บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการวิจัยสาขาต่าง ๆ ไม่ว่าจะเป็นทางด้านวิทยาศาสตร์ สังคมศาสตร์ อุตสาหกรรม เกษตรกรรม การตลาด และทางด้านธุรกิจ จำเป็นต้องอาศัยระเบียบวิธีทางสถิติเข้ามาช่วยในการค้นคว้าวิจัย และดำเนินงานอย่างเป็นระบบเพื่อให้งานวิจัยมีความถูกต้องและน่าเชื่อถือ การค้นคว้าหาคำตอบเพื่อคาดคะเนเหตุการณ์ในอนาคต หรือที่เรียกว่าการพยากรณ์เป็นระเบียบวิธีทางสถิติแขนงหนึ่งที่เข้าไปมีบทบาทสำคัญในการวิจัยสาขาต่าง ๆ ซึ่งนักวิจัยส่วนใหญ่จะเลือกใช้วิธีการวิเคราะห์ความถดถอย (regression analysis) ในการหารูปแบบความสัมพันธ์ระหว่าง 2 ตัวแปรหรือมากกว่า เพื่อนำไปสู่การพยากรณ์ค่าจริง

วิธีการวิเคราะห์ความถดถอยที่นิยมใช้กันอย่างแพร่หลายคือ การวิเคราะห์ความถดถอยเชิงเส้น (linear regression analysis) ซึ่งมีตัวแบบในรูปทั่วไปดังนี้

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

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตามขนาด n x 1

X เป็นเมทริกซ์ของตัวแปรอิสระขนาด n x p

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่ไม่ทราบค่าขนาด p x 1

$\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ของความคลาดเคลื่อนสุ่มขนาด n x 1

n เป็นขนาดตัวอย่าง

และ p เป็นจำนวนพารามิเตอร์สัมประสิทธิ์การถดถอยเชิงเส้น

การพิจารณาตัวแปรอิสระที่ใช้ในตัวแบบการถดถอยเชิงเส้นนั้น ถ้าใช้ตัวแปรอิสระเพียงตัวแปรเดียวจะเรียกว่าการวิเคราะห์ความถดถอยเชิงเส้นอย่างง่าย (simple linear regression) แต่ในสภาพความเป็นจริงตัวแปรอิสระที่มีอิทธิพลต่อตัวแปรตามอาจมีได้มากกว่า 1 ตัวแปร ทำให้ต้องใช้ตัวแปรอิสระหลายตัวแปรในตัวแบบการถดถอยเชิงเส้น การวิเคราะห์ในกรณีนี้เรียกว่าการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ (multiple linear regression) ซึ่งจะช่วยให้การพยากรณ์ถูกต้องและแม่นยำมากขึ้น แต่ในการใช้ตัวแปรอิสระจำนวนมากในตัวแบบการถดถอยอาจ

ไม่ส่งผลดีเสมอไป เพราะนอกจากในบางครั้งจะไม่ช่วยลดความคลาดเคลื่อนในการพยากรณ์แล้ว ยังอาจเป็นการเพิ่มความคลาดเคลื่อนในการพยากรณ์ก็เป็นได้ ดังนั้นจึงจำเป็นต้องมีการคัดเลือกตัวแบบที่ดีที่สุดเพื่อใช้ในการพยากรณ์ โดยวิธีที่ใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นที่ดีที่สุดสามารถจำแนกได้ 2 แนวทาง คือ [1]

1) ใช้ค่าสถิติในการพิจารณาความเหมาะสมของตัวแบบทุกรูปแบบที่เป็นไปได้ จะเรียกวิธีนี้ว่า วิธีพิจารณาทุกรูปแบบ (all possible regression) ค่าสถิติที่นิยมใช้ในวิธีนี้ ได้แก่ ค่าสัมประสิทธิ์การกำหนด (R^2) ค่าสัมประสิทธิ์การกำหนดที่ปรับแล้ว (R^2_{adj}) และค่าสถิติของมอลโลวส์ (*Mallows* C_p) เป็นต้น

2) ใช้วิธีการคัดเลือกตัวแปรอิสระ (variable selection) โดยการเพิ่ม และ/หรือลดตัวแปรอิสระจากตัวแบบการถดถอย ซึ่งวิธีที่นิยมใช้กันอย่างแพร่หลายได้แก่ วิธีการคัดเลือกตัวแปรแบบไปข้างหน้า (forward selection) วิธีการกำจัดตัวแปรแบบถอยหลัง (backward elimination) และวิธีการถดถอยแบบขั้นบันได (stepwise regression) เป็นต้น

