัมประสิทธิ์การถดถอยพหุคูณด้วยวิธีริดจ์รีเกรสชันเป็นตัวประมาณที่มีความเอนเอียง (bias) และจะต้องมีการประมาณค่าพารามิเตอร์ k ซึ่งได้มีผู้เสนอวิธีการประมาณค่าพารามิเตอร์ k หลายวิธี เช่น วิธีของ Hoerl, A.E., Kennard, and Baldwin, K.F. (HKB) วิธีของ Tze-San-Lee (TZE) เป็นต้น ซึ่งวิธีประมาณค่าพารามิเตอร์เหล่านี้เป็นวิธีที่ค่อนข้างยุ่งยากซับซ้อน และค่าพารามิเตอร์ที่คำนวณได้ยังอาจจะไม่เป็นค่า k ที่เหมาะสมที่สุด [3] ในปี พ.ศ. 2534 จิราพุธ พู่มนตรี ได้ทำการวิจัยเปรียบเทียบการประมาณค่าพารามิเตอร์ k ด้วยวิธี HKB, TZE และการค้นหาแบบทวิ (Binary Search) ซึ่งเป็นวิธีการในทางคอมพิวเตอร์ที่ใช้ในการค้นหาข้อมูลที่มีลักษณะเรียงลำดับ ผลสรุปที่ได้คือ วิธีการค้นหาแบบทวิจะให้ค่าประมาณพารามิเตอร์ k ดีที่สุด และในงานวิจัยของ ธันยากร ต้นชวจันช์ ได้ชี้ให้เห็นว่าวิธีการค้นหาแบบทวินั้นไม่เหมาะสมที่จะนำมาใช้ เนื่องจากวิธีค้นหาแบบทวิมีข้อจำกัด คือ จะใช้ได้กับข้อมูลที่มีลักษณะแบบไม่ต่อเนื่องและใช้ในการค้นหาข้อมูลที่มีลักษณะเรียงลำดับเท่านั้น ซึ่ง

---

[1] ธันยากร ต้นชวจันช์, "การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุด วิธีริดจ์รีเกรสชัน และวิธีที่ใช้หลักการของริดจ์และสไตน์ ในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ" (วิทยานิพนธ์ปริญญามหาบัณฑิตสาขาวิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538), หน้า 1-2.

[2] เรื่องเดียวกัน, หน้า 2

[3] เรื่องเดียวกัน, หน้า 19

ซับยากร ต้นชลขันธ์ ได้นำเสนอวิธีการประมาณค่าพารามิเตอร์ k แบบลำดับ (Sequential Search) ที่เป็นวิธีการค้นหาข้อมูลในทางคอมพิวเตอร์เช่นเดียวกับวิธีการค้นหาแบบทวิ แต่วิธีการนี้สามารถใช้ได้กับข้อมูลที่มีลักษณะทั้งแบบต่อเนื่องและแบบไม่ต่อเนื่อง และข้อมูลที่จะค้นหาจะเรียงลำดับหรือไม่เรียงลำดับก็ได้ ดังนั้นในงานวิจัยนี้ผู้วิจัยจึงได้เลือกใช้วิธีการค้นหาแบบลำดับประมาณค่าพารามิเตอร์ k