การใช้เกณฑ์ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นที่ดีที่สุดใน 2 แนวทางดังกล่าวข้างต้น ล้วนเป็นวิธีที่นิยมใช้กันอย่างแพร่หลาย แต่เป็นที่น่าสนใจสำหรับนักสถิติหลายท่านในการหาวิธีการใหม่ ๆ สำหรับคัดเลือกตัวแบบการถดถอยเชิงเส้นที่ดีที่สุดภายใต้แนวทางอื่น ๆ เพื่อให้ได้วิธีการคัดเลือกตัวแบบการถดถอยที่มีประสิทธิภาพและเหมาะสมกว่าวิธีที่นิยมใช้ในปัจจุบัน การนำแนวทางของเบส์ (Bayesian Approach) มาใช้ในการหาเกณฑ์หรือวิธีการคัดเลือกตัวแบบการถดถอยเชิงเส้นที่ดีที่สุดเป็นแนวทางหนึ่งที่นักสถิติให้ความสนใจกันมากในปัจจุบัน เนื่องจากแนวทางของเบส์มีการนำข้อมูลหรือความรู้ในอดีตมาใช้ ดังนั้นการนำแนวทางของเบส์มาประยุกต์ใช้น่าจะมีแนวโน้มทำให้ได้วิธีการคัดเลือกตัวแบบที่มีประสิทธิภาพและเหมาะสมกว่าวิธีการภายใต้แนวทางอื่น ๆ

ในปี ค.ศ.1978 ชวาร์ซ (Schawrz) ได้เสนอเกณฑ์ในการคัดเลือกตัวแบบโดยใช้การพิจารณาค่าสูงสุดของความน่าจะเป็นภายหลัง (posterior probability) มาเป็นส่วนประกอบในการพิจารณา ซึ่งเรียกว่าเกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion (BIC)) วิธีการนี้ถือเป็นวิธีการพื้นฐานในการหาเกณฑ์ในการคัดเลือกตัวแบบภายใต้แนวทางของเบส์ หลักการของวิธีการนี้จะกำหนดการการแจกแจงก่อน (prior distribution) เป็นการแจกแจงแบบ

---

[1] ทรงศิริ แต้สมบัติ, **การวิเคราะห์การถดถอย** (กรุงเทพมหานคร : มหาวิทยาลัยเกษตรศาสตร์ ,2541), หน้า 205.

สม่ำเสมอ (uniform distribution) และอนุมานความน่าจะเป็นภายหลังจากความน่าจะเป็นก่อน (prior probability) รวมกับฟังก์ชันความควรจะเป็น (likelihood function) เพื่อหาค่าเกณฑ์ที่ใช้ในการพิจารณาหาตัวแบบที่ดีที่สุด

ในปี ค.ศ.1993 จอร์จ (George) และ แมคคัลลอช (McCulloch) ได้เสนอวิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS)) โดยใช้หลักการคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็น (Stochastic Search Variable Selection (SSVS)) ซึ่งใช้วิธีการสุ่มแบบกิบส์ (Gibbs Sampling) วิธีการนี้จะมีการกำหนดความน่าจะเป็นให้กับตัวแบบโดยใช้หลักการคัดเลือกตัวแปรโดยการค้นหาด้วยความน่าจะเป็นเพื่อกำหนดตัวแบบที่เหมาะสมกับข้อมูล เมื่อทราบการแจกแจงความน่าจะเป็นของพารามิเตอร์ในตัวแบบ และนอกจากนั้นในปี ค.ศ.1996 จอร์จ (George) และ แมคคัลลอช (McCulloch) ยังได้มีการนำเสนอแนวคิดวิธีการคัดเลือกตัวแปรของเบส์สำหรับตัวแบบเชิงเส้นนัยทั่วไป (generalized linear models) ซึ่งยังคงใช้หลักการเดิมอีกด้วย