การวิเคราะห์ความถดถอยพหุคูณในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ นอกจากจะใช้วิธีริดจ์รีเกรสชันแล้วยังมีวิธีการที่ใช้หลักการของโครงข่ายประสาทเทียม (Artificial Neural Networks) ซึ่งเป็นอีกสาขาหนึ่งของหลักสูตรวิทยาการคอมพิวเตอร์ (Computer Sciences) ที่สามารถนำมาใช้ในการวิเคราะห์กรณีนี้ได้ เนื่องจากมีความยืดหยุ่นสูงจนสามารถจำลองกระบวนการของปัญหาใดๆ ได้ และไม่ว่าข้อมูลนำเข้าจะเป็นฟังก์ชันใดโครงข่ายประสาทเทียมก็สามารถประมาณฟังก์ชันผลลัพธ์ของโครงข่ายได้อย่างเที่ยงตรง รวมทั้งมีความสามารถในการนำชุดข้อมูลนำเข้าและชุดผลลัพธ์ที่มีความซับซ้อนมากจนไม่สามารถจำลองแบบในเชิงความน่าจะเป็นได้ และมีความสามารถในการตอบสนองต่อข้อมูลที่ไม่เคยเห็น กล่าวโดยสรุปคือ โครงข่ายประสาทเทียมสามารถที่จะแปลงรูปแบบความสัมพันธ์ ได้ด้วยตัวเองแสดงให้เห็นความซับซ้อนและความสัมพันธ์ที่ไม่เชิงเส้น และโครงข่ายประสาทเทียมเป็นวิธีที่เหมาะสมสำหรับการพยากรณ์ เพราะว่ามีความสามารถในการปรับปรุงตัวแบบด้วยตัวเอง แต่โครงข่ายประสาทเทียมยังมีข้อจำกัดที่เห็นได้อย่างชัดเจนมากที่สุด คือ ไม่มีการอธิบายตัวแบบและทดสอบสมมติฐานในการพยากรณ์ เนื่องจากสัมประสิทธิ์การถดถอยเป็นแบบไม่เชิงเส้น ซึ่งตัวแบบความน่าจะเป็นปกติจะใช้ไม่ได้ นั่นคือ โครงข่ายประสาทเทียมไม่มีพารามิเตอร์ที่มีคุณสมบัติทางสถิติ นอกจากนี้ข้อจำกัดอีกข้อหนึ่งคือ โครงข่ายประสาทเทียมมีฟังก์ชันความคลาดเคลื่อนกำลังสองได้หลายค่า เนื่องจากการที่โครงข่ายประสาทเทียมพยายามปรับตัวแบบให้มีความคลาดเคลื่อนเท่ากับหรือต่ำกว่าที่กำหนด หลักการทำงานของโครงข่ายประสาทเทียมคือ จะทำการสร้างกระบวนการ (process) ซึ่งได้มาโดยการเรียนรู้จากตัวอย่าง (learning by examples) ถ้าตัวอย่างที่นำมาฝึก (train) หรือสอนแก่โครงข่ายมีจำนวนมากพอและวิธีการสอนดีพอ กระบวนการที่สร้างขึ้นก็จะมีความถูกต้องเพียงพอสำหรับการนำไปใช้งานจริงได้

ในงานวิจัยนี้จะใช้โครงข่ายประสาทเทียมที่มีแบบจำลองเพอร์เซปตรอน 3 ชั้น ซึ่งประกอบด้วย ชั้นนำเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นผลลัพธ์ (Output Layer) เหตุผลที่ใช้ชั้นซ่อนเพียงชั้นเดียวเนื่องจากในการศึกษาของ Minsky และ Papert ได้แสดงให้เห็นว่าโครงข่ายเพียงชั้นเดียว (Single-layer net) ก็สามารถเรียนรู้เฉพาะปัญหาเชิงเส้น และเป็นการง่ายที่จะขยายผลนี้ไปแสดงในโครงข่ายหลายชั้น (Multilayer nets) และจากบทความของ จุฬารัตน์ ต้นประเสริฐ และธิติพงษ์ ต้นประเสริฐ ได้อ้างว่ามีทฤษฎีที่กล่าวไว้ว่า ชั้นซ่อนเพียงชั้นเดียวก็เพียงพอสำหรับการจำลองปัญหาที่ไม่เชิงเส้นใดๆ แต่ในบางกรณีการเพิ่มชั้นซ่อนก็สามารถช่วยให้การเรียนรู้ง่ายขึ้น ซึ่งในการวิเคราะห์การถดถอยพหุคูณเป็นเพียงกรณีหนึ่งของการถดถอยเชิงเส้นเท่านั้น ดังนั้นชั้นซ่อนเพียงชั้นเดียวก็เพียง

พอแล้ว ส่วนการเชื่อมต่อระหว่างชั้นเป็นแบบ Fully Connected Feedforward นั่นคือทุกนิวรอนในชั้นนำเข้าจะส่งสัญญาณไปยังทุกนิวรอนในชั้นซ่อน และทุกนิวรอนในชั้นซ่อนจะส่งสัญญาณไปยังทุกนิวรอนในชั้นผลลัพธ์ และใช้ขั้นตอนการเรียนรู้ (Learning algorithm) เป็นแบบการแพร่ย้อนกลับ (Backpropagation) ที่ใช้กันอย่างกว้างขวางในโครงข่ายประสาทเทียมมีหลักการที่สำคัญคือจะปรับน้ำหนักของแต่ละนิวรอนตามความผิดพลาด (error) ของโครงข่ายประสาทเทียมโดยการหาอนุพันธ์ (derivative) ของความผิดพลาดหรือกล่าวได้ว่าเป็นวิธีการปรับน้ำหนักเพื่อให้ได้ค่าที่เหมาะสมและใช้ความผิดพลาดของหน่วยผลลัพธ์เมื่อเปรียบเทียบกับค่าเป้าหมายมาเป็นตัวชี้นำในการปรับน้ำหนัก [4]

จากเหตุผลที่กล่าวไว้ข้างต้น จึงเป็นสิ่งที่น่าสนใจว่าเมื่อนำวิธีวิลคอกซันที่ใช้วิธีประมาณค่าพารามิเตอร์ k แบบค้นหาลำดับกับวิธีที่ใช้หลักการของโครงข่ายประสาทเทียม มาหาค่าพยากรณ์ในการวิเคราะห์ความถดถอยพหุคูณเมื่อเกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระภายใต้การแจกแจงของความคลาดเคลื่อน จำนวนตัวแปรอิสระ ขนาดตัวอย่าง และระดับของสัมประสิทธิ์การแปรผัน วิธีการใดจะให้ค่าพยากรณ์ที่ถูกต้องมากกว่ากัน

วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบความถูกต้องของค่าพยากรณ์เมื่อใช้วิธีวิลคอกซันและวิธีที่ใช้หลักการของโครงข่ายประสาทเทียมในการวิเคราะห์ความถดถอยพหุคูณ ในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ

ธรรมเนียมการวิจัย

วิธีที่ใช้หลักการของโครงข่ายประสาทเทียมจะให้ค่าพยากรณ์ที่ถูกต้องมากกว่าในกรณีที่ความคลาดเคลื่อนมีการแจกแจงปกติปลอมปน และการแจกแจงแบบเบ้ เมื่อขนาดตัวอย่างมีค่าสูง ตัวแปรอิสระมีความสัมพันธ์สูง และวิธีที่ใช้หลักการของโครงข่ายประสาทเทียมให้ผลใกล้เคียงกับวิธีวิลคอกซัน เมื่อความคลาดเคลื่อนมีการแจกปกติ ขนาดตัวอย่างมีค่าสูง และตัวแปรอิสระมีความสัมพันธ์สูง

---

[4] อุทารัตน์ ดันประเสริฐ และชิดพงษ์ ดันประเสริฐ, " โครงข่ายประสาทเทียม (Artificial Neural Networks) อีกแนวทางหนึ่งของการแก้ปัญหาด้วยคอมพิวเตอร์ (ตอนที่ 2)," สาร NECTEC (ตุลาคม-ธันวาคม 2538) : 19-23.