ในปี ค.ศ.1997 ราฟเทอร์รี่ (Raftery) เมดิแกน (Madigan) และโฮเอ็ททิง (Hoeting) ได้เสนอวิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method (BMA)) สำหรับตัวแบบการถดถอยเชิงเส้น ซึ่งเป็นวิธีการพิจารณาหาค่าความน่าจะเป็นภายหลัง (posterior probability) สำหรับทุก ๆ ตัวแบบที่เราสนใจ และนำตัวแบบทุกตัวแบบที่เราสนใจมาเฉลี่ยกันโดยใช้ความน่าจะเป็นภายหลังของแต่ละตัวแบบเป็นตัวถ่วงน้ำหนัก เพื่อหาค่าพยากรณ์ที่เหมาะสม วิธีการเฉลี่ยตัวแบบของเบส์นี้มีการคำนึงถึงหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบ (Model Uncertainly) ซึ่งจากงานวิจัยของนักสถิติหลายท่าน เช่น เมดิแกน และยอร์ค (Madigan and York, 1995) และราฟเทอร์รี่ (Raftery, 1996) ได้ชี้ให้เห็นว่าความไม่แน่นอนของตัวแบบเป็นสิ่งที่สำคัญ และควรนำมาพิจารณาในการคัดเลือกตัวแบบ สำหรับการหาปริภูมิตัวแบบที่จะนำมาใช้ในวิธีการเฉลี่ยตัวแบบของเบส์นั้น ราฟเทอร์รี่ เมดิแกน และโฮเอ็ททิง (Raftery , Madigan and Hoeting, 1997) ได้นำเสนอไว้ 2 แนวทาง ซึ่งทำให้วิธีการเฉลี่ยตัวแบบของเบส์สามารถจำแนกได้ดังนี้ คือ

1) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window) (BMA_{occ})
2) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition(MC^3)) (BMA_{MC3})

โดยการวิจัยครั้งนี้ผู้วิจัยจะทำการศึกษาวิธีการเฉลี่ยตัวแบบของเบส์ทั้ง 2 แนวทางดังกล่าวข้างต้น

ในการวิจัยครั้งนี้ผู้วิจัยได้สนใจทำการศึกษาเปรียบเทียบวิธีการที่ใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ 5 วิธี คือ

1) วิธีการคัดเลือกตัวแบบที่ดีที่สุด โดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC))
2) วิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS))
3) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window) (BMA_{OCC})
4) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition(MC^3)) (BMA_{MC3})
5) วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

โดย 4 วิธีการแรกนั้นล้วนเป็นวิธีการภายใต้แนวทางของเบส์ ซึ่งแต่ละวิธีได้มีการนำแนวทางของเบส์ไปประยุกต์ใช้ในแบบต่าง ๆ จึงเป็นที่น่าสนใจว่าวิธีการใดจะให้ผลที่ดีที่สุดในการคัดเลือกตัวแบบการถดถอย และที่ผู้วิจัยได้นำวิธีการพื้นฐานคือ วิธีการถดถอยแบบขั้นบันได ซึ่งเป็นวิธีที่นิยมใช้อย่างแพร่หลาย และมีประสิทธิภาพดีในการคัดเลือกสมการถดถอยที่ดีที่สุดเมื่อการวิเคราะห์ความถดถอยเป็นไปตามข้อตกลงเบื้องต้น[1] มาเปรียบเทียบกับวิธีการภายใต้แนวทางของเบส์นั้น ทั้งนี้เพื่อศึกษาดูแนวโน้มว่าวิธีการคัดเลือกตัวแบบภายใต้แนวทางของเบส์ จะมีความเหมาะสมกว่าวิธีการพื้นฐานหรือไม่ นอกจากนี้การนำวิธีการเฉลี่ยตัวแบบของเบส์มาศึกษาเปรียบเทียบกับวิธีการอื่น ๆ ยังจะทำให้ทราบว่าการพิจารณาหลักการเกี่ยวกับความไม่แน่นอนของตัวแบบจะทำให้การพยากรณ์มีประสิทธิภาพมากขึ้นหรือไม่ ดังนั้นถ้าผลการศึกษาวิจัยพบว่าวิธีการใดมีประสิทธิภาพในการพยากรณ์มากที่สุดก็ควรจะนำไปใช้ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณที่ดีที่สุดต่อไป

---

[1] จะเด็จ สวรรค์ตรานนท์, "การเปรียบเทียบวิธีที่ใช้สำหรับการเลือกสมการถดถอยที่ดีที่สุด," (วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาสถิติ บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, 2530), หน้า 212.