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ขอบเขตการวิจัย

1. <u>ความคลาดเคลื่อนมีการแจกแจงปกติ (Normal Distribution)</u>

การแจกแจงของความคลาดเคลื่อนเป็นไปตามข้อสมมติเบื้องต้นของสมการถดถอย คือ $\underset{\sim}{\varepsilon} \sim N(\underset{\sim}{0}, \sigma^2 I_n)$ ซึ่งฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} \qquad , \ \sigma > 0$$

1.1 จำนวนตัวแปรอิสระที่ใช้ในการศึกษา มี 2 ระดับ คือ 3 และ 5

1.2 ขนาดตัวอย่างที่ทำการศึกษามี 3 ขนาด คือ 30, 50 และ 100 โดยที่ในแต่ละระดับจะสร้างข้อมูลเพิ่มขึ้น 20 เพื่อใช้ในการทดสอบตัวแบบ

<u>หมายเหตุ</u> ที่เริ่มใช้ขนาดตัวอย่างเท่ากับ 30 เนื่องจากในทางปฏิบัติถือว่าขนาดตัวอย่างเท่ากับ 30 ลู่เข้าสู่การแจกแจงปกติ แต่ในทางทฤษฎีจะใช้ที่ขนาดตัวอย่างเท่ากับ 100 ขึ้นไป

1.3 ระดับสัมประสิทธิ์การแปรผัน (C.V.) ที่ศึกษามี 3 ระดับ คือ 10%, 30% และ 50%

1.3.1 กรณีตัวแปรอิสระเท่ากับ 3 ระดับพหุสัมพันธ์ของตัวแปรที่ศึกษามี 3 ระดับ คือ

ระดับต่ำ ρ = 0.1, 0.3

ระดับปานกลาง ρ = 0.5, 0.7

ระดับสูง ρ = 0.9, 0.99

เมื่อ ρ คือความสัมพันธ์ระหว่าง x_1 กับ x_2, x_1 กับ x_3, x_2 กับ x_3

1.3.2 กรณีตัวแปรอิสระเท่ากับ 5 ระดับพหุสัมพันธ์ของตัวแปรที่ศึกษามี 3 ระดับ คือ

ระดับต่ำ ρ = (0.1, 0.1), (0.3, 0.3)

ระดับปานกลาง ρ = (0.5, 0.5), (0.7, 0.7)

ระดับสูง ρ = (0.9, 0.9), (0.99, 0.99)

เมื่อ ρ ของค่าแรกในวงเล็บ คือ ความสัมพันธ์ระหว่าง x_1 กับ x_2, x_1 กับ x_3, x_2 กับ x_3
และ ρ ของค่าหลังในวงเล็บ คือ ความสัมพันธ์ระหว่าง x_4 กับ x_5

<u>หมายเหตุ</u> ในกรณีนี้ไม่ได้พิจารณาเมื่อตัวแปรอิสระ x_4 หรือ x_5 มีความสัมพันธ์กับตัวแปรอิสระ x_1, x_2 และ x_3 ซึ่งอาจมีผลกับค่าความคลาดเคลื่อนที่ได้ ซึ่งในที่นี้จะพิจารณาในกรณีที่เราเพิ่มตัวแปรอิสระเข้าไปมีผลทำให้ค่าความคลาดเคลื่อนที่ได้เปลี่ยนแปลงไปอย่างไร

2. ความคลาดเคลื่อนมีการแจกแจงปกติปลอมปน (Scale-Contaminated Normal Distribution) ฟังก์ชันการแจกแจงอยู่ในรูปของ

$$F = (1-p)N(\underset{\sim}{1},\sigma^2) + p\,N(\underset{\sim}{1},c^2\sigma^2)$$

เมื่อ c คือ สเกลแฟคเตอร์ (scale factor) ถ้าสเกลแฟคเตอร์มีค่าสูงจะทำให้เกิดค่าสังเกตที่ผิดปกติ มีค่าสูงด้วย ในงานวิจัยนี้จะศึกษาที่ c = 3 และ 10 และ p คือเปอร์เซ็นต์การปลอมปน (percent of contamination)ในกรณีนี้จะใช้ p = 5 และ 10