พลตรีกรรม ...
จุฬาลงกรณ์มหาวิทยาลัย

1.2 วัตถุประสงค์ของการวิจัย

1) เพื่อศึกษาวิธีการคัดเลือกสมการถดถอยที่ดีที่สุด ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ 5 วิธี คือ

1.1) วิธีการคัดเลือกตัวแบบที่ดีที่สุด โดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC))

1.2) วิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS))

1.3) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window) (BMA_{OCC})

1.4) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition(MC^3)) (BMA_{MC3})

1.5) วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

2) เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์จากตัวแบบที่ได้จากวิธีต่าง ๆ

1.3 ข้อตกลงเบื้องต้น

1) รูปแบบทั่วไปของสมการถดถอยเชิงเส้นพหุคูณมีรูปแบบดังสมการ (1.1)

2) ตัวแปรอิสระแต่ละตัวถือว่าเป็นค่าคงที่

3) ความคลาดเคลื่อนสุ่มเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น σ เหมือนกันและเป็นอิสระซึ่งกันและกัน

4) ตัวประมาณค่าเวกเตอร์พารามิเตอร์สัมประสิทธิ์ความถดถอย β ของตัวแบบความถดถอยเชิงเส้นพหุคูณ คือ

4.1) ตัวประมาณค่ากำลังสองน้อยสุด (Least Square Estimator : $\hat{\underset{\sim}{\beta}}$) คือ

$$\hat{\underset{\sim}{\beta}} = (x'x)^{-1}x'\underset{\sim}{y}$$

4.2) ตัวประมาณค่าความควรจะเป็นสูงสุด (Maximum Likelihood Estimator : $\hat{\underset{\sim}{\beta}}_{ML}$) คือ

$$\hat{\underset{\sim}{\beta}}_{ML} = (x'x)^{-1}x'\underset{\sim}{y}$$

1.4 สมมติฐานของการวิจัย

วิธีการคัดเลือกสมการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์ ในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณนั้น วิธีการเฉลี่ยของเบส์จะให้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากกว่าวิธีอื่น ๆ ทั้งนี้เพราะค่าพยากรณ์ที่ได้จากวิธีนี้มาจากตัวแบบที่เป็นไปได้เฉลี่ยกันภายใต้หลักการของเบส์ ไม่ใช่เป็นเพียงรูปแบบเดียวที่เลือกมาจากทุกรูปแบบที่เป็นไปได้ดังเช่นวิธีอื่น ๆ ส่วนวิธีการภายใต้แนวทางของเบส์จะให้ค่าพยากรณ์ที่มีความถูกต้องและแม่นยำมากกว่าวิธีการพื้นฐาน เนื่องจากวิธีการภายใต้แนวทางของเบส์มีการนำข้อมูลหรือความรู้ในอดีตของพารามิเตอร์มาใช้

1.5 ขอบเขตของการวิจัย

1) ตัวแบบของการถดถอยเชิงเส้นพหุคูณที่สนใจศึกษามีรูปแบบดังสมการ (1.1) ซึ่งรูปแบบการถดถอยนอกจากจะเป็นเชิงเส้นในพารามิเตอร์แล้ว ยังเป็นเชิงเส้นในตัวแปรอิสระด้วย

2) ตัวแบบการถดถอยในการวิจัยครั้งนี้เป็นตัวแบบติดกลุ่ม (nested models) [1]

---

[1] ตัวแบบติดกลุ่ม (nested models) หมายถึง 2 ตัวแบบจะติดกัน ถ้าในแต่ละพจน์ของตัวแบบแรกเป็นส่วนหนึ่งของตัวแบบที่สอง ซึ่งตัวแบบที่สองจะมีพจน์มากกว่าตัวแบบแรกอย่างน้อย 1 เทอม ตัวแบบที่สองที่มีความซับซ้อนมากกว่าตัวแบบแรกจะเรียกว่าตัวแบบสมบูรณ์ (complete model) และตัวแบบแรกที่เป็นตัวแบบอย่างง่ายของตัวแบบที่สอง เรียกว่าตัวแบบลดรูป (reduced model)

3) การวิจัยครั้งนี้กำหนดให้ $\underset{\sim}{\beta}' = (1, 1, \ldots, 1)_{1xp}$ ในประชากรทุกรูปแบบที่ศึกษา โดยที่ p เป็นจำนวนพารามิเตอร์ในตัวแบบการถดถอย

4) ขนาดตัวอย่าง (n) ที่ศึกษาคือ 25 50 75 และ 100

5) จำนวนตัวแปรอิสระที่ศึกษาคือ 3 5 10 12 และ 15 ตัวแปร โดยตัวแปรอิสระแต่ละตัวแปรสร้างจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1

6) ระดับนัยสำคัญ (α) คือ 0.01 และ 0.05

7) ในการวิจัยครั้งนี้จะศึกษาเมื่อความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5 10 20 และ 25

8) ค่าคงที่สำหรับวิธี BVS และวิธี BMA_{MC3} มี 4 ระดับ คือ (1,5) (1,10) (10,100) และ (10,500)

8) ข้อมูลที่ใช้ในการศึกษาวิจัยครั้งนี้ใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โล กระทำซ้ำ 500 รอบในแต่ละสถานการณ์

1.6 ขั้นตอนในการศึกษาวิจัย

ขั้นตอนในการศึกษาวิจัยมีดังนี้ คือ

1) กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง จำนวนตัวแปรอิสระ ระดับนัยสำคัญ และค่าคงที่สำหรับวิธี BVS และวิธี BMA_{MC3}

2) สร้างข้อมูลตัวแปรตามจากตัวแปรอิสระ และความคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา โดยให้ตัวแปรตามมีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ

3) สร้างตัวแบบโดยวิธีทั้ง 5 วิธี คือ

3.1) วิธีการคัดเลือกตัวแบบที่ดีที่สุด โดยใช้เกณฑ์ข้อสนเทศของเบส์ (Bayes Information Criterion Method (BIC))

3.2) วิธีการคัดเลือกตัวแปรของเบส์ (Bayesian Variable Selection Method (BVS))

3.3) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยใช้การค้นหาปริภูมิตัวแบบด้วยวิธีออกส์แคม วินโดว์ (Occam 's Window) (BMA_{occ})

3.4) วิธีการเฉลี่ยตัวแบบของเบส์ (Bayesian Model Averaging Method) โดยการหาองค์ประกอบของตัวแบบด้วยเทคนิคมอนติคาร์โลโดยใช้ลูกโซ่มาร์คอฟ (Markov Chain Monte Carlo Model Composition(MC^3)) (BMA_{MC3})

3.5) วิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method (SR))

4) หาค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละตัวแบบ

5) สรุปผลในรูปของตารางและรูปภาพ

1.7 ประโยชน์ที่คาดว่าจะได้รับ

1) เป็นแนวทางในการคัดเลือกตัวแบบที่เหมาะสมเพื่อใช้ในการพยากรณ์ สำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ

2) เป็นแนวทางในการศึกษาวิธีการคัดเลือกตัวแบบการถดถอยที่ดีที่สุดวิธีอื่น ๆ ภายใต้แนวทางของเบส์ สำหรับการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณต่อไป

1.8 เกณฑ์ที่ใช้ในการตัดสินใจ

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการคัดเลือกตัวแบบการถดถอยวิธีใดจะมีประสิทธิภาพมากที่สุดพิจารณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error (AMSE)) และเกณฑ์ที่ใช้ประกอบการตัดสินใจจะใช้เกณฑ์ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Square Error (RDAMSE)) ซึ่งมีสูตรดังนี้

$$MSE_i = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n - p}$$

$$AMSE = \frac{\sum_{j=1}^{500} MSE_i}{500}$$

เมื่อ y_i เป็นค่าสังเกตที่ i

$\hat{y}_i$ เป็นค่าพยากรณ์ที่ i

p เป็นจำนวนพารามิเตอร์ในตัวแบบการถดถอย

n เป็นขนาดตัวอย่าง

MSE_j เป็นค่าคลาดเคลื่อนกำลังสองเฉลี่ยของการทำซ้ำรอบที่ j

และ $AMSE$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำ 500 รอบ

$$RDAMSE_i = \frac{\left(AMSE_i - AMSE_{min}\right)}{AMSE_{min}} \times 100\%$$

เมื่อ $AMSE_i$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากวิธีที่ i

และ $AMSE_{min}$ เป็นค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยที่มีค่าต่ำสุดจากทุกวิธี