3. ความคลาดเคลื่อนมีการแจกแจงแบบเบ้

ผู้วิจัยจะศึกษาในกรณีการแจกแจงลอกนอร์มอล (Lognormal Distribution) ซึ่งมีฟังก์ชันความหนาแน่นอยู่ในรูปของ

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\} \qquad , \; x>0$$

ผู้วิจัยสนใจศึกษาเมื่อค่าเฉลี่ย (μ) เท่ากับ 1 และความแปรปรวน (σ^2) เท่ากับ 0.05, 0.3 และ 0.7 ซึ่งสอดคล้องกับค่า C.V. = 22.64%, 59.15% และ 100.69% ตามลำดับ สาเหตุที่เลือกใช้ C.V. = 22.64% โดยไม่เลือกค่า C.V. ที่มีค่าต่ำกว่านี้เพราะเมื่อพิจารณากราฟการแจกแจงลอกนอร์มอล ถ้า C.V. มีค่าต่ำกว่านี้กราฟของการแจกแจงจะค่อนเข้าสู่การแจกแจงปกติมากขึ้น [5]

เกณฑ์การตัดสินใจ

การเปรียบเทียบประสิทธิภาพของวิธีการทั้งสอง ในการวิจัยนี้จะพิจารณาจาก

1. ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ย (Average Mean Square Error : AMSE) โดยใช้ข้อมูล 20 ชุดสุดท้ายจากการ generate มาหาค่าความคลาดเคลื่อน สาเหตุเนื่องจากการหาตัวแบบที่เหมาะสมของวิธีที่ใช้หลักการของโครงข่ายประสาทเทียมนั้นจะหยุดดำเนินการเมื่อความคลาดเคลื่อนในการเรียนรู้เท่ากับหรือต่ำกว่าที่กำหนดไว้ ดังนั้นจึงต้องใช้ข้อมูลที่ตัวแบบไม่เคยเห็น แต่ยังคงใช้การ Simulate ขึ้นมาภายใต้สถานการณ์เดียวกันนำมาหาค่าความคลาดเคลื่อนของตัวแบบ

---

[5] ธันยากร ต้นรอดจันธ์, "การเปรียบเทียบการประมาณค่าสัมประสิทธิ์การถดถอยพหุโดยวิธีกำลังสองน้อยที่สุด วิธีริดจ์รีเกรสชัน และวิธีที่ใช้หลักการของริดจ์และลาโซ ในกรณีที่เกิดพหุสัมพันธ์ระหว่างตัวแปรอิสระ" (วิทยานิพนธ์ปริญญามหาบัณฑิตสาขาวิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2538), หน้า 4-6.

2. เปอร์เซ็นต์อัตราส่วนผลต่างค่าเฉลี่ยของค่า AMSE ของทั้งสองวิธี ซึ่งสามารถเขียนได้ดังนี้

$$DIFF = \frac{AMSE_{(i)} - AMSE_{(min)}}{AMSE_{(min)}} \times 100 \qquad ; i = 1,2$$

เมื่อ $DIFF$ คือ เปอร์เซ็นต์อัตราส่วนผลต่างค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

$AMSE_{(min)}$ คือ ค่า AMSE ของวิธีที่ให้ค่าต่ำสุด

และ $AMSE_{(i)}$ คือ ค่า AMSE ของวิธีที่เหลือ

ประโยชน์ที่คาดว่าจะได้รับ

1. ผลการศึกษาเป็นแนวทางในการศึกษาเกี่ยวกับการหาค่าพยากรณ์ที่ถูกต้อง ในการวิเคราะห์ความฉลาดฉลาดพหุคูณ ในกรณีตัวแปรอิสระมีพหุสัมพันธ์เมื่อความคลาดเคลื่อนมีการแจกแจงปกติปลอมปน การแจกแจงแบบเบ้ และการแจกแจงปกติ

2. ผลการศึกษาเปรียบเทียบเป็นแนวทางในการเลือกใช้วิธีการพยากรณ์ ในแต่ละสถานการณ์ได้อย่างเหมาะสม

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย