การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นเมื่อตัวอย่างมีขนาดเล็ก

นางสาวภัทรสุดา สุดแสน

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาสถิติศาสตรมหาบัณฑิต

สาขาวิชาสถิติ ภาควิชาสถิติ

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2548

ISBN 974-14-2260-1

A COMPARISION OF MODEL SELECTION CRITERIA ON LINEAR REGRESSION WITH SMALL SAMPLE SIZE

Miss Pattarasuda Sudsaen

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science Program in Statistics
Department of Statistics
Faculty of Commerce and Accountancy
Chulalongkorn University
Academic Year 2005
ISBN 974-14-2260-1

หัวข้อวิทยานิพนธ์ การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นเมื่อตัวอย่างมีขนาดเล็ก
โดย นางสาวภัทรสุดา สุดแสน
สาขาวิชา สถิติ
อาจารย์ที่ปรึกษา รองศาตราจารย์ ดร.ธีระพร วีระถาวร

คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญามหาบัณฑิต

.... คณบดีคณะพาณิชยศาสตร์และการบัญชี
(ผู้ช่วยศาสตราจารย์ ดร.ดนุชา คุณพนิชกิจ)

คณะกรรมการสอบวิทยานิพนธ์

.... ประธานกรรมการ
(รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา)

.... อาจารย์ที่ปรึกษา
(รองศาตราจารย์ ดร.ธีระพร วีระถาวร)

.... กรรมการ
(รองศาสตราจารย์ ผกาวดี ศิริรังษี)

ภัทรสุดา สุดแสน : การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นเมื่อตัวอย่างมีขนาดเล็ก (A COMPARISION OF MODEL SELECTION CRITERIA ON LINEAR REGRESSION WITH SMALL SAMPLE SIZE) อ. ที่ปรึกษา : รศ.ดร.ธีระพร วีระถาวร , 118 หน้า. ISBN 974-14-2260-1

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นเมื่อตัวอย่างมีขนาดเล็กโดยจะเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอย 4 เกณฑ์ ได้แก่ เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้(Corrected Akaike's Information Criterion (AIC_C)) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (Corrected Schwarz's Information Criterion (SIC_C)) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้(Corrected Hannan and Quinn's Information Criterion (HQ_C)) และเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลล์แบล็คที่ปรับแก้(Corrected Kullback's Information Criterion (KIC_C)) เกณฑ์การตัดสินใจคือค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Squares Error (AMSE)) และใช้อัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Squares Error (RDAMSE)) เพื่อเปรียบเทียบประสิทธิภาพของเกณฑ์การคัดเลือกตัวแบบทั้ง 4 เกณฑ์ การแจกแจงของความคลาดเคลื่อนสุ่มที่ศึกษาคือการแจกแจงแบบปกติที่มีค่าเฉลี่ย 0 และความแปรปรวนเป็น 1, 2, 4, 8, และ 16 ขนาดตัวอย่างที่ใช้คือ 12, 15, 18, 21, 24, 27 และ 30 และจำนวนตัวแปรอิสระในตัวแบบการถดถอยคือ 3, 5 และ 7 ตัวแปร ข้อมูลที่ใช้ในการวิจัยใช้โปรแกรมคอมพิวเตอร์จำลองด้วยเทคนิคมอนติคาร์โลกระทำซ้ำ 1,000 ครั้ง ในแต่ละสถานการณ์ ซึ่งผลการวิจัยสรุปได้ดังนี้

ปัจจัยที่มีผลต่อค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (**AMSE**) ของทุกเกณฑ์การคัดเลือกตัวแบบได้แก่ จำนวนตัวแปรอิสระและความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) และขนาดตัวอย่าง ตามลำดับ โดยที่ **AMSE แปรผันตาม**จำนวนตัวแปรอิสระและความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) แต่ **AMSE แปรผกผัน**กับขนาดตัวอย่าง ปัจจัยดังกล่าวข้างต้นส่งผลกระทบกับเกณฑ์การคัดเลือกตัวแบบดังกรณีต่อไปนี้

1. **กรณีขนาดตัวอย่างเท่ากับ 12 ถึง 15**

เกณฑ์การคัดเลือกตัวแบบที่ดีที่สุดคือ เกณฑ์ HQ_c รองลงมาคือ เกณฑ์ AIC_c,SIC_c และ KIC_c ตามลำดับ สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่มและจำนวนตัวแปรอิสระ

2. **กรณีขนาดตัวอย่างเท่ากับ 18**

เกณฑ์การคัดเลือกตัวแบบที่ดีที่สุดมีสองเกณฑ์คือ เกณฑ์ HQ_c และเกณฑ์ AIC_c และรองลงมาคือเกณฑ์ KIC_c และ SIC_c โดยที่เกณฑ์ KIC_c มีค่า AMSE ต่ำกว่า SIC_c เพียงเล็กน้อย สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่มและจำนวนตัวแปรอิสระ

3. **กรณีขนาดตัวอย่างเท่ากับ 21 ถึง 30**

เกณฑ์การคัดเลือกตัวแบบที่ให้ค่า AMSE น้อยที่สุดคือ เกณฑ์ AIC_c รองลงมาคือ เกณฑ์ HQ_c, KIC_c และ SIC_c ตามลำดับ สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่มและจำนวนตัวแปรอิสระ

เมื่อพิจารณาค่า RDAMSE พบว่า ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบจะแบ่งออกเป็นสองกลุ่มอย่างเห็นได้ชัด คือ กลุ่มที่ 1 ได้แก่ เกณฑ์ HQ_c และเกณฑ์ AIC_c มีประสิทธิภาพใกล้เคียงกันและดีที่สุดสำหรับทุกขนาดตัวอย่างและความแปรปรวนของความคลาดเคลื่อนสุ่ม ส่วนกลุ่มที่ 2 ได้แก่ เกณฑ์ KIC_c และ SIC_c ซึ่งมีประสิทธิภาพใกล้เคียงกัน**แต่แตกต่างและต่ำกว่ากลุ่มแรก**สำหรับทุกขนาดตัวอย่างและความแปรปรวนของความคลาดเคลื่อนสุ่ม

ภาควิชา สถิติ ลายมือชื่อนิสิต.....................

สาขาวิชา สถิติ ลายมือชื่ออาจารย์ที่ปรึกษา.....................

ปีการศึกษา 2548

4682360026 : MAJOR STATISTICS

KEY WORD : AIC_C / HQ_C / KIC_C / LINEAR REGRESSION / MODEL SELECTION / SIC_C / SMALL SAMPLE SIZE

PATTARASUDA SUDSAEN : A COMPARISION OF MODEL SELECTION CRITERIA ON LINEAR REGRESSION WITH SMALL SAMPLE SIZE.THESIS ADVISOR : ASSOC.PROF.THEERAPORN VERATHAWORN, Ph.D. 118 pp. ISBN 974-14-2260-1

The purpose of this research is to compare model selection criteria for linear regression with small sample size. The four model selection criteria in this comparison are composed of Corrected Akaike's Information Criterion (AIC_C), Corrected Schwarz's Information Criterion (SIC_C), Corrected Hannan and Quinn's Information Criterion (HQ_C) and Corrected Kullback's's Information Criterion (KIC_C). The criterion of comparison is the Average of Mean Squares Error (AMSE) using Ratio of Different Average Mean Squares Error (RDAMSE) to compare the efficiency of these model selection criteria. The distribution of random errors are normal distribution with zero mean and variances are 1, 2, 4, 8 and 16, respectively. The sample sizes used in this study are 12, 15, 18, 21, 24, 27 and 30, respectively. The number of independent variables in regression model are 3, 5, and 6, respectively. The data for this experiment are generated through the Monte Carlo simulation technique, repeating 1,000 times for each case. The results of this research are as follow :

The factors that affected AMSE of all model selection criteria are the number of independent variables, the variance of random errors and sample size, respectively .The AMSE vary with, most to least ,respectively,the number of independent variables and the variance of random errors but AMSE is converse to sample size.These factors affected model selection criteria as follow.

1. In case of sample size is 12 to 15

For all the number of independent variables and all variance of random errors, the HQ_C is the best, the AIC_C , the SIC_C and the KIC_C , respectively.

2. In case of sample size is 18

For all the number of independent variables and all variance of random errors, the HQ_C and the AIC_C are the best, the KIC_C and the SIC_C, respectively. The KIC_C give AMSE slightly lower than the SIC_C.

3. In case of sample size is 21 to 30

For all the number of independent variables and all variance of random errors, the AIC_C is the best, the HQ_C , the SIC_C and the KIC_C , respectively.

Furthermore, the four model selection criteria efficiency is separated into two groups by RDAMSE. The first group is composed of the HQ_C and the AIC_C .Their RDAMSE are close and they are the best for all cases. The second group is composed of the SIC_C and the KIC_C. Their RDAMSE are close but their RDAMSE are clearly different from the first group for all cases.

Department Statistics .... Student's signature Pattar Sudsa

Field of study Statistics .... Advisor's signature Theeraporn Verath

Academic year 2005

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยความช่วยเหลืออย่างดียิ่งจาก รองศาสตราจารย์ ดร.ธีระพร วีระถาวร อาจารย์ที่ปรึกษาวิทยานิพนธ์ ที่กรุณาให้คำแนะนำ ปรึกษา รวมทั้งตรวจสอบและแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดี ผู้วิจัยใคร่ขอกราบขอบพระคุณเป็นอย่างสูง

ผู้เขียนขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ซึ่งประกอบด้วย รองศาสตราจารย์ ดร.สุพล ดุรงค์วัฒนา และรองศาสตราจารย์ ผกาวดี ศิริรังษี ที่ได้กรุณาช่วยตรวจและแก้ไขวิทยานิพนธ์ฉบับนี้ให้สมบูรณ์ยิ่งขึ้น

ผู้เขียนขอขอบคุณเจ้าหน้าที่ห้องสมุด คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ที่ได้อำนวยความสะดวกในด้านตำราต่างๆ เพื่อศึกษา ค้นคว้า ประกอบการทำวิทยานิพนธ์ และขอขอบคุณเพื่อน ๆ ทุกคน ที่ให้ความช่วยเหลือ และเป็นกำลังใจให้ผู้เขียน โดยเฉพาะอย่างยิ่ง นายสมลักษณ์ ศิริชื่นวิจิตร นิสิตปริญญาโท ภาควิชาสถิติ ที่ให้คำแนะนำ ปรึกษาในด้านการเขียนโปรแกรม

สุดท้ายนี้ผู้เขียนขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัวของผู้เขียน ที่ส่งเสริมสนับสนุนด้านการศึกษาและคอยเป็นกำลังใจให้ผู้เขียนมาโดยตลอด

สารบัญ

หน้า

สารบัญตาราง

ตารางที่ หน้า

สารบัญรูป

รูปที่ หน้า

บทที่ 1

บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันการวิจัยในด้านต่าง ๆ ไม่ว่าจะเป็นวิทยาศาสตร์ สังคมศาสตร์ เศรษฐศาสตร์ เป็นต้น จำเป็นต้องใช้วิธีการทางสถิติเข้ามาร่วมในการค้นคว้าหาคำตอบและพยากรณ์ ซึ่งวิธีการที่นิยมกันมากคือการวิเคราะห์ความถดถอย(regression analysis) ซึ่งเป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตามทั้งในรูปแบบที่เป็นเชิงเส้นและไม่เป็นเชิงเส้น รูปแบบที่นิยมกันแพร่หลายคือการวิเคราะห์ความถดถอยเชิงเส้น (linear regression analysis)

การนำตัวแบบการถดถอยเชิงเส้นเพื่อใช้การพยากรณ์ให้มีความแม่นยำและถูกต้อง จำเป็นต้องเลือกตัวแบบที่ให้ค่าความคลาดเคลื่อนในการพยากรณ์น้อยซึ่งการที่จะได้ตัวแบบที่เหมาะสมดังกล่าวก็น่าจะมาจากการใช้วิธีการคัดเลือกตัวแบบที่เหมาะสมกับข้อมูลด้วย ซึ่งในปัจจุบันมีเกณฑ์การคัดเลือกตัวแบบอยู่หลายเกณฑ์ด้วยกันและที่นิยมใช้กันอย่างแพร่หลาย เช่น เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะ (Akaike 's Information Criterion (AIC)) หรือเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของเบส์ (Bayesian Information Criterion (BIC)) เกณฑ์การคัดเลือกเหล่านี้จะมีคุณสมบัติที่ดีเมื่อตัวอย่างมีขนาดใหญ่เรียกว่า Asymptotically Efficient กล่าวคือ เป็นเกณฑ์การคัดเลือกซึ่งจะเลือกตัวแบบที่มีค่าความคลาดเคลื่อนต่ำที่สุด หรือ ความน่าจะเป็นในการจะเลือกตัวแบบที่ถูกต้องเข้าใกล้หนึ่งเมื่อขนาดตัวอย่างมีขนาดใหญ่

แต่ในทางกลับกันในการเก็บรวบรวมข้อมูลหรือการวิจัยในบางด้าน เช่น ในการทดลองทางวิทยาศาสตร์อาจมีตัวอย่างขนาดเล็กหรือสัดส่วนระหว่างจำนวนตัวแปรอิสระและจำนวนตัวอย่างมีค่าสูง เมื่อนำข้อมูลเหล่านี้ไปวิเคราะห์ความถดถอยและคัดเลือกตัวแบบโดยใช้เกณฑ์การคัดเลือกที่นิยมใช้กันทั่วไป ซึ่งเหมาะสมสำหรับตัวอย่างขนาดใหญ่มากกว่าจะส่งผลให้มีโอกาสที่จะได้ตัวแบบที่ไม่เหมาะสมสูงขึ้น กล่าวคืออาจได้ตัวแบบที่ประกอบด้วยตัวแปรอิสระมากกว่าตัวแบบจริง เรียกว่า เกิดปัญหา overfitting หรืออาจได้ตัวแบบที่มีตัวแปรอิสระน้อยกว่าตัวแบบจริง เรียกว่า เกิดปัญหา underfitting

ปัญหาของ overfitting คือค่าพยากรณ์ที่ได้จากตัวแบบนั้นมีความแปรปรวนสูง เนื่องมาจากตัวแปรอิสระที่เกินมา (Extraneous Variables) ส่วนปัญหาของ underfitting คือให้ค่าพยากรณ์ที่ไม่แม่นยำเนื่องจากขาดรายละเอียดของตัวแปรอิสระที่มีความสำคัญที่ขาดหายไป

เนื่องจากปัญหาที่เกิดขึ้นในเกณฑ์การคัดเลือกตัวแบบสำหรับตัวอย่างขนาดเล็ก จึงได้มีผู้คิดค้นและปรับปรุงเกณฑ์การคัดเลือกตัวแบบสำหรับตัวอย่างขนาดเล็กขึ้นมา ดังต่อไปนี้

ในปี ค.ศ.1989 เฮอร์วิช และไซ (Hurvich and Tsai) ได้ศึกษาประสิทธิภาพของ AIC เมื่อมีขนาดตัวอย่างเล็กผลปรากฏว่ามีประสิทธิภาพต่ำเมื่อขนาดตัวอย่างลดลงเพราะ AIC มีความเอนเอียงเข้าสู่ด้านลบจึงทำให้มีแนวโน้มที่จะคัดเลือกตัวแบบที่มีตัวแปรอิสระมากกว่าตัวแบบจริง หรือเกิด overfittingนั่นเอง เขาจึงได้เสนอเกณฑ์ข้อสนเทศที่ได้รับการปรับแก้ความเอนเอียงนี้ซึ่งเรียกว่า Corrected for Akaike Information Criterion (AIC_C) เพื่อใช้เป็นเกณฑ์ในการคัดเลือกตัวแบบเมื่อ ขนาดตัวอย่างเล็กและนอกจากนั้นยังพบว่าเมื่อขนาดตัวอย่างใหญ่ AIC_C ก็มีประสิทธิภาพเท่าเทียมกับ AIC

ในปี ค.ศ.1998 แมคไควร์รีและไซ (Mcquarrie and Tsi) ได้เสนอเกณฑ์การคัดเลือกตัวแบบที่ปรับแก้ความเอนเอียงสำหรับตัวอย่างขนาดเล็กของเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (Hannan and Quinn) ซึ่งเรียกว่า Corrected for Hannan and Quinn Information Criterion (HQ_C) โดยมีแนวคิดในการสร้างเกณฑ์การคัดเลือกตัวแบบใหม่โดยอาศัยความสัมพันธ์ระหว่าง AIC กับ AIC_C

ในปี ค.ศ. 1999 แมคไควร์รี (McQuirrie) ได้ศึกษาประสิทธิภาพของ SIC เมื่อขนาดตัวอย่างเล็กพบว่ามีลักษณะเป็นเช่นเดียวกันกับ AIC เขาจึงได้เสนอเกณฑ์ข้อสนเทศที่ได้รับการปรับแก้ความเอนเอียงนี้ซึ่งเรียกว่า Corrected for Schwarz Information Criterion (SIC_C) เพื่อใช้เป็นเกณฑ์ในการคัดเลือกตัวแบบเมื่อขนาดตัวอย่างเล็กโดยมีแนวคิดในการสร้างเกณฑ์การคัดเลือกตัวแบบใหม่โดยอาศัยความสัมพันธ์ระหว่าง AIC กับ AIC_C

ในปี ค.ศ. 1999 โจเซฟ.อี.คาวานอท (Joseph E. Cavanaugh) ได้เสนอเกณฑ์การคัดเลือกตัวแบบโดยอาศัยแนวคิดของ Kullback's Symmetric Divergence คือ เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็ค(KIC) ซึ่งมีประสิทธิภาพดีเมื่อตัวอย่างขนาดใหญ่ แต่ยังมีข้อบกพร่องในกรณีขนาดตัวอย่างเล็กเช่นเดียวกับเกณฑ์การคัดเลือกวิธีอื่นๆ เช่น AIC และ SIC

ดังนั้นในปี ค.ศ.2004 เขาจึงได้เสนอเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์-แบล็คที่ปรับแก้(Corrected for Kullback Information Criterion (KIC_C)) เพื่อใช้เป็นเกณฑ์ในการคัดเลือกตัวแบบเมื่อขนาดตัวอย่างเล็ก

ในการวิจัยนี้ผู้วิจัยจึงสนใจศึกษา เกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นสำหรับตัวอย่างขนาดเล็ก โดยเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นที่สนใจศึกษา ได้แก่

1) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้
   (Corrected Akaike's Information Criterion (AIC_C))
2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้

(Corrected Schwarz's Information Criterion (SIC_C))

3) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของเฮนแนนและควินน์ที่ปรับแก้ (Corrected Hannan and Quinn's Information Criterion (HQ_C))

4) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (Corrected Kullback's's Information Criterion (KIC_C))

1.2 วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาเกณฑ์การคัดเลือกตัวแบบในการวิเคราะห์การถดถอยของตัวแบบเชิงเส้นสำหรับตัวอย่างขนาดเล็ก โดยมีเกณฑ์ที่ใช้ในการศึกษาคือ AIC_C , SIC_C , HQ_C และ KIC_C

2. เพื่อเปรียบเทียบความถูกต้องในการพยากรณ์ของการคัดเลือกตัวแบบจากเกณฑ์การคัดเลือกตัวแบบทั้ง

4 เกณฑ์การคัดเลือก

1.3 สมมติฐานของการวิจัย

เกณฑ์การคัดเลือกตัวแบบในการวิเคราะห์ความถดถอยของตัวแบบเชิงเส้นสำหรับตัวอย่างขนาดเล็ก มีประสิทธิภาพมากที่สุดน่าจะเป็นเกณฑ์ KIC_C เพราะเป็นเกณฑ์การคัดเลือกตัวแบบที่คำนึงถึงการหาระยะทางระหว่างตัวแบบที่แท้จริง(true model) กับตัวแบบที่นำมาพิจารณา (candidate model) แบบสมมาตรซึ่งน่าจะทำให้ความคลาดเคลื่อนในการพยากรณ์ต่ำกว่าวิธีอื่นๆ ที่นำมาเปรียบเทียบกัน

1.4 ข้อตกลงเบื้องต้น

1. รูปแบบทั่วไปของตัวแบบการถดถอยเชิงเส้นแบบพหุคูณแบบติดกลุ่มอยู่ในรูปเชิงเส้นของพารามิเตอร์และตัวแปรอิสระมีรูปแบบดังนี้

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

เมื่อ $\underset{\sim}{y}$ คือเวกเตอร์ของตัวแปรตามขนาด $n\times 1$

X คือเมทริกซ์ของตัวแปรอิสระขนาด $n\times(p+1)$

$\underset{\sim}{\beta}$ คือเวกเตอร์ของพารามิเตอร์ของสัมประสิทธิ์การถดถอยขนาด $(p+1)\times 1$

$\underset{\sim}{\varepsilon}$ คือเวกเตอร์ของความคลาดเคลื่อนขนาด $n\times 1$ โดยที่ $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0},\sigma^2 I_n)$

I_n คือเมทริกซ์เอกลักษณ์ขนาด $n\times n$

n คือขนาดตัวอย่าง

และ p คือจำนวนตัวแปรอิสระในตัวแบบ

2. ตัวประมาณค่าพารามิเตอร์สัมประสิทธิ์การถดถอยของ $\underset{\sim}{\beta}$ ของตัวแบบการถดถอยเชิงเส้นข้างต้นคือตัวประมาณค่าควรจะเป็นสูงสุด (Maximum Likelihood Estimator)

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

3. ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่

4. ความคลาดเคลื่อนเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น σ เหมือนกันและเป็นอิสระซึ่งกันและกัน

1.5 คำจำกัดความต่าง ๆ ที่ใช้ในการวิจัย

<u>ตัวอย่างขนาดเล็ก</u>(Small Sample Size (n)) หมายถึง ขนาดตัวอย่างที่นำมาใช้ในการศึกษามีจำนวนไม่เกิน 30

<u>ตัวแปรตาม</u> (Dependent Variable (y)) หมายถึง ตัวแปรที่เราสนใจที่เราต้องการทำการศึกษา

<u>ตัวแปรอิสระ</u> (Independent Variable (x)) หมายถึง ตัวแปรที่เราคิดว่ามีความสัมพันธ์กับตัวแปรตามที่เราต้องการศึกษา หรือมีผลกระทบต่อสิ่งที่เราต้องการศึกษา

<u>ความคลาดเคลื่อนยกกำลังสองเฉลี่ย</u> (MSE) หมายถึง ความคลาดเคลื่อนที่เกิดขึ้นจากการใช้สมการในการพยากรณ์สำหรับประมาณค่าตัวแปรตามโดยใช้ฟังก์ชันความสูญเสียแบบความผิดพลาดยกกำลังสอง (square error loss)

<u>ตัวแบบติดกลุ่ม</u> (Nested Models) หมายถึง ตัวแบบ 2 ตัวแบบจะติดกันถ้าในแต่ละพจน์ของตัวแบบแรกเป็นส่วนหนึ่งของตัวแบบที่สอง ซึ่งตัวแบบที่สองจะมีพจน์มากกว่าตัวแบบแรกอย่างน้อย 1 พจน์ ตัวแบบที่สองที่มีความซับซ้อนมากกว่าตัวแบบแรกจะเรียกว่า ตัวแบบที่สมบูรณ์ (completed model) และตัวแบบแรกที่เป็นตัวแบบที่ง่ายกว่าของตัวแบบที่สองเรียกว่า ตัวแบบลดรูป(reduced model) เช่น

ตัวแบบเต็มรูป (full model) ของสมการการถดถอยเมื่อมีตัวแปรอิสระเริ่มต้น 3 ตัว คือ

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i ; i = 1, 2, ..., n$$

จะมีตัวแบบลดรูปที่ติดกลุ่มกับตัวแบบข้างต้นอยู่ทั้งหมด 7 ตัวแบบ ได้แก่

$$y = \beta_0 + \varepsilon_i \qquad ; i = 1, 2, ..., n$$

$$y = \beta_0 + \beta_1 x_1 + \varepsilon_i \qquad ; i = 1, 2, ..., n$$

$$y = \beta_0 + \beta_1 x_2 + \varepsilon_i \qquad ; i = 1, 2, ..., n$$

$$y = \beta_0 + \beta_1 x_3 + \varepsilon_i \qquad ; i = 1, 2, ..., n$$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i \qquad ; i = 1, 2, ..., n$$

$$y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon_i \qquad ; i = 1, 2, ..., n$$

$$y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon_i \qquad ; i = 1, 2, ..., n$$

1.6 ขอบเขตของการวิจัย

1. ในการวิจัยครั้งนี้ศึกษาการถดถอยเชิงเส้นแบบติดกลุ่มซึ่งตัวแบบนี้จะอยู่ในรูปเชิงเส้นของพารามิเตอร์และตัวแปรอิสระโดยมีรูปแบบทั่วไปของตัวแบบที่ต้องการศึกษาเป็นไปตามข้อตกลงเบื้องต้น

2. ในการวิจัยครั้งนี้ศึกษาความคลาดเคลื่อนมีค่าเฉลี่ย (μ) เท่ากับ 0 โดยกำหนดให้ค่าความแปรปรวน(σ^2) เท่ากับ 1 , 2 , 4 , 8 และ 16

3. การวิจัยครั้งนี้กำหนดให้ $\underset{\sim}{\beta}' = (1, 1, 1, \ldots, 1)_{1\times(p+1)}$

4. จำนวนตัวแปรอิสระเริ่มต้นที่ใช้ศึกษาเท่ากับ 3 , 5 และ 7

5. ขนาดตัวอย่างที่ศึกษา $n =$ 12 , 15 , 18 , 21, 24, 27 และ 30

6. การวิจัยในครั้งนี้เป็นการวิจัยเชิงทดลอง โดยทำการสร้างแบบจำลองข้อมูลจากการทำซ้ำจำนวน 1,000 รอบ ด้วยวิธีการจำลองแบบมอนติคาร์โล (Monte Carlo Simulation)และเขียนโดยใช้โปรแกรม Borland Delphi 7

1.7 ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนในการดำเนินการการวิจัยมีดังนี้

1. กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง และจำนวนตัวแปรอิสระที่ใช้ในการวิจัย

2. สร้างข้อมูลตัวแปรอิสระ และค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่กำหนด

3. สร้างข้อมูลตัวแปรตามจากตัวแปรอิสระและค่าคลาดเคลื่อน โดยให้ตัวแปรตามมีลักษณะความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ

4. กำหนดตัวแบบเริ่มต้นซึ่งเป็นตัวแบบที่เต็มรูป (full model) ในการคัดเลือกตัวแบบที่เหมาะสม

5. กำหนดตัวแบบที่เป็นไปได้ทั้งหมด ที่ติดกลุ่มของตัวแบบเริ่มต้นเพื่อใช้ในการคัดเลือกตัวแบบตัวแบบที่เหมาะสม

6. ทำการคัดเลือกตัวแบบที่เหมาะสมจากทั้ง 4 เกณฑ์ คือ

   6.1 เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (Corrected Akaike's Information Criterion (AIC_C))

6.2 เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (Corrected Schwarz's Information Criterion (SIC_C))

6.3 เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ (Corrected Hannan and Quinn's Information Criterion (HQ_C))

6.4 เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (Corrected Kullback's's Information Criterion (KIC_C))

7. คำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำจำนวน 1,000 รอบ ของตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบการถดถอยทั้ง 4 เกณฑ์

8. ทำการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบการถดถอยทั้ง 4 เกณฑ์ โดยสรุปผลในรูปของตารางและรูปภาพ

1.8 เกณฑ์การตัดสินใจ

เกณฑ์การคัดเลือกตัวแบบการถดถอยว่าวิธีใดจะมีประสิทธิภาพมากที่สุด จะพิจาณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(Average of Mean Squares Error (AMSE)) และเกณฑ์ที่ใช้ในการประกอบการตัดสินใจเพื่อดูประสิทธิภาพจะของเกณฑ์ต่างๆ จะใช้ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Squares Error (RDAMSE)) ซึ่งมีสูตรดังนี้

$$AMSE = \frac{\sum_{k=1}^{1,000} MSE_k^{\ *}}{1,000}$$

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

โดยที่ y_i แทนค่าสังเกตที่ i

$\hat{y}_i$ แทนค่าพยากรณ์ที่ i

p แทนจำนวนตัวแปรอิสระในตัวแบบ

n แทนขนาดตัวอย่าง

[1] $MSE_k^{\ *}$ แทนค่าคลาดเคลื่อนกำลังสองเฉลี่ยในการวิเคราะห์ความแปรปรวน(ANOVA) ของการทำซ้ำรอบที่ k

และ $AMSE$ แทนค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำ 1,000 รอบ

---

[1] $MSE_k^{\ *}$ เป็นตัวประมาณที่ไม่เอนเอียงของ σ^2 และเป็นค่าที่เน้นความสำคัญของการพยากรณ์ของตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ p ตัว ดังนั้นเมื่อต้องการเปรียบเทียบระหว่างตัวแบบการถดถอยที่สงสัย ตัวแบบที่ให้ค่า $MSE_k^{\ *}$ น้อยกว่าย่อมจะให้ค่าพยากรณ์ที่ดีกว่า

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

เมื่อ $AMSE_i$ แทนค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ยจากวิธีที่ i , $i=1,2,3,4$
และ $AMSE_{\min}$ แทนค่าเฉลี่ยความคลาดเคลื่อนกำลังสองเฉลี่ยมีค่าต่ำสุดจาก 4 วิธี

จะพิจารณาว่าเกณฑ์การคัดเลือกตัวแบบใดให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุดจะเป็นเกณฑ์การคัดเลือกตัวแบบที่ดีที่สุด

1.9 ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

เพื่อเป็นแนวทางในการศึกษาและคัดเลือกตัวแบบการถดถอยเชิงเส้นในกรณีที่ขนาดตัวอย่างเล็กเพื่อใช้ในการพยากรณ์

บทที่ 2

ทฤษฎีและวิธีการทางสถิติที่เกี่ยวข้อง

การศึกษาเกณฑ์การคัดเลือกตัวแบบการถดถอยเป็นการพิจารณาคัดเลือกตัวแบบที่ให้ค่าพยากรณ์ที่ถูกต้องและแม่นยำมากที่สุด ซึ่งปัจจุบันมีแนวคิดเบื้องต้นในการคัดเลือกตัวแบบการถดถอย 2 แนวคิด คือ

1. <u>แนวคิดของการทดสอบสมมุติฐานเพื่อทำการคัดเลือกตัวแปรอิสระในการสร้างตัวแบบ</u> เช่น Forward Selection, Backward Elimination และ Stepwise Regression เป็นต้น ซึ่งแนวคิดนี้ต้องมีการกำหนดระดับนัยสำคัญในการทดสอบทำให้เป็นข้อเสียเปรียบเนื่องจากตัวแบบที่ได้อาจแตกต่างกันเมื่อระดับนัยสำคัญแตกต่างกัน

2. <u>แนวคิดของเกณฑ์ข้อสนเทศ(Information Criterion (IC))</u> ซึ่งเป็นวิธีคัดเลือกตัวแบบที่เหมาะสมโดยพิจารณาถึงความสมดุลระหว่างความกลมกลืน (Goodness of Fit) และความเพียงพอของข้อมูลที่ได้จากตัวแบบ โดยทั่วไปเกณฑ์ข้อสนเทศ(Information Criterion) จะประกอบด้วยลอการิทึมของฟังก์ชันความควรจะเป็น(Log-Likelihood Function (L)) และ complexity penalty parameter (λ)ซึ่งอยู่ในรูปสมการดังนี้

$$IC = -2(L) + \lambda$$

เมื่อ λ เป็นค่าคงที่

เกณฑ์การคัดเลือกตัวแบบที่เป็นที่นิยมและใช้กันอย่างแพร่หลายได้แก่

2.1 <u>เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะ</u> (AIC) อาไคเคะได้นำเสนอไว้ในปี ค.ศ. 1973 ซึ่งเป็นเกณฑ์ที่ใช้ในการคัดเลือกตัวแบบ โดยใช้แนวคิดของ คูลส์แบล็ค – ไลท์เบลอร์ (Kullback – Leibler :1951) มาใช้ในการพิจารณาหาตัวแบบที่เหมาะสมเพื่อใช้ในการพยากรณ์ด้วยตัวแบบที่ให้ค่า AIC ต่ำที่สุดจะเป็นตัวแบบที่ดีที่สุด

2.2 <u>เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซ (SIC)หรือเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของเบส์(Bayesian Information Criterion (BIC))</u> ซึ่งชวาร์ซ (Schwarz) เสนอไว้ในปี ค.ศ. 1978 นั้นได้นำแนวคิดของเบส์มาประยุกต์ใช้ โดยสมมติว่าความน่าจะเป็นก่อน (prior probability) ของทุกตัวแบบเหมือนกัน ในการพิจารณาการคัดเลือกตัวแบบที่เหมาะสมจะพิจารณาจากความน่าจะเป็นภายหลัง (posterior proability) โดยพิจารณาจากค่าสูงสุดของความน่าจะเป็นภายหลัง หรือพิจารณาที่ตัวแบบที่ให้ค่า BIC ต่ำที่สุดจะเป็นตัวแบบที่ดีที่สุด

นอกจากนี้ยังมีเกณฑ์ข้อสนเทศอื่นๆ อีกมาก เช่น เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็ค(Kullback Information Criterion (KIC)) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของฟิชเชอร์ (Fisher Information Criterion (FIC)) เป็นต้น ซึ่งในการวิจัยครั้งนี้ได้ศึกษาเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นโดยข้อสนเทศที่ปรับแก้เพื่อใช้ในกรณีที่ขนาดตัวอย่างเล็กที่นำมาพิจารณามี 4 วิธี ดังนี้

1) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (Corrected Akaike's Information Criterion (AIC_C))
2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (Corrected Schwarz's Information Criterion (SIC_C))
3) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ (Corrected Hannan and Quinn's Information Criterion (HQ_C))
4) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (Corrected Kullback's's Information Criterion (KIC_C))

ซึ่งรายละเอียดของเกณฑ์การคัดเลือกตัวแบบการถดถอยแต่ละเกณฑ์มีดังนี้

2.1 การวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis)

ตัวแบบและข้อตกลงเบื้องต้นที่ใช้ในการวิจัยครั้งนี้คือตัวแบบการถดถอยเชิงเส้นพหุคูณซึ่งมีรูปแบบทั่วไปดังนี้

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

เมื่อ $\underset{\sim}{y}$ คือเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X คือเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p+1)$

$\underset{\sim}{\beta}$ คือเวกเตอร์ของพารามิเตอร์ของสัมประสิทธิ์การถดถอยขนาด $(p+1) \times 1$

$\underset{\sim}{\varepsilon}$ คือเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$ โดยที่ $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0}, \sigma^2 I_n)$

I_n คือเมทริกซ์เอกลักษณ์ขนาด $n \times n$

n คือขนาดตัวอย่าง

และ p คือจำนวนตัวแปรอิสระในตัวแบบ

<u>ข้อตกลงเบื้องต้นของการวิเคราะห์การถดถอยเชิงเส้นพหุคูณมีดังนี้</u>

1) ความคลาดเคลื่อนเป็นตัวแปรสุ่มที่มีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ^2 นั่นคือ ε_i $i.i.d$ $N(0, \sigma^2)$

2) รูปแบบการถดถอยเป็นแบบเชิงเส้นของพารามิเตอร์

3) ตัวแปรอิสระแต่ละตัวเป็นค่าคงที่และไม่มีพหุสัมพันธ์กัน

เนื่องจากตัวแบบที่ศึกษาเป็นตัวแบบการถดถอยเชิงเส้นและค่าคลาดเคลื่อนมีการแบบปกติ ดังนั้นในการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอยด้วยวิธีการความควรจะเป็นสูงสุด (maximum likelihood estimation) และวิธีการกำลังสองน้อยที่สุด(least square estimation) จะได้ตัวประมาณเดียวกัน ซึ่งจะได้ว่า

ตัวประมาณค่าควรจะเป็นสูงสุด (maximum likelihood estimator($\hat{\underset{\sim}{\beta}}_M$)) อยู่ในรูปของ

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

ซึ่งเป็นตัวประมาณที่ใช้สำหรับเกณฑ์การคัดเลือกตัวแบบทั้ง 4 เกณฑ์ของการวิจัยนี้ และจะกล่าวรายละเอียดของการประมาณด้วยวิธีความควรจะเป็นสูงสุดในหัวข้อถัดไป

2.2 **การประมาณด้วยวิธีความควรจะเป็นสูงสุด**(Maximum Likelihood Estimation)

ตัวแบบซึ่งมีรูปแบบดังสมการ

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

มีฟังก์ชันความควรจะเป็นดังนี้

$$\begin{aligned}
ML &= l\left(\underset{\sim}{\beta}_M, \sigma^2 \middle| \underset{\sim}{y}\right) \\
&= f\left(y_1 \middle| \beta_1, \sigma^2\right) \times f\left(y_2 \middle| \beta_2, \sigma^2\right) \times \ldots \times f\left(y_n \middle| \beta_p, \sigma^2\right) \\
&= (2\pi\sigma^2)^{\frac{-n}{2}} \exp\left(\frac{-\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)}{2\sigma^2}\right)
\end{aligned}$$

จะได้ว่า

$$\begin{aligned}
\ln l\left(\underset{\sim}{\beta}_M, \sigma^2 \middle| \underset{\sim}{y}\right) &= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right) \\
&= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)
\end{aligned}$$

จาก $\frac{\partial}{\partial \underset{\sim}{\beta}_M}\ln l\left(\underset{\sim}{\beta}_M, \sigma^2 \middle| \underset{\sim}{y}\right) = 0$ และ $\frac{\partial}{\partial\sigma^2}\ln l\left(\underset{\sim}{\beta}_M, \sigma^2 \middle| \underset{\sim}{y}\right) = 0$

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

และ $\hat{\sigma}^2 = \frac{1}{n}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)$

2.3 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะที่ปรับแก้ (AIC_C)

ในปี ค.ศ.1973 อาไคเคะได้นำเสนอเกณฑ์ที่ใช้ในการคัดเลือกตัวแบบ เพื่อใช้เป็นเครื่องมือในการหาตัวแบบที่ให้ค่าพยากรณ์ที่แม่นยำที่สุด นั่นคือ เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะ (AIC) ซึ่งเป็นเกณฑ์การคัดเลือกตัวแบบที่พิจารณาจากการประมาณความคลาดเลื่อนรวมกับข้อสนเทศ(information) ของค่าสังเกต โดยใช้แนวคิดจากค่าต่ำสุดของข้อสนเทศของ คูลส์แบล็ค – ไลท์เบลอร์ (Kullback – Leibler :1951) ข้อสนเทศของคูลส์แบล็ค – ไลท์เบลอร์ ใช้เป็นเครื่องมือในการวัดระยะทางระหว่างตัวแบบที่แท้จริง(true model)กับตัวแบบที่นำมาพิจารณา (candidate model) กล่าวคือ AIC เป็นตัวประมาณที่ไม่เอนเอียงโดยประมาณของค่าคาดหวังของข้อสนเทศของ คูลส์แบล็ค – ไลท์เบลอร์

กำหนดให้ ตัวแบบที่แท้จริง มีรูปแบบดังนี้

$$\underset{\sim}{y} = X_0\underset{\sim}{\beta_0} + \underset{\sim}{\varepsilon}$$

เมื่อ $\underset{\sim}{y}$ แทนเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X_0 แทนเมทริกซ์ของตัวแปรอิสระขนาด $n \times (p_0 + 1)$

$\underset{\sim}{\beta}_0$ แทนเวกเตอร์ของพารามิเตอร์ของสัมประสิทธิ์การถดถอยขนาด $(p_0 + 1) \times 1$

p_0 แทนจำนวนตัวแปรอิสระในตัวแบบที่แท้จริง

และ $\underset{\sim}{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$ โดยที่ $\underset{\sim}{\varepsilon}$ $i.i.d.$ $N_n(\underset{\sim}{0}, \sigma^2 I_n)$

ส่วนตัวแบบที่นำมาพิจารณา มีรูปแบบดังนี้

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

เมื่อ $\underset{\sim}{y}$ แทนเวกเตอร์ของตัวแปรตามขนาด $n \times 1$

X แทนเมทริกซ์ของตัวแปรอิสระที่คงที่ขนาด $n \times (p + 1)$

$\underset{\sim}{\beta}$ แทนเวกเตอร์ของพารามิเตอร์ของสัมประสิทธิ์การถดถอยขนาด $(p + 1) \times 1$

p แทนจำนวนตัวแปรอิสระในตัวแบบที่นำมาพิจารณา

และ $\underset{\sim}{\varepsilon}$ แทนเวกเตอร์ของความคลาดเคลื่อนขนาด $n \times 1$ โดยที่ $\underset{\sim}{\varepsilon}$ $i.i.d.$ $N_n(\underset{\sim}{0}, \sigma^2 I_n)$

โดยมีข้อตกลงเบื้องต้นคือ ตัวแบบที่แท้จริงต้องอยู่ในวงศ์ของตัวแบบที่นำมาพิจารณา(candidate family) หรือตัวแบบที่แท้จริงต้องติดกลุ่มกับตัวแบบที่นำมาพิจารณา เพื่อเป็นการรับประกันว่าเวกเตอร์ของพารามิเตอร์ของตัวแบบที่แท้จริงจะอยู่ในปริภูมิเวกเตอร์ ($\theta_0 \in \Theta$)

จากข้อสนเทศของ คูลส์แบล็ค – ไลท์เบลอร์ จะได้

$$d(\theta_0,\theta_p) = E_{\theta_0}\left[-2\ln f(y|\theta_p)\right]$$

เมื่อ θ_0 คือเวกเตอร์ของพารามิเตอร์ของตัวแบบแท้จริงขนาด $k_0 \times 1$ เมื่อ $k_0 = p_0 + 1$

θ_k คือเวกเตอร์ของพารามิเตอร์ของตัวแบบที่นำมาพิจารณาขนาด $k \times 1$ เมื่อ $k = p+1$

$f(y|\theta_0)$ คือฟังก์ชันความหนาแน่นของ ตัวแบบที่แท้จริง

$f(y|\theta_k)$ คือฟังก์ชันความหนาแน่นของ ตัวแบบที่นำมาพิจารณา

และ $E_{\theta_0}\{.\}$ คือค่าคาดหวังภายใต้ตัวแบบที่แท้จริง

เราสามารถประมาณ $d(\theta_0,\theta_k)$ ได้ด้วย $d(\theta_0,\hat{\theta}_k)$ จะได้ว่า

$$d(\theta_0,\hat{\theta}_k) = E_{\theta_0}\left[-2\ln f(y|\hat{\theta}_k)\right]$$

แต่การที่จะหาค่า $d(\theta_0,\hat{\theta}_k)$ นั้นเป็นไปไม่ได้เพราะเราไม่ทราบข้อมูลเกี่ยวกับ θ_0 ดังนั้นอาไคเคะจึงได้เสนอ $-2\ln f(y|\hat{\theta}_k)$ เป็นตัวประมาณที่เอนเอียงของ $d(\theta_0,\hat{\theta}_k)$ และได้เสนอตัวปรับแก้ความไม่เอนเอียง (Bias Adjustment)ดังนั้น

$$\because\ B_1(k,\theta_0) = E_{\theta_0}\left[d(\theta_0,\hat{\theta}_k)\right] - E_{\theta_0}\left[-2\ln f(y|\hat{\theta}_k)\right] \approx 2(k+1)$$

จะได้ว่า

$$AIC = -2\ln f(y|\hat{\theta}_k) + 2(k+1)$$

เนื่องจาก AIC เป็นตัวประมาณที่ไม่เอนเอียงโดยประมาณของ $d(\theta_0,\hat{\theta}_k)$ เมื่อขนาดตัวอย่างมากและจำนวนตัวแปรน้อยหรือเป็นตัวประมาณไม่เอนเอียงที่ลู่เข้าสู่อนันต์ (asymptotically biased estimator) แต่ในทางกลับกันถ้าขนาดตัวอย่าง (n) เล็กจะทำให้ AIC เป็นตัวประมาณเอนเอียงของ $d(\theta_0,\hat{\theta}_k)$ เพราะ จะมีค่าน้อยกว่าตัวปรับแก้ความเอนเอียง $B_1(k,\theta_0)$ อย่างมาก จึงทำให้ AIC มีความเอนเอียงด้านลบ

ดังนั้นเพื่อกำจัดความเอนเอียงนี้ ในปี ค.ศ.1989 เฮอวิชและไซ ได้เสนอเกณฑ์ข้อสนเทศที่ได้รับการปรับแก้ความเอนเอียงนี้ซึ่งเรียกว่า เกณฑ์ AIC$_C$ โดยการหาค่าที่แท้จริงของตัวปรับแก้ความเอนเอียง $B_1(k,\theta_0)$ กล่าวคือ

$$B_1(k,\theta_0)=\frac{2n(k+1)}{(n-k-2)}$$

$$\therefore\ \ AIC_c=-2\ln f\left(y\middle|\hat{\theta}_k\right)+\frac{2n(k+1)}{n-k-2}$$

จึงสรุปได้ว่า AIC$_C$ เป็นตัวประมาณไม่เอนเอียงอย่างแน่นอน(An Exactly Unbiased Estimator) ของ $d(\theta_0,\hat{\theta}_k)$ ดังนั้น AIC$_C$ จึงเป็นเกณฑ์ที่น่าจะเหมาะกับตัวอย่างขนาดเล็กมากกว่า AIC และสามารถเขียน AIC$_C$ ได้ใหม่สำหรับการแจกแจงแบบปกติตามรูปแบบทั่วไปที่กำหนดในข้อตกลงเบื้องต้น ดังนี้

$$AIC_c=-2\ln f(y|\hat{\theta}_k)+\frac{2n(k+1)}{n-k-2}$$

$$=-2\ln\left[(2\pi\sigma^2)^{\frac{-n}{2}}\exp\left(\frac{-\left(\underset{\sim}{y}-X\underset{\sim M}{\hat{\beta}}\right)'\left(\underset{\sim}{y}-X\underset{\sim M}{\hat{\beta}}\right)}{2\sigma^2}\right)\right]+\frac{2n(k+1)}{n-k-2}$$

$$=-2\left[-\frac{n}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\left(\underset{\sim}{y}-X\underset{\sim M}{\hat{\beta}}\right)'\left(\underset{\sim}{y}-X\underset{\sim M}{\hat{\beta}}\right)\right]+\frac{2n(k+1)}{n-k-2}$$

$$=-2\left[-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{1}{2\sigma^2}\left(\underset{\sim}{y}-X\underset{\sim M}{\hat{\beta}}\right)'\left(\underset{\sim}{y}-X\underset{\sim M}{\hat{\beta}}\right)\right]+\frac{2n(k+1)}{n-k-2}$$

$$=-2\left[-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\sigma^2-\frac{n}{2}\right]+\frac{2n(k+1)}{n-k-2}$$

$$=-2\left[-\frac{n}{2}(\ln(2\pi)+1)-\frac{n}{2}\ln\hat{\sigma}^2\right]+\frac{2n(k+1)}{n-k-2}$$

$$=n(\ln(2\pi)+1)+n\ \ln\hat{\sigma}^2+\frac{2n(k+1)}{n-k-2}$$

เนื่องจากเทอมของ $n(\ln(2\pi)+1)$ เป็นค่าคงที่สำหรับทุกตัวแบบที่นำมาพิจารณา ดังนั้นจึงละไว้ได้ และเขียน AIC$_C$ ได้ดังนี้

$$AIC_c = n\ln\left(\hat{\sigma}^2\right) + \frac{2n(k+1)}{n-k-2}$$

2.4 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของชวาร์ซที่ปรับแก้ (SIC_C)

ในปี ค.ศ. 1978 ชวาร์ซ(Schwarz) ได้เสนอเกณฑ์การคัดเลือกตัวแบบ ซึ่งใช้เป็นเกณฑ์ในการคัดเลือกตัวแบบทางสถิติที่ใช้กันอย่างแพร่หลาย นั่นคือเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซ(SIC , BIC,BSC)

เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของชวาร์ซได้นำแนวคิดของเบส์มาพัฒนาดังรายละเอียดต่อไปนี้

กำหนด

$p\left(X \middle| M_j, \underset{\sim}{\beta}_j\right)$ แทนความน่าจะเป็นของ X เมื่อกำหนดตัวแบบ M_j และเวกเตอร์พารามิเตอร์ $\underset{\sim}{\beta}_j$

$p(M_j)$ แทนความน่าจะเป็นก่อนของตัวแบบ M_1 ,…,M_j เมื่อ $1 \leq j \leq L$ เมื่อ L คือจำนวนตัวแบบที่นำมาพิจารณาทั้งหมดซึ่งมี 2^p ตัวแบบ

$p\left(\underset{\sim}{\beta}_j \middle| M_j\right)$ แทนความน่าจะเป็นก่อนบน $\underset{\sim}{\beta}$ เมื่อกำหนดตัวแบบ M_j

จากทฤษฎีของเบส์ (Bayes's Theorem) จะได้ความน่าจะเป็นร่วมภายหลังของ M_j และ $\underset{\sim}{\beta}_j$ ดังนี้

$$p\left(M_j, \underset{\sim}{\beta}_j \middle| X\right) = \frac{p(M_j)\,p\left(\underset{\sim}{\beta}_j \middle| M_j\right) p\left(X \middle| M_j, \underset{\sim}{\beta}_j\right)}{p(X)}$$

เมื่อ $p(X)$ แทนการแจกแจงส่วนริม (Marginal Distribution) ของ X

ในการคัดเลือกตัวแบบจะเลือกตัวแบบ M_j ซึ่งมีความน่าจะเป็นภายหลังมากที่สุด การแจกแจงภายหลัง (Posterior Distribution)ของตัวแบบ M_j มีรูปแบบดังนี้

$$p\left(M_j \middle| X\right) = \int p\left(M_j, \underset{\sim}{\beta}_j \middle| X\right) d\underset{\sim}{\beta}_j$$

เนื่องจาก $p\left(M_j, \underset{\sim}{\beta}_j \middle| X\right) = L\left(\underset{\sim}{\beta}_j \middle| X\right)$ ดังนั้นสามารถเขียนสูตร $p\left(M_j \middle| X\right)$ ได้ดังนี้

$$p(M_j|X) = p(X)^{-1}\, p(M_j) \int L\left(\underset{\sim}{\beta}_j \middle| X\right) p\left(\underset{\sim}{\beta}_j \middle| M_j\right) d\underset{\sim}{\beta}_j$$

เนื่องจากเราสามารถหาค่าต่ำสุดของ $-2\ln p(M_j|X)$ แทนการหาค่าสูงสุดของ $p(M_j|X)$ ได้ ดังนั้นจะได้ว่า

$$-2\ln p(M_j|X) = 2\ln p(X) - 2\ln p(M_j) - 2\ln\left[\int L\left(\underset{\sim}{\beta}_j\middle|X\right)p\left(\underset{\sim}{\beta}_j\middle|M_j\right)d\underset{\sim}{\beta}_j\right]$$

จะเห็นว่าเทอม $2\ln p(X)$ เป็นค่าคงที่สำหรับทุกตัวแบบ จึงสามารถละไว้ได้ เพราะฉะนั้น

$$-2\ln p(M_j|X) \approx -2\ln p(M_j) - 2\ln\left[\int L\left(\underset{\sim}{\beta}_j\middle|X\right)p\left(\underset{\sim}{\beta}_j\middle|M_j\right)d\underset{\sim}{\beta}_j\right]$$

$$\approx -2\ln L\left(\underset{\sim}{\beta}_j\middle|X\right) + (\dim(\underset{\sim}{\beta}_j))\ln(n))$$

$$\therefore \qquad SIC = -2\ln L\left(\underset{\sim}{\beta}_j\middle|X\right) + (\dim(\underset{\sim}{\beta}_j))\ln(n))$$

หรือ $$SIC = -2\ln L\left(\underset{\sim}{\beta}_j\middle|X\right) + k\ln(n))$$

ดังนั้น SIC สำหรับการแจกแจงแบบปกติตามรูปแบบทั่วไปที่กำหนดในข้อตกลงเบื้องต้นสามารถเขียนได้ใหม่ดังนี้

$$SIC = -2\ln L\left(\underset{\sim}{\beta}_j\middle|X\right) + k\ln(n))$$

$$= -2\ln\left[(2\pi\sigma^2)^{\frac{-n}{2}}\exp\left(\frac{-\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)}{2\sigma^2}\right)\right] + k\ln(n)$$

$$= -2\left[-\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)\right] + k\ln(n)$$

$$= -2\left[-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)'\left(\underset{\sim}{y} - X\hat{\underset{\sim}{\beta}}_M\right)\right] + k\ln(n)$$

$$= -2\left[-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{n}{2}\right] + k\ln(n)$$

$$= -2\left[-\frac{n}{2}(\ln(2\pi) + 1) - \frac{n}{2}\ln\hat{\sigma}^2\right] + k\ln(n)$$

$$= n(\ln(2\pi) + 1) + n\ln\hat{\sigma}^2 + k\ln(n)$$

จะได้ว่า

(2.4.1) $$SIC = n(\ln(2\pi)+1) + n\ln\hat{\sigma}^2 + k\ln(n)$$

จากสมการ (2.4.1) จะเห็นได้ว่าเทอมของ $n(\ln(2\pi)+1)$ เป็นค่าคงที่สำหรับทุกตัวแบบที่นำมาพิจารณา ดังนั้นจึงละไว้ได้และเขียน SIC ได้ดังนี้

(2.4.2) $$SIC = n\ln(\hat{\sigma}^2) + k\ln(n)$$

จากการศึกษาพบว่า SIC เป็นตัวประมาณที่คงเส้นคงวาเมื่อตัวอย่างมีขนาดใหญ่ แต่จากการศึกษาของแมคไควร์รี (McQuarrie)ในปี ค.ศ.1999 ในกรณีตัวอย่างขนาดเล็กพบว่า $n\ln(\hat{\sigma}^2)$ มีค่าลดลงอย่างรวดเร็วกว่า $k\ln(n)$ มาก ทำให้ค่า SIC ไม่มีประสิทธิภาพในการคัดเลือกตัวแบบ ดังนั้นจึงได้เสนอเกณฑ์ SIC_C โดยอาศัยแนวคิดของเฮอวิชและไซ ที่ใช้ในการหา AIC_c โดยพิจารณาความสัมพันธ์ระหว่าง AIC กับ AIC_c ดังนี้

(2.4.3) $$AIC_c = n\ln\hat{\sigma}^2 + 2(k+1)$$

และ

(2.4.4) $$AIC_c = n\ln\hat{\sigma}^2 + \frac{2n(k+1)}{n-k-2}$$

ดังนั้นจะเห็นว่า AIC กับ AIC_c แตกต่างกันเฉพาะเทอมที่ 2 ที่เรียกว่า Penalty function ซึ่งความสัมพันธ์ระหว่าง Penalty function ของทั้งสอง คือ Penalty function ของ AIC ถูกปรับมาเป็น AIC_c โดยคูณด้วย $\frac{n}{n-k-2}$ นั่นเอง ดังนั้นเขาจึงนำแนวคิดนี้มาใช้ในการปรับแก้ SIC และสรุปได้ว่า SIC_C มีรูปแบบสมการดังนี้

$$SIC_c = n\ln(\hat{\sigma}^2) + \frac{nk\ln(n)}{n-k-2}$$

2.5 เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ (HQ_C)

ในปี ค.ศ.1979 แฮนแนนและควินน์ (Hannan and Quinn) ได้เสนอเกณฑ์การคัดเลือกตัวแบบสำหรับการวิเคราะห์อนุกรมเวลา (Autoregression Model) ไว้ซึ่งสามารถใช้งานได้อย่างมีประสิทธิภาพและสามารถนำมาประยุกต์ใช้ในการวิเคราะห์การถดถอยเชิงเส้นได้

แฮนแนนและควินน์ได้สร้างเกณฑ์การคัดเลือกตัวแบบที่มีรูปแบบคล้ายกับ AIC แต่ยังต้องคงคุณสมบัติความคงเส้นคงวาอยู่เมื่อพิจารณา Penalty function ของ AIC ที่อยู่ในรูปแบบ αk

ดังนั้นเขาจึงได้เสนอเกณฑ์การคัดเลือกตัวแบบที่มีรูปแบบดังนี้

$$\psi(k) = n\ln(\hat{\sigma}^2) + kC_n$$

เมื่อ C_n ของ AIC คือ 2

และ C_n ของ SIC คือ $\ln(n)$

แฮนแนนและควินน์รักษาคุณสมบัติความคงเส้นคงวาได้โดยที่ C_n ต้องมีค่าน้อยกว่า $\ln(n)$ ดังนั้นได้เสนอเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ (HQ) ที่มีรูปแบบดังนี้

$$\psi(k) = n\ln(\hat{\sigma}^2) + 2ck(\ln\ln(n)) \qquad \text{เมื่อ } c > 1$$

ในปี ค.ศ. 1980 แฮนแนนพบว่า เกณฑ์ HQ ยังคงคุณสมบัติคงเส้นคงวาแม้ c = 1 ดังนั้นเกณฑ์ HQ ที่นิยมใช้ทั่วไป จึงมีรูปแบบดังนี้

$$HQ = n\ln(\hat{\sigma}^2) + 2k(\ln\ln(n))$$

ต่อมาในปี ค.ศ. 1998 แมคไควร์รีและไซ (McQuarrie and Tsi) ได้พัฒนาเกณฑ์การคัดเลือกตัวแบบเพื่อใช้งานในกรณีที่ขนาดตัวอย่างเล็ก ดังนั้นเขาจึงได้เสนอเกณฑ์ HQ_C โดยอาศัยแนวคิดของเฮอวิชและไซที่ใช้ในการหา AIC_c โดยพิจารณาความสัมพันธ์ระหว่าง AIC กับ AIC_c ซึ่งแตกต่างกันเฉพาะ เทอมที่ 2 ที่เรียกว่า Penalty function และความสัมพันธ์ระหว่าง Penalty function ของทั้งสอง คือ Penalty function ของ AIC ถูกปรับมาเป็น AIC_c โดยคูณด้วย $\dfrac{n}{n-k-2}$ นั่นเอง ดังนั้นเขาจึงนำแนวคิดนี้มาใช้ในการปรับแก้ HQ ทำให้ได้ว่า HQ_C มีรูปสมการดังนี้

$$HQ_C = n\ln(\hat{\sigma}^2) + \frac{2nk(\ln\ln(n))}{n-k-2}$$

2.6 เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (KIC_C)

ในปี ค.ศ. 1999 โจเซฟ.อี.คาวานอท (Joseph.E.Cavanaugh) ได้เสนอเกณฑ์การคัดเลือกตัวแบบที่มีความแม่นยำ โดยมีพื้นฐานแนวคิดจาก Kullback's Symmetric Divergence ซึ่งแตกต่างจากเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะ (AIC) เพราะ AIC จะวัดระยะทางโดยใช้ข้อสนเทศของคูลส์แบล็คแบบระบุทิศทาง (Kullback's Directed Divergence) ซึ่งจะไม่มีคุณสมบัติสมมาตรเพราะการวัดระยะทางในทิศทางหนึ่งอาจได้ผลแตกต่างจากอีกทิศทางหนึ่ง ดังนั้นเขาจึงได้เสนอเกณฑ์ที่อาศัยพื้นฐานการวัดระยะทางที่สมมาตรคือข้อสนเทศของคูลส์แบล็คแบบสมมาตร (Kullback's Symmetric Divergence(KSD)) ซึ่งมีรายละเอียดดังนี้

KSD มีรูปแบบดังนี้

$$I(\theta_0,\theta_k) = E_{\theta_0}\left\{\ln\frac{f(y|\theta_0)}{f(y|\theta_k)}\right\}$$

เมื่อ $I(\theta_0,\theta_k)$ เป็น KSD ระหว่าง ตัวแบบที่แท้จริงกับตัวแบบที่นำมาพิจารณาภายใต้ตัวแบบที่แท้จริง

$f(y|\theta_0)$ เป็นฟังก์ชันความหนาแน่นของตัวแบบที่แท้จริง

$f(y|\theta_k)$ เป็นฟังก์ชันความหนาแน่นของตัวแบบที่นำมาพิจารณา

และ $E_{\theta_0}\{.\}$ เป็นค่าคาดหวังภายใต้ตัวแบบที่แท้จริง

ในทำนองเดียวกัน $I(\theta_k,\theta_0)$ เป็น KSD ระหว่างตัวแบบที่นำมาพิจารณาแท้จริงกับตัวแบบที่แท้จริงภายใต้ตัวแบบที่นำมาพิจารณา จะมีรูปแบบดังนี้

$$I(\theta_k,\theta_0) = E_{\theta_k}\left\{\ln\frac{f(y|\theta_k)}{f(y|\theta_0)}\right\}$$

ซึ่งจะได้ว่า $I(\theta_0,\theta_k) \neq I(\theta_k,\theta_0)$ เว้นแต่ $\theta_k = \theta_0$

KSDมีรูปแบบดังนี้

$$\begin{aligned} J(\theta_0,\theta_k) &= E_{\theta_0}\left(\ln\frac{f(y|\theta_0)}{f(y|\theta_k)}\right) + E_{\theta_k}\left(\ln\frac{f(y|\theta_k)}{f(y|\theta_0)}\right) \\ &= I(\theta_0,\theta_k) + I(\theta_k,\theta_0) \end{aligned}$$

จากการศึกษาพบว่า $I(\theta_0,\theta_k)$ จะแสดงความคลาดเคลื่อนที่เกิดจากความแปรปรวนได้ ในขณะที่ $I(\theta_k,\theta_0)$ จะแสดงความคลาดเคลื่อนที่เกิดจากความเอนเอียงของการประมาณ ดังนั้น เมื่อนำคุณสมบัติทั้งสองประการมาพิจารณาร่วมกันก็น่าจะได้ข้อสนเทศที่ดีในการคัดเลือกตัวแบบ

$$\because \quad d(\theta_0,\theta_k) = E_{\theta_0}\{-2\ln f(y|\theta_k)\}$$

$$\therefore \quad 2I(\theta_0,\theta_k) = d(\theta_0,\theta_k) - d(\theta_0,\theta_0)$$

เนื่องจากค่าของ $d(\theta_0,\theta_0)$ ไม่ขึ้นกับ θ_k ดังนั้นค่าของ $I(\theta_0,\theta_k)$ จึงขึ้นอยู่กับ $d(\theta_0,\theta_k)$ เท่านั้น จึงได้ว่า

$$2J(\theta_0,\theta_k) = (d(\theta_0,\theta_k) - d(\theta_0,\theta_0)) + (d(\theta_k,\theta_0) - d(\theta_k,\theta_k))$$

ส่วน $K(\theta_0,\theta_k)$ ที่สมมูลกับ $J(\theta_0,\theta_k)$ มีรูปแบบดังนี้

$$K(\theta_0,\theta_k) = d(\theta_0,\theta_k) + (d(\theta_k,\theta_0) - d(\theta_k,\theta_k))$$

และจะใช้ $K(\theta_0,\hat{\theta}_k)$ เป็นตัวประมาณของ $K(\theta_0,\theta_k)$

เมื่อ $K(\theta_0,\hat{\theta}_k) = d(\theta_0,\hat{\theta}_k) + (d(\hat{\theta}_k,\theta_0) - d(\hat{\theta}_k,\hat{\theta}_k))$

โดยที่ $d(\theta_0,\hat{\theta}_k) = E_{\theta_0}\left\{-2\ln f(y|\theta_k)\right\}\Big|\theta_k = \hat{\theta}_k$

$d(\hat{\theta}_k,\theta_0) = E_{\theta_k}\left\{-2\ln f(y|\theta_0)\right\}\Big|\theta_k = \hat{\theta}_k$

และ $d(\hat{\theta}_k,\hat{\theta}_k) = E_{\theta_k}\left\{-2\ln f(y|\theta_k)\right\}\Big|\theta_k = \hat{\theta}_k$

การหาค่า $d(\theta_0,\hat{\theta}_k)$ นั้นเป็นไปไม่ได้เพราะเราไม่ทราบข้อมูลเกี่ยวกับ θ_0 ในทำนองเดียวกับ AIC จึงใช้ $\left\{-2\ln f(y|\hat{\theta}_k)\right\}$ เป็นตัวประมาณของ $K(\theta_0,\hat{\theta}_k)$ แทน แต่พบว่ายังมีความเอนเอียงอยู่จึงทำการปรับแก้ความเอนเอียงด้วยตัวปรับแก้ความเอนเอียง ดังนี้

$$\begin{aligned} E_{\theta_0}(K(\theta_0,\hat{\theta}_k)) - E_{\theta_0}(-2\ln f(y|\hat{\theta}_k)) &= E_{\theta_0}(d(\theta_0,\hat{\theta}_k)) - E_{\theta_0}(-2\ln f(y|\hat{\theta}_k)) \\ &\quad + E_{\theta_0}(d(\hat{\theta}_k,\theta_0)) - E_{\theta_0}(d(\hat{\theta}_k,\hat{\theta}_k)) \end{aligned}$$

จะได้ว่า

$$E_{\theta_0}(K(\theta_0,\hat{\theta}_k)) = E_{\theta_0}(-2\ln f(y|\hat{\theta}_k)) + B_1(k,\theta_0) + B_2(k,\theta_0)$$

เมื่อ $B_1(k,\theta_0) = E_{\theta_0}(d(\theta_0,\hat{\theta}_k)) - E_{\theta_0}(-2\ln f(y|\hat{\theta}_k)) \approx 2(k+1)$

และ $B_2(k,\theta_0) = E_{\theta_0}(d(\hat{\theta}_k,\theta_0)) - E_{\theta_0}(d(\hat{\theta}_k,\hat{\theta}_k)) \approx (k+1)$

ดังนั้น

$$KIC = -2\ln f(y|\hat{\theta}_k) + 3(k+1)$$

เนื่องจาก KIC เป็นตัวประมาณไม่เอนเอียงโดยประมาณของ $K(\theta_0,\hat{\theta}_k)$ เมื่อขนาดตัวอย่างมากและจำนวนตัวแปรน้อยหรือเป็นตัวประมาณที่ไม่เอนเอียงที่ลู่เข้าสู่อนันต์ แต่ในทางกลับกันถ้าขนาดตัวอย่างเล็กจะทำให้ KIC เป็นตัวประมาณเอนเอียงของ $K(\theta_0,\hat{\theta}_k)$ เช่นเดียวกับ AIC

ในปี ค.ศ.2004 โจเซฟ.อี.คาวานอทได้เสนอเกณฑ์ข้อสนเทศที่ได้รับการปรับแก้ความเอนเอียงนี้ซึ่งเรียกว่าเกณฑ์ KIC_C โดยการหาค่าที่แท้จริงของตัวปรับแก้ความเอนเอียง $B_2(k,\theta_0)$ กล่าวคือ

$$B_2(k,\theta_0) = n\ln\frac{n}{2} - n\psi\left(\frac{n-k}{2}\right)$$

เมื่อ $\psi(.)$ คือdigamma หรือ phi function

และสามารถเขียน KIC_c ได้ดังสมการต่อไปนี้

$$KIC_c = -2\ln f(y|\hat{\theta}_k) + \frac{2n(k+1)}{n-k-2} - n\psi\left(\frac{n-k}{2}\right) + n\ln\frac{n}{2}$$

ในการคำนวณ phi function ค่อนข้างมีความยุ่งยากเพราะอยู่ในรูปสมการดังนี้

$$n\ln\frac{n}{2} - n\psi\left(\frac{n-k}{2}\right) = n\ln\left(\frac{n}{n-k}\right) + \frac{n}{n-k} + O(n^{-2})$$

$$\approx n\ln\left(\frac{n}{n-k}\right) + \frac{n}{n-k}$$

เพราะฉะนั้น

$$KIC_c = -2\ln f(y|\hat{\theta}_k) + n\ln\left(\frac{n}{n-k}\right) + \frac{n((n-k)(2k+3)-2)}{(n-k-2)(n-k)}$$

ดังนั้น KIC_C สำหรับการแจกแจงแบบปกติตามรูปแบบทั่วไปที่กำหนดในข้อตกลงเบื้องต้นสามารถเขียนได้ใหม่ดังนี้

$$KIC_c = -2\ln f(y|\hat{\theta}_k) + n\ln\left(\frac{n}{n-k}\right) + \frac{n((n-k)(2k+3)-2)}{(n-k-2)(n-k)}$$

$$\because -2\ln f(y|\hat{\theta}_k) = -2\ln\left[(2\pi\sigma^2)^{\frac{-n}{2}}\exp\left(\frac{-\left(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\beta}_M\right)'\left(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\beta}_M\right)}{2\sigma^2}\right)\right]$$

$$= -2\left[-\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\beta}_M\right)'\left(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\beta}_M\right)\right]$$

$$= -2\left[-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\beta}_M\right)'\left(\underset{\sim}{y} - \underset{\sim}{X}\,\hat{\beta}_M\right)\right]$$

$$= -2\left[-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{n}{2}\right]$$

$$= -2\left[-\frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\hat{\sigma}^2\right]$$

$$= n(\ln(2\pi)+1) + n\ln\hat{\sigma}^2$$

$$\therefore\ KIC_c = n(\ln(2\pi)+1) + n\ln\hat{\sigma}^2 + n\ln\left(\frac{n}{n-k}\right) + \frac{n((n-k)(2k+3)-2)}{(n-k-2)(n-k)}$$

เนื่องจากเทอมของ $n\,(\ln(2\pi)+1)$ เป็นค่าคงที่สำหรับทุกตัวแบบที่นำมาพิจารณา จึงละไว้ได้ และเขียน KIC_C ได้ดังนี้

$$KIC_c = n\ln\hat{\sigma}^2 + n\ln\left(\frac{n}{n-k}\right) + \frac{n((n-k)(2k+3)-2)}{(n-k-2)(n-k)}$$

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็กเพื่อหาตัวแบบที่ให้ค่าพยากรณ์ถูกต้องมากที่สุด โดยใช้ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยเป็นเกณฑ์ในการตัดสินใจ ซึ่งเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณประกอบด้วย 4 เกณฑ์ ดังนี้

1) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (Corrected Akaike's Information Criterion (AIC_C))
2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (Corrected Schwarz's Information Criterion (SIC_C))
3) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ (Corrected Hannan and Quinn's Information Criterion (HQ_C))
4) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (Corrected Kullback's's Information Criterion (KIC_C))

ผู้วิจัยจะใช้เทคนิคมอนติคาร์โล (Monte Carlo Technique) ในการหาข้อสรุปของปัญหาที่ศึกษา และใช้โปรแกรม Borland Delphi 7 ในการประมวลผลและวิเคราะห์ข้อมูล ซึ่งขั้นตอนในการวิจัยมีรายละเอียดดังนี้

3.1 การจำลองข้อมูลด้วยวิธีมอนติคาร์โล

วิธีมอนติคาร์โลเป็นเทคนิคในการจำลองแบบทางคณิตศาสตร์ที่นิยมใช้กันอย่างแพร่หลาย โดยการจำลองตัวเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษาที่ยังไม่แน่ใจในผลลัพธ์ที่จะเกิดขึ้น เพราะเลขสุ่มมีประโยชน์หลายประการ ดังนี้

1) การเลือกตัวอย่างจะเกิดความไม่เอนเอียงในการสำรวจหรือทดลองในเรื่องนั้นๆ ทั้งนี้เพราะเลขสุ่มมาจากแนวคิดเกี่ยวกับการคำนวณความน่าจะเป็น

2) การใช้เลขสุ่มจะทำให้ได้มาซึ่งรูปแบบต่างๆ หรือวิธีการที่สลับซับซ้อนโดยการสร้างสถานการณ์การจำลอง (simulation)

3) การใช้เลขสุ่มอาจทำการศึกษาคุณสมบัติทางทฤษฎีของกระบวนการทางสถิติที่มีความสำคัญต่อการประมาณค่า ตลอดจนนำไปสู่การอธิบายเกี่ยวกับอำนาจการทดสอบทางสถิติ (power of statistical tests)

4) เพื่อหาคำตอบในปัญหาทางคณิตศาสตร์ โดยพิจารณาการแจกแจงความน่าจะเป็นของปัญหานั้นๆ

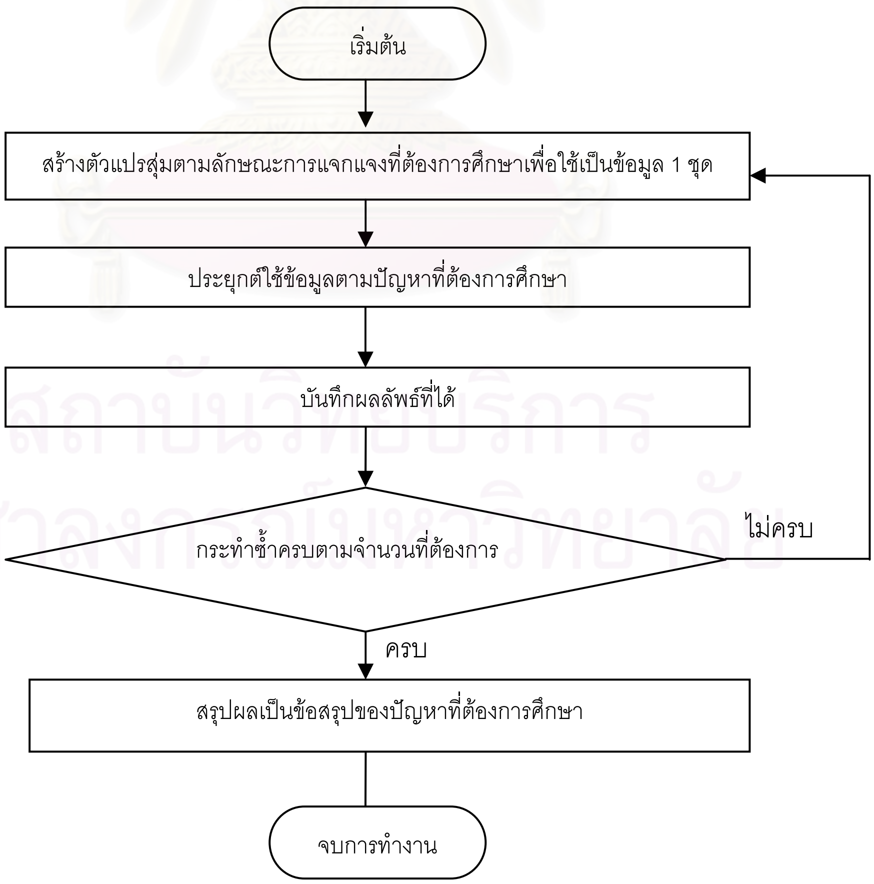
หลักการของเทคนิคมอนติคาร์โลคือการนำเลขสุ่มมาประยุกต์ในการแก้ปัญหาต่างๆ ที่สนใจศึกษาถึงผลสรุปของปัญหานั้นๆ โดยมีขั้นตอนที่สำคัญ 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 **การสร้างเลขสุ่ม**(generate random number)การสร้างเลขสุ่มจะกำหนดให้มีการแจกแจงแบบยูนิฟอร์มในช่วง [0,1] และเป็นอิสระซึ่งกันและกัน จากนั้นนำเลขสุ่มนี้ไปสร้างตัวแปรตามลักษณะการแจกแจงที่ต้องการในปัญหาที่ศึกษา เพื่อเป็นข้อมูลของปัญหานั้นๆ

ขั้นตอนที่ 2 **การประยุกต์ปัญหาที่ต้องการศึกษาโดยใช้เลขสุ่ม** ขั้นตอนนี้ขึ้นอยู่กับปัญหาที่ศึกษา ซึ่งขั้นตอนที่นำเลขสุ่มมาใช้ในการหาค่าต่างๆ ตามปัญหาที่ต้องการตามสูตรการคำนวณในปัญหาที่ศึกษา

ขั้นตอนที่ 3 **การทดลองกระทำซ้ำ** คือการทดลองโดยใช้กระบวนการของการสุ่ม (random process) มากระทำในลักษณะซ้ำๆ กัน (replication) จำนวนหลายครั้ง เพื่อหาคำตอบของปัญหาที่ต้องการ โดยถือว่าการทำซ้ำๆ กันนั้นเป็นวิธีการเก็บรวบรวมข้อมูลให้มีข้อมูลจำนวนมากเพื่อลดความไม่แน่นอนของคำตอบ

จากขั้นตอนเทคนิคมอนติคาร์โล สามารถเขียนแผนผังได้ดังนี้



<u>รูปที่ 3.1 แสดงผังงานสำหรับขั้นตอนของวิธีมอนติคาร์โล</u>

3.2 แผนการทดลอง

ผู้วิจัยได้กำหนดสถานการณ์ต่างๆ สำหรับการวิจัยครั้งนี้ไว้ดังนี้

1) เลือกตัวอย่างสุ่มเพื่อใช้เป็นค่าคลาดเคลื่อนสุ่ม (ε_i) จากประชากรที่มีการแจกแจงเดียวกัน โดยการวิจัยครั้งนี้สนใจศึกษาเฉพาะตัวอย่างสุ่มที่มาจากการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ 1 , 2 , 4 , 8 และ 16

2) กำหนดขนาดตัวอย่างที่ศึกษา คือ 12 , 15 , 18 , 21 , 24 , 27 และ 30

3) กำหนดจำนวนตัวแปรอิสระที่ศึกษา คือ 3 , 5 และ 7

4) กำหนดให้ $\underset{\sim}{\beta}' = (1, 1, 1, \ldots, 1)_{1\times(p+1)}$

3.3 ขั้นตอนในการดำเนินการวิจัย

ขั้นตอนในการดำเนินการการวิจัยมีดังนี้

1) กำหนดลักษณะการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง และจำนวนตัวแปรอิสระที่ใช้ในการวิจัย

2) สร้างข้อมูลตัวแปรอิสระ และค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่กำหนด

3) สร้างข้อมูลตัวแปรตามจากตัวแปรอิสระและค่าคลาดเคลื่อน โดยให้ตัวแปรตามมีลักษณะความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ

4) กำหนดตัวแบบเริ่มต้นซึ่งเป็นตัวแบบที่เต็มรูป (full model) ในการคัดเลือกตัวแบบที่เหมาะสม

5) กำหนดตัวแบบที่เป็นไปได้ทั้งหมด ที่ติดกลุ่มของตัวแบบเริ่มต้นเพื่อใช้ในการคัดเลือกตัวแบบตัวแบบที่เหมาะสม

6) ทำการคัดเลือกตัวแบบที่เหมาะสมจากทั้ง 4 เกณฑ์ คือ

6.1) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (AIC_C)

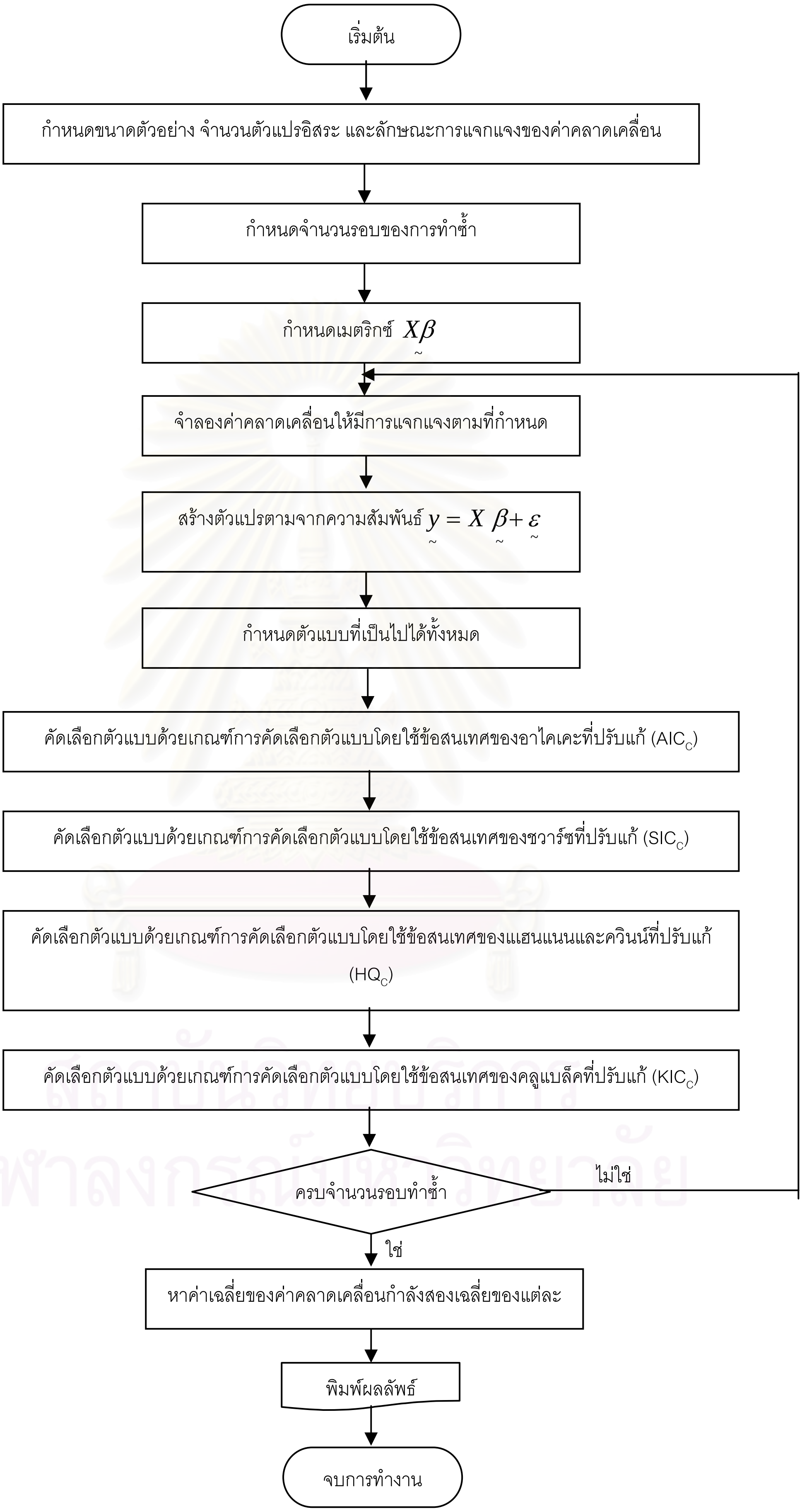
6.2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (SIC_C)

6.3) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ (HQ_C)

6.4) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (KIC_C)

7) คำนวณค่าคลาดเคลื่อนกำลังสองเฉลี่ยจากการทำซ้ำจำนวน 1,000 รอบ ของตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบการถดถอยทั้ง 4 เกณฑ์

8) ทำการเปรียบเทียบค่าคลาดเคลื่อนกำลังสองเฉลี่ยของตัวแบบที่ได้จากเกณฑ์การคัดเลือกตัวแบบการถดถอยทั้ง 4 เกณฑ์ โดยสรุปผลในรูปของตารางและรูปภาพผังแสดงขั้นตอนในการดำเนินการวิจัย ดังแสดงในรูปที่ 3.2



รูปที่ 3.2 แสดงแผนผังสำหรับขั้นตอนในการดำเนินการวิจัย

รายละเอียดของแต่ละขั้นตอนมีดังนี้

ขั้นตอนที่ 1

กำหนดลักษณะของการแจกแจงของค่าคลาดเคลื่อน ขนาดตัวอย่าง และจำนวนตัวแปรอิสระ โดยกำหนดตามแผนการทดลองที่กล่าวมาแล้วข้างต้น

ขั้นตอนที่ 2

การสร้างข้อมูลเพื่อการวิจัยครั้งนี้ประกอบด้วย ข้อมูลตัวแปรอิสระ ข้อมูลค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษา และข้อมูลตัวแปรตาม โดยมีรายละเอียดดังนี้

1) <u>ข้อมูลตัวแปรอิสระ</u>

เนื่องจากการวิจัยครั้งนี้ศึกษาเฉพาะการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณ ในกรณีที่ตัวแปรอิสระเป็นไปตามข้อตกลงเบื้องต้น กล่าวคือ ตัวแปรอิสระเป็นค่าคงที่ไม่มีพหุสัมพันธ์กัน กำลังสูงสุดของตัวแปรอิสระเป็น 1 และไม่มีอันตรกิริยาระหว่างตัวแปรอิสระ ดังนั้นในการสร้างข้อมูลตัวแปรอิสระจะสร้างจากการแจกแจงพหุแบบปกติ (multivariate normal distribution) ที่มีเวกเตอร์ค่าเฉลี่ยเท่ากับ 0 และเมทริกซ์ความแปรปรวนร่วมเป็นเมทริกซ์เอกลักษณ์ (identity matrix) ซึ่งจะทำให้ตัวแปรอิสระแต่ละตัวแปรมีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 และตัวแปรอิสระทุกตัวที่ได้จะไม่มีพหุสัมพันธ์กัน โดยสร้างจากโปรแกรมย่อย creat_x (n , p,x_out()) ที่สร้างขึ้นโดยโปรแกรม Borland Delphi 7 โดยสามารถดูรายละเอียดได้จากภาคผนวก

2) <u>ข้อมูลค่าคลาดเคลื่อน</u>

การวิจัยครั้งนี้สนใจศึกษาเฉพาะที่ค่าคลาดเคลื่อนสุ่มแต่ละค่ามีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนเท่ากับ σ^2 เหมือนกันและเป็นอิสระซึ่งกันและกัน โดยเริ่มจากการสร้างเลขสุ่มให้มีการแจกแจงแบบปกติมาตรฐานจากโปรแกรมย่อย datanormal(n,z()) จากนั้นจึงทำการแปลงเลขสุ่มที่ได้ให้มีการการแจกแจงปกติที่มีค่าเฉลี่ยและความแปรปรวนตามที่ต้องการ โดยใช้โปรแกรมย่อย creat_err(n,variance,error()) ซึ่งรายละเอียดสามารถดูได้จากภาคผนวก

3) <u>ข้อมูลตัวแปรตาม</u>

เมื่อมีข้อมูลตัวแปรอิสระ และข้อมูลค่าคลาดเคลื่อนที่มีลักษณะการแจกแจงตามที่ต้องการศึกษาแล้ว จากนั้นจะทำการสร้างข้อมูลตัวแปรตามให้มีความสัมพันธ์เชิงเส้นในพารามิเตอร์กับตัวแปรอิสระ ซึ่งมีรูปแบบความสัมพันธ์ดังนี้

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

เมื่อ $\underset{\sim}{y}$ เป็นเวกเตอร์ของตัวแปรตาม

X เป็นเมทริกซ์ของตัวแปรอิสระ

$\underset{\sim}{\beta}$ เป็นเวกเตอร์ของพารามิเตอร์ที่กำหนดขึ้น ซึ่งในการวิจัยครั้งนี้กำหนดให้

$\underset{\sim}{\beta}' = (1,1,1,\ldots,1)_{1\times(p+1)}$ นั่นคือ กำหนดให้ตัวแปรอิสระทุกตัวมีอิทธิพลเท่ากันในตัวแบบเริ่มต้น

และ $\underset{\sim}{\varepsilon}$ เป็นเวกเตอร์ค่าคลาดเคลื่อน ซึ่ง $\underset{\sim}{\varepsilon} \sim N_n(\underset{\sim}{0},\sigma^2 \mathrm{I}_n)$

ขั้นตอนที่ 3

กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูป (full model) ในการคัดเลือกตัวแบบที่เหมาะสม

ขั้นตอนที่ 4

กำหนดตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบเริ่มต้นที่กำหนดไว้ เช่น ในตัวแบบเริ่มต้นมีตัวแปรอิสระจำนวน p ตัวแปร จะได้ว่าตัวแบบที่เป็นไปได้ทั้งหมดมีจำนวน 2^p ตัวแบบ สร้างจากโปรแกรมย่อย model_all_out (n,p,x_input(),model_all()) ซึ่งรายละเอียดสามารถดูได้จากภาคผนวก

ขั้นตอนที่ 5

หลังจากที่ข้อมูลครบแล้วจะคัดเลือกตัวแบบที่เหมาะสมจากทั้ง 4 วิธี ดังต่อไปนี้

1) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (AIC_C) โดยใช้ฟังก์ชัน aic(n1,k1,sigma)

2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (SIC_C) โดยใช้ฟังก์ชัน sic(n1,k1,sigma)

4) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้(HQ_C) โดยใช้ฟังก์ชัน aic(n1,k1,sigma)

5) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (KIC_C) โดยใช้ฟังก์ชัน aic(n1,k1,sigma)

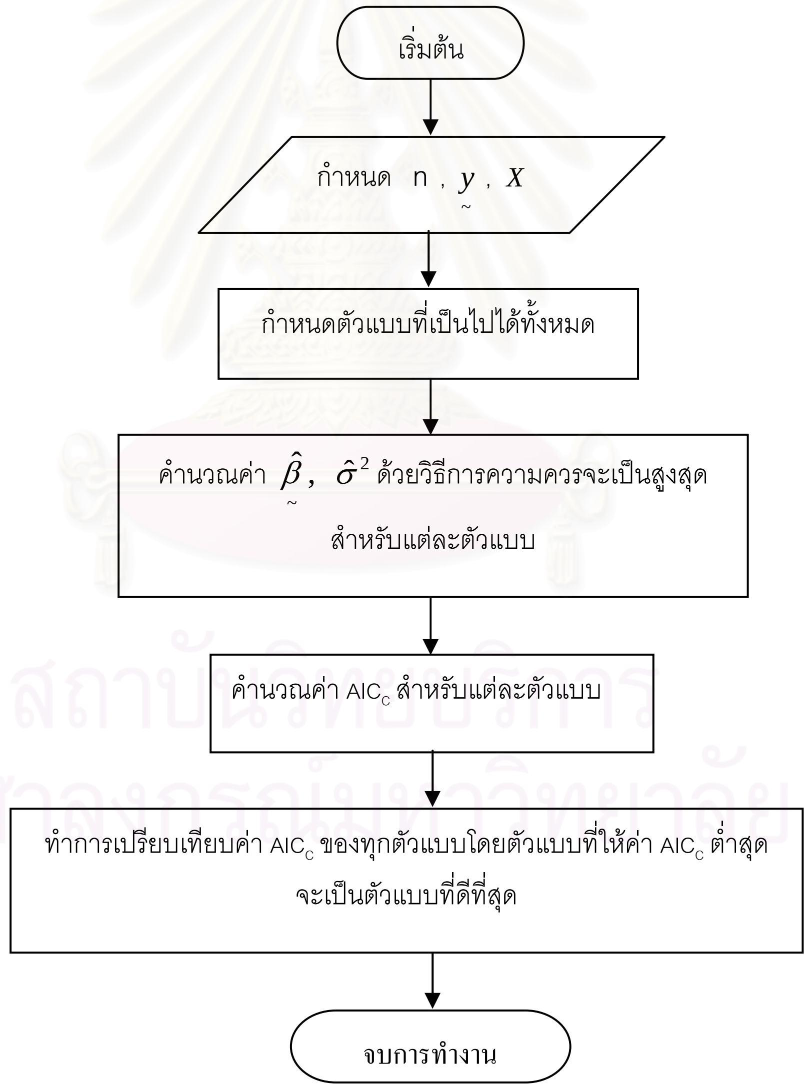
ซึ่งฟังก์ชัน aic() sic() hq() และ kic() สามารถดูรายละเอียดได้จากภาคผนวก

วิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบแต่ละเกณฑ์มีรายละเอียด ดังนี้

1) **เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้**(corrected Akaike's Information Criterion(AIC_C))

การคัดเลือกตัวแบบที่ดีที่สุดโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ จะเริ่มจากการสร้างตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบเริ่มต้นที่กำหนดไว้ จากนั้นจะคำนวณค่า AIC_C สำหรับแต่ละตัวแบบ โดยตัวแบบที่ให้ค่า AIC_C ต่ำที่สุดจะเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยากที่สุด

ผังงานแสดงขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะที่ปรับแก้แล้ว ดังแสดงในรูปที่ 3.3



รูปที่ 3.3 แสดงผังงานขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (AIC_C)

2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้(Corrected Schwarz's Information Criterion (SIC_C))

การคัดเลือกตัวแบบที่ดีที่สุดโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ จะเริ่มจากการสร้างตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบเริ่มต้นที่กำหนดไว้ จากนั้นจะคำนวณค่า SIC_C สำหรับแต่ละตัวแบบ โดยตัวแบบที่ให้ค่า SIC_C ต่ำที่สุดจะเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยากที่สุด

ผังงานแสดงขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของชวาร์ซที่ปรับแก้แล้ว ดังแสดงในรูปที่ 3.4

รูปที่ 3.4 แสดงผังงานขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (SIC_C)

3) **เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้**
(Corrected Hannan and Quinn's Information Criterion (HQ_C))

การคัดเลือกตัวแบบที่ดีที่สุดโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ จะเริ่มจากการสร้างตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบเริ่มต้นที่กำหนดไว้ จากนั้นจะคำนวณค่า HQ_C สำหรับแต่ละตัวแบบ โดยตัวแบบที่ให้ค่า HQ_C ต่ำที่สุดจะเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยากที่สุด

ผังงานแสดงขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ ดังแสดงในรูปที่ 3.5

รูปที่ 3.5 แสดงผังงานขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้ (HQ_C)

4) **เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้**(Corrected Kullback's's Information Criterion (KIC_C))

การคัดเลือกตัวแบบที่ดีที่สุดโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ จะเริ่มจากการสร้างตัวแบบที่เป็นไปได้ทั้งหมดจากตัวแบบเริ่มต้นที่กำหนดไว้ จากนั้นจะคำนวณค่า KIC_C สำหรับแต่ละตัวแบบ โดยตัวแบบที่ให้ค่า KIC_C ต่ำที่สุดจะเป็นตัวแบบที่ให้ค่าพยากรณ์ถูกต้องและแม่นยากที่สุด

ผังงานแสดงขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ ดังแสดงในรูปที่ 3.6

<u>รูปที่ 3.6 แสดงผังงานขั้นตอนของวิธีการคัดเลือกตัวแบบที่เหมาะสมของเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้</u> (KIC_C)

ขั้นตอนที่ 6

คำนวณหาค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) จากการทำซ้ำจำนวน 1,000 รอบ ของทั้ง 4 วิธีเพื่อใช้เป็นเกณฑ์ในการตัดสินใจ และคำนวณค่าส่วนเบี่ยงเบนมาตรฐานของค่าคลาดเคลื่อนกำลังสอง รวมถึงค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(RDAMSE) ของทั้ง 4 วิธีดังกล่าวข้างต้น เพื่อเป็นประกอบการตัดสินใจเกี่ยวกับประสิทธิภาพของวิธีการ

ขั้นตอนที่ 7

ทำการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบที่ได้จากการคัดเลือกตัวแบบการถดถอยทั้ง 4 วิธี โดยสรุปผลในรูปของตารางและรูปภาพเพื่อแสดงการเปรียบเทียบและศึกษาแนวโน้มของแต่ละวิธี

สำหรับรายละเอียดของโปรแกรมทั้งหมดสามารถดูได้จากภาคผนวก

บทที่ 4

ผลการวิจัย

งานวิจัยนี้เป็นการศึกษาเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณที่ดีที่สุดเมื่อตัวอย่างมีขนาดเล็ก เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ที่ได้จากการคัดเลือกตัวแบบการถดถอยซึ่งเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณที่นำมาพิจารณามี 4 เกณฑ์ ดังนี้

1) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (Corrected Akaike's Information Criterion (AIC_C))
2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (Corrected Schwarz's Information Criterion (SIC_C))
3) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของเฮนแนนและควินน์ที่ปรับแก้ (Corrected Hannan and Quinn's Information Criterion (HQ_C))
4) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (Corrected Kullback's's Information Criterion (KIC_C))

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการคัดเลือกตัวแบบการถดถอยวิธีใดจะมีความถูกต้องและแม่นยำมากที่สุด จะพิจารณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Squares Error (AMSE)) ซึ่งวิธีการคัดเลือกตัวแบบใดให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด จะเป็นวิธีการคัดเลือกตัวแบบที่ดีที่สุด และใช้อัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Squares Error (RDAMSE)) ในการประกอบการตัดสินใจเกี่ยวกับประสิทธิภาพของเกณฑ์ต่างๆ ที่ใช้

ขอบเขตของการวิจัยจะเป็นการศึกษาเฉพาะตัวแบบการถดถอยเชิงเส้นพหุคูณที่ประกอบด้วยจำนวนตัวแปรอิสระ 3, 5 และ 7 ตัวแปร ซึ่งใช้สถานการณ์ในกรณีที่ความคลาดเคลื่อนสุ่มมีการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 , 2 ,4 , 8 และ 16 กำหนดขนาดตัวอย่างเป็น 12 , 15 , 18 , 21 , 24 และ 30 ซึ่งการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยในการวิจัยครั้งนี้พิจารณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) และเกณฑ์ค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย(RDAMSE) จากการทำซ้ำจำนวน 1,000 รอบ ในแต่ละสถานการณ์

การนำเสนอผลการวิจัย ผู้วิจัยได้ใช้สัญลักษณ์ต่างๆ ในตารางและการสรุปผล โดยมีความหมายดังนี้

1) σ^2 แทนความแปรปรวนของค่าคลาดเคลื่อนสุ่ม

2) n แทนขนาดตัวอย่าง

3) AIC_C แทนเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (Corrected Akaike's Information Criterion)

4) SIC_C แทนเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (Corrected Schwarz's Information Criterion)

5) HQ_C แทนเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของเฮนแนนและควินน์ที ปรับแก้(Corrected Hannan and Quinn's Information Criterion)

6) KIC_C แทนเกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (Corrected Kullback's's Information Criterion)

ค่าที่แสดงในตารางผลการวิจัยสำหรับแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมา ได้แก่ ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(AMSE) ค่าส่วนเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยซึ่งแสดงในวงเล็บ และค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (RDAMSE) ในบรรทัดที่สามโดยแถบสีเข้มแสดงถึงเกณฑ์การคัดเลือกตัวแบบที่ดีที่สุด ตามลำดับ

ตารางที่ 4.1 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 1

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
1	12	1.1813	1.2842	1.0944	1.4035
		(0.2803)	(0.3091)	(0.2180)	(0.3274)
		7.9404	17.3228	0.0000	28.2438
	15	1.0733	1.1769	1.0411	1.2440
		(0.2431)	(0.2785)	(0.2049)	(0.2896)
		3.0900	13.0439	0.0000	19.4890
	18	1.0358	1.1284	1.0358	1.1056
		(0.1983)	(0.2546)	(0.1983)	(0.2462)
		0.0000	8.9399	0.0000	6.7388
	21	1.0159	1.0813	1.0277	1.0577
		(0.1898)	(0.2419)	(0.1929)	(0.2304)
		0.0000	6.4376	1.1615	4.1146
	24	1.0080	1.0396	1.0174	1.0282
		(0.1747)	(0.2299)	(0.1828)	(0.2215)
		0.0000	3.1349	0.9325	2.0040
	27	1.0010	1.0194	1.0053	1.0095
		(0.1651)	(0.2154)	(0.1762)	(0.2124)
		0.0000	1.8382	0.4296	0.8492
	30	0.9890	1.0020	0.9930	0.9956
		(0.1609)	(0.2052)	(0.1671)	(0.1965)
		0.0000	1.3125	0.4044	0.6673

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.2 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 2

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
2	12	2.2263	2.3644	2.1447	2.4298
		(0.4771)	(0.5242)	(0.4088)	(0.5486)
		3.8047	10.2439	0.0000	13.2932
	15	2.0984	2.2352	2.0819	2.2698
		(0.4476)	(0.4893)	(0.3922)	(0.5080)
		0.7925	7.3635	0.0000	9.0254
	18	2.0600	2.1854	2.0600	2.1715
		(0.3809)	(0.4705)	(0.3809)	(0.4747)
		0.0000	6.0874	0.0000	5.4126
	21	2.0349	2.1367	2.0565	2.1187
		(0.3997)	(0.4628)	(0.3772)	(0.4475)
		0.0000	5.0027	1.0615	4.1181
	24	2.0212	2.0911	2.0397	2.0689
		(0.3448)	(0.4492)	(0.3608)	(0.4349)
		0.0000	3.4583	0.9153	2.3600
	27	2.0164	2.0546	2.0242	2.0444
		(0.3331)	(0.4255)	(0.3595)	(0.4181)
		0.0000	1.8945	0.3868	1.3886
	30	2.0093	2.0418	2.0131	2.0255
		(0.3106)	(0.4082)	(0.3380)	(0.3996)
		0.0000	1.6175	0.1891	0.8063

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.3 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 4

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
4	12	4.3684	4.5750	4.2120	4.6365
		(0.8807)	(0.9923)	(0.7426)	(0.9815)
		3.7132	8.6182	0.0000	10.0783
	15	4.1652	4.3557	4.0823	4.4491
		(0.7693)	(0.9343)	(0.7026)	(0.9241)
		2.0307	6.6972	0.0000	8.9851
	18	4.1150	4.3146	4.1150	4.2879
		(0.6946)	(0.9169)	(0.6946)	(0.8777)
		0.0000	4.8505	0.0000	4.2017
	21	4.0792	4.2480	4.1073	4.2245
		(0.0759)	(0.8827)	(0.6744)	(0.8639)
		0.0000	4.1381	0.6889	3.5620
	24	4.0442	4.1946	4.0923	4.1776
		(0.6612)	(0.8226)	(0.6478)	(0.8384)
		0.0000	3.7189	1.1894	3.2986
	27	4.0108	4.1278	4.0568	4.1088
		(0.6473)	(0.7975)	(0.6114)	(0.8041)
		0.0000	2.9171	1.1469	2.4434
	30	4.0070	4.1185	4.0410	4.0960
		(0.6163)	(0.7516)	(0.5896)	(0.7836)
		0.0000	2.7826	0.8485	2.2211

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.4 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 8

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
8	12	8.3832	8.5878	8.1910	8.6534
		(1.6389)	(1.8404)	(1.4285)	(1.8207)
		2.3465	4.8443	0.0000	5.6452
	15	8.2457	8.4574	8.1090	8.5014
		(1.5815)	(1.5604)	(1.3283)	(1.7623)
		1.6858	4.2965	0.0000	4.8391
	18	7.9771	8.3061	7.9771	8.2278
		(1.2883)	(1.6305)	(1.2883)	(1.5764)
		0.0000	4.1243	0.0000	3.1427
	21	7.9045	8.2186	7.9376	8.1517
		(1.2355)	(1.5788)	(1.2526)	(1.5219)
		0.0000	3.9737	0.4187	3.1273
	24	7.8136	8.1218	7.8587	8.0918
		(1.2205)	(1.5561)	(1.1953)	(1.4824)
		0.0000	3.9444	0.5772	3.5605
	27	7.6400	7.8952	7.7690	7.8780
		(1.1620)	(1.4954)	(1.1537)	(1.4291)
		0.0000	3.3403	1.6885	3.1152
	30	7.4350	7.6746	7.5357	7.6305
		(1.1019)	(1.4091)	(1.0633)	(1.3453)
		0.0000	3.2226	1.3544	2.6295

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

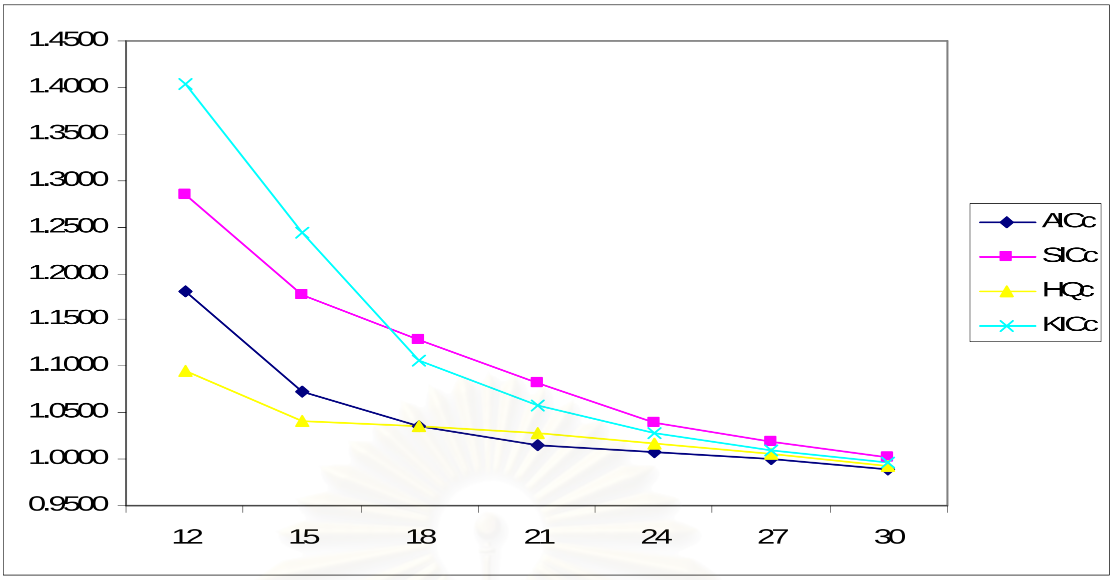
1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.5 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 16

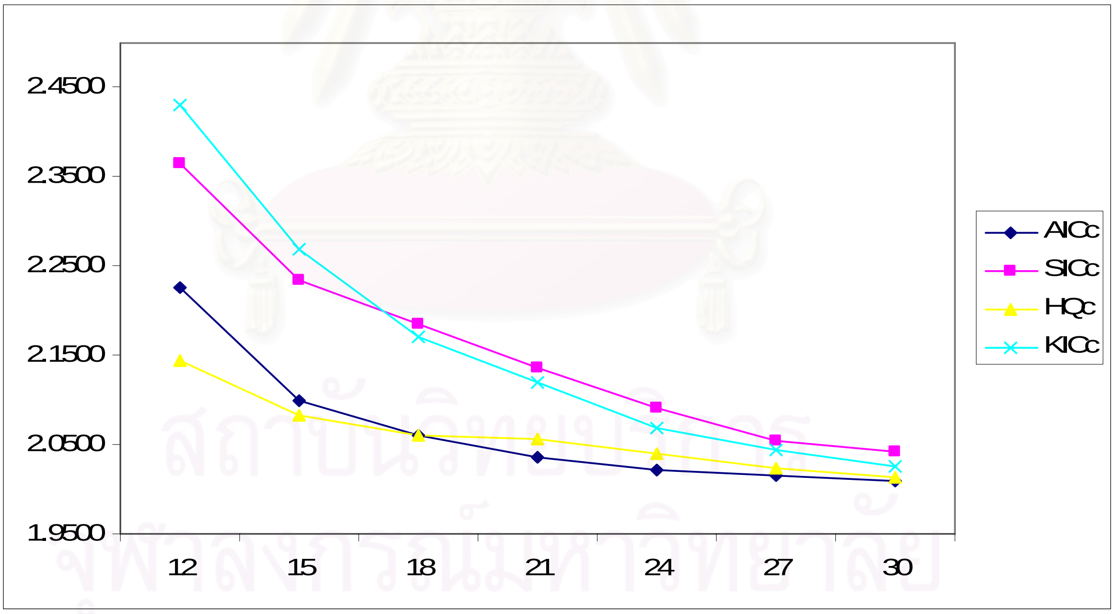
σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
16	12	15.8297	15.9360	15.6050	16.1320
		(3.0947)	(3.4151)	(2.7215)	(3.3942)
		1.4339	2.1211	0.0000	3.3771
	15	15.7290	15.8747	15.5670	15.9350
		(3.0168)	(3.2416)	(2.5499)	(3.3033)
		1.0407	1.9766	0.0000	2.3640
	18	15.5320	15.7625	15.5320	15.7339
		(2.5084)	(3.0942)	(2.5084)	(3.0146)
		0.0000	1.4840	0.0000	1.2999
	21	15.4500	15.6430	15.4964	15.6123
		(2.4148)	(3.0050)	(2.4453)	(2.9148)
		0.0000	1.2492	0.3003	1.0505
	24	15.2460	15.5376	15.3975	15.4780
		(2.3814)	(2.9770)	(2.3420)	(2.8356)
		0.0000	1.9126	0.9937	1.5217
	27	14.9840	15.3440	15.1840	15.2580
		(2.2791)	(2.9062)	(2.2548)	(2.7678)
		0.0000	2.4026	1.3348	1.8286
	30	14.6770	15.0730	14.8574	14.9640
		(2.1751)	(2.7674)	(2.0964)	(2.6382)
		0.0000	2.6981	1.2291	1.9554

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

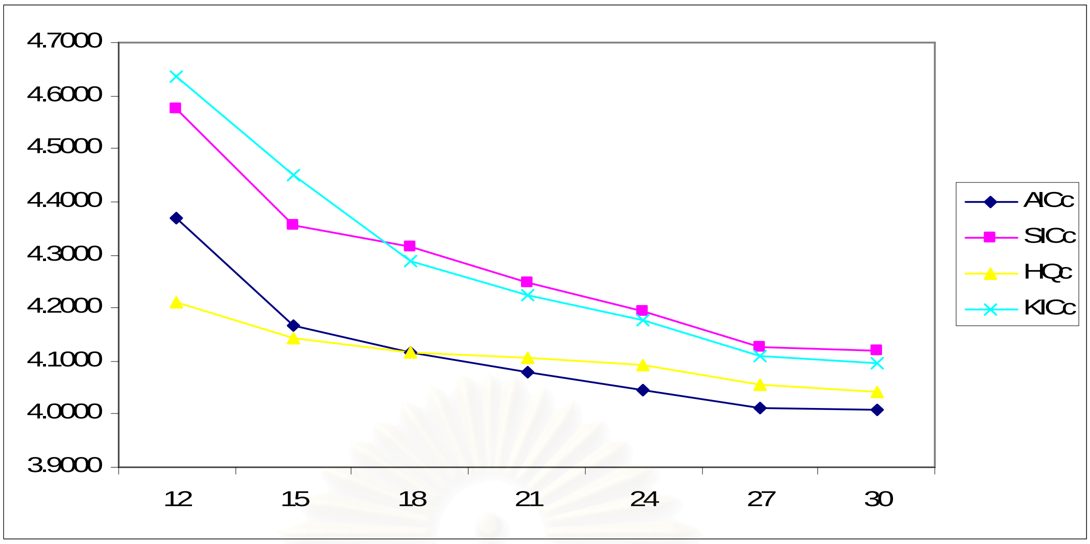
1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE



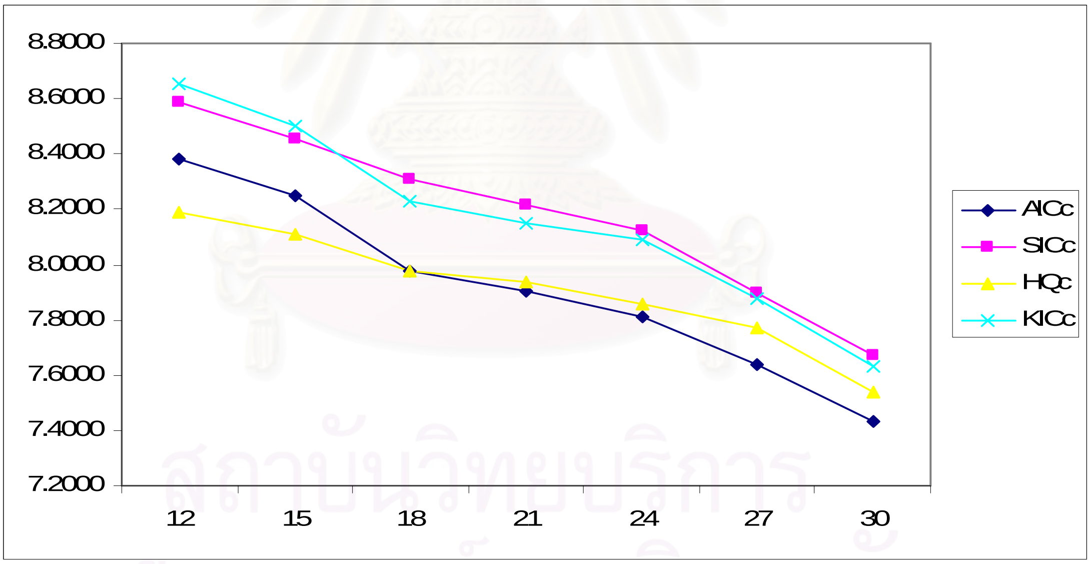
รูปที่ 4.1 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 1



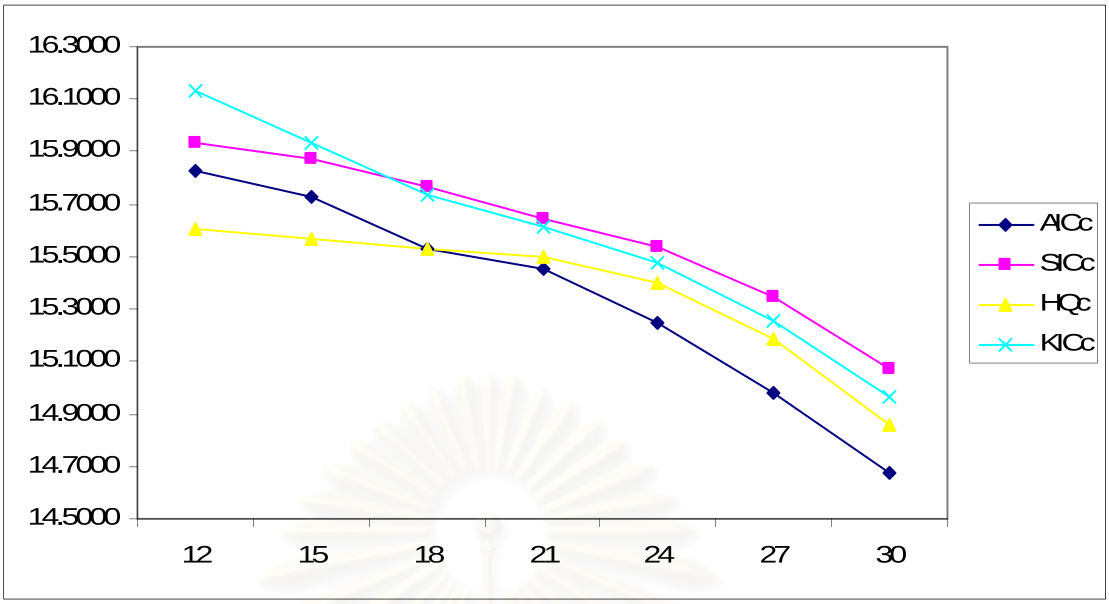
รูปที่ 4.2 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 2



รูปที่ 4.3 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 4



รูปที่ 4.4 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 8



รูปที่ 4.5 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 16

จากผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 ตัวแปร เมื่อความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เท่ากับ 1 , 2 , 4 ,8 และ 16 (ตารางที่ 4.1 – 4.5 และรูปที่ 4.1 – 4.5) พบว่า ค่า AMSE ของแต่ละวิธีเมื่อเรียงลำดับจากน้อยไปมากขึ้นอยู่กับขนาดตัวอย่าง ดังรายละเอียดต่อไปนี้

เมื่อขนาดตัวอย่างเท่ากับ 12 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ HQ_C , AIC_C ,SIC_C และ KIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยที่ค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้น เมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 15 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ HQ_C , AIC_C ,SIC_C และ KIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 18 พบว่าค่า AMSE ของเกณฑ์ HQ_C และเกณฑ์ AIC_C มีค่าเท่ากันและเป็นค่าที่น้อยที่สุด สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) การที่เกณฑ์ HQ_C และเกณฑ์ AIC_C มีค่าเท่ากันเนื่องจากทั้งสองเกณฑ์มี penalty function ที่คล้ายกันทำให้ประสิทธิภาพในการคัดเลือกตัวแบบใกล้เคียงกันและเมื่อขนาดตัวอย่างเท่ากับ 18 จะเป็นตำแหน่งที่ค่าใกล้เคียงกันมากที่สุดจึงทำให้การคัดเลือกได้ตัวแบบเดียวกันหรือใกล้เคียงกันมาก ส่วนค่า AMSE ของเกณฑ์ KIC_C จะน้อยกว่าค่า AMSE ของเกณฑ์ SIC_C เพียงเล็กน้อยสำหรับทุกความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) และเมื่อพิจารณาค่า AMSE ของแต่ละเกณฑ์พบว่า มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 21 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 24 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 27 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 30 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

สรุปคือ เมื่อขนาดตัวอย่างเท่ากับ 12 ถึง 15 เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพมากที่สุดคือ เกณฑ์ HQ_C รองลงมาคือ เกณฑ์ AIC_C ,SIC_C และ KIC_C ตามลำดับ เมื่อขนาดตัวอย่างเท่ากับ 18 เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพดีที่สุดมีสองเกณฑ์คือ เกณฑ์ HQ_C และเกณฑ์ AIC_C และรองลงมาคือเกณฑ์ KIC_C และ SIC_C ซึ่งมีประสิทธิภาพใกล้เคียงกัน เมื่อขนาดตัวอย่างเท่ากับ 21 ถึง 30 เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพมากที่สุดคือ เกณฑ์ AIC_C รองลงมาคือ เกณฑ์ HQ_C , KIC_C และ SIC_C ตามลำดับ

เมื่อพิจารณาค่า RDAMSE เพื่อบ่งบอกประสิทธิภาพของเกณฑ์แต่ละเกณฑ์พบว่า

เมื่อขนาดตัวอย่างเท่ากับ 12 เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C และ SIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 3.38% - 28.24% และเกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 2.12% - 17.34% ส่วนเกณฑ์เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่า AIC_C ไม่มากนักคือประมาณ 1.44% - 7.94% โดยประสิทธิภาพของเกณฑ์ HQ_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆเมื่อความแปรปรวนเพิ่มขึ้น

เมื่อขนาดตัวอย่างเท่ากับ 15 เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C และ SIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 2.36% -

19.49% และเกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 1.98% - 13.04% ส่วนเกณฑ์เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่า AIC_C ไม่มากนักคือประมาณ 1.04% - 3.09% โดยประสิทธิภาพของเกณฑ์ HQ_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 18</u> เกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C มากพอสมควร กล่าวคือ เกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 1.48% - 8.93% และเกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 1.29% - 6.73% โดยประสิทธิภาพของเกณฑ์ HQ_C และ AIC_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 21</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C ดังนี้ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 1.25% - 6.44% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 1.05% - 4.11% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.3% - 1.16% โดยประสิทธิภาพของเกณฑ์ AIC_C จะแตกต่างจากเกณฑ์อื่นๆมากที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 24</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C ดังนี้ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 1.9% - 3.94% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 1.52% - 3.65% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.57% - 1.19% โดยประสิทธิภาพของเกณฑ์ AIC_C จะแตกต่างจากเกณฑ์อื่นๆมากที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 27</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C น้อยลง กล่าวคือเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 1.83% - 3.34% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 0.84% - 3.11% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยมากคือประมาณ 0.38% - 1.68%

<u>เมื่อขนาดตัวอย่างเท่ากับ 30</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C , KIC_C และ HQ_C น้อยลงดังนี้ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 1.31% - 3.22% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 0.67% - 2.62% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.18% - 1.35%

สรุปคือ <u>ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบจะแบ่งออกเป็นสองกลุ่มอย่างเห็นได้ชัด คือ กลุ่มที่ 1 ได้แก่ เกณฑ์ HQ_C และเกณฑ์ AIC_C มีประสิทธิภาพ</u>

ใกล้เคียงกันและดีที่สุดสำหรับทุกขนาดตัวอย่างและความแปรปรวน ส่วนกลุ่มที่ 2 ได้แก่ เกณฑ์ KIC_C และ SIC_C ซึ่งมีประสิทธิภาพใกล้เคียงกันแต่ต่ำกว่ากลุ่มแรกเสมอ โดยประสิทธิภาพในการคัดเลือกตัวแบบของกลุ่มที่ 1 จะดีกว่ากลุ่มที่ 2 ประมาณ 1% - 30 % ซึ่งประสิทธิภาพในการคัดเลือกตัวแบบของทั้งสองกลุ่มจะแตกต่างกันมากเมื่อขนาดตัวอย่าง 12 ถึง 18 และประสิทธิภาพจะใกล้เคียงกันขึ้นเรื่อยเมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะการเพิ่มขนาดตัวอย่างจะทำให้ทุกเกณฑ์มีความแม่นยำในการคัดเลือกตัวแบบสูงขึ้น นอกจากนี้พบว่าประสิทธิภาพในการคัดเลือกตัวแบบของทั้งสองกลุ่มจะแตกต่างกันมากเมื่อแปรปรวนน้อยคือ ความแปรปรวนเป็น 1 และความแปรปรวนเป็น 2 และประสิทธิภาพจะใกล้เคียงกันมากขึ้นเมื่อความแปรปรวนเพิ่มขึ้นเพราะเมื่อความแปรปรวนมากขึ้น จะส่งผลให้ข้อมูลมีการกระจายสูงทำให้ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์แต่ละเกณฑ์ลดน้อยลง

จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า AMSE ได้แก่ ขนาดตัวอย่างและความแปรปรวนของค่าความคลาดเคลื่อน โดยที่ค่า AMSE จะแปรผกผันกับขนาดตัวอย่าง กล่าวคือการเพิ่มขนาดตัวอย่างสำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) จะส่งผลให้ค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความเบี่ยงเบนที่ไม่ทราบสาเหตุลงได้ แต่ค่า AMSE จะแปรผันตามกับความแปรปรวนของค่าความคลาดเคลื่อน กล่าวคือเมื่อความแปรปรวนเพิ่มขึ้นส่งผลให้ค่า AMSE ของแต่ละเกณฑ์จะมีแนวโน้มเพิ่มขึ้นเนื่องจากเพราะค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

ส่วนอัตราการลดลงของค่า AMSE ของแต่ละเกณฑ์จะไม่แตกต่างกันมากนัก โดยเมื่อพิจารณาจะพบว่า อัตราการลดลงของเกณฑ์ HQ_C และ SIC_C ก็จะเป็นในลักษณะเดียวกันเนื่องจากเป็นเกณฑ์ที่มีคุณสมบัติความคงเส้นคงวาเหมือนกัน ส่วนอัตราการลดลงของเกณฑ์ AIC_C และ KIC_C จะเป็นในลักษณะเดียวกันเนื่องจากทั้งสองเกณฑ์ไม่มีคุณสมบัติความคงเส้นคงวา และเมื่อพิจารณาอัตราการลดลงในแต่ละระดับความแปรปรวนพบว่าเมื่อความแปรปรวนมีค่าน้อยคือที่ระดับความแปรปรวนเท่ากับ 1 และความแปรปรวนเท่ากับ 2 อัตราการลดลงของค่า AMSE ของทุกวิธีในช่วงขนาดตัวอย่าง 12 ถึง 18 จะค่อนข้างมาก และอัตราการลดลงของค่า AMSE ของทุกวิธีในช่วงขนาดตัวอย่าง 21 จนถึง 30 จะค่อนข้างน้อยจนค่าจะใกล้เคียงกันมากเมื่อขนาดตัวอย่างเพิ่มขึ้น ส่วนเมื่อความแปรปรวนปานกลางและมากคือที่ระดับความแปรปรวนเท่ากับ 4 , 8 และ 16 พบว่าอัตราการลดลงของค่า AMSE ของทุกวิธีในทุกช่วงขนาดตัวอย่างจะค่อนข้างน้อย และสม่ำเสมอแล้วมีแนวโน้มลู่เข้าสู่กันเมื่อเพิ่มขนาดตัวอย่างให้สูงขึ้นเพราะทุกเกณฑ์เป็นตัวประมาณไม่เอนเอียงเมื่อใกล้อนันต์

ตารางที่ 4.6 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 1

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
1	12	1.6224	1.8445	1.3881	2.0645
		(0.3850)	(0.4440)	(0.2765)	(0.4816)
		16.8792	32.8795	0.0000	48.7285
	15	1.3343	1.6462	1.2587	1.6916
		(0.3022)	(0.3895)	(0.2477)	(0.3938)
		6.0062	30.7857	0.0000	34.3926
	18	1.0730	1.3555	1.0730	1.2964
		(0.2054)	(0.3058)	(0.2054)	(0.2887)
		0.0000	26.3281	0.0000	20.8201
	21	1.0513	1.2335	1.0647	1.2175
		(0.1964)	(0.2759)	(1.1998)	(0.2652)
		0.0000	17.3309	1.2746	15.8090
	24	1.0176	1.1277	1.0251	1.0844
		(0.1764)	(0.2493)	(0.1842)	(0.2336)
		0.0000	10.8196	0.7370	6.5645
	27	1.0100	1.0876	1.0122	1.0494
		(0.1665)	(0.2301)	(0.1774)	(0.2208)
		0.0000	7.6832	0.2178	3.9010
	30	0.9934	1.0436	0.9959	1.0156
		(0.1616)	(0.2137)	(0.1676)	(0.2005)
		0.0000	5.0534	0.2517	2.2347

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.7 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 2

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
2	12	2.6137	2.9468	2.3815	3.0118
		(0.5601)	(0.6533)	(0.4539)	(0.6801)
		9.7502	23.7371	0.0000	26.4665
	15	2.3951	2.7183	2.2946	2.8228
		(0.5109)	(0.5950)	(0.4323)	(0.6317)
		4.3798	18.4651	0.0000	23.0193
	18	2.2647	2.6535	2.2647	2.6183
		(0.4187)	(0.5713)	(0.4187)	(0.5724)
		0.0000	17.1678	0.0000	15.6135
	21	2.1680	2.5380	2.2000	2.4180
		(0.4258)	(0.5497)	(0.4035)	(0.5107)
		0.0000	17.0664	1.4760	11.5314
	24	2.0566	2.3093	2.0935	2.1982
		(0.3509)	(0.4960)	(0.3703)	(0.4621)
		0.0000	12.2873	1.7942	6.8852
	27	2.0532	2.1914	2.0784	2.1295
		(0.3392)	(0.4538)	(0.3691)	(0.4335)
		0.0000	6.7310	1.2274	3.7162
	30	2.0131	2.1263	2.0493	2.0786
		(0.3112)	(0.4250)	(0.3441)	(0.4101)
		0.0000	5.6232	1.7982	3.2537

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.8 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 4

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
4	12	4.5387	4.8636	4.2456	4.9672
		(0.9150)	(1.0549)	(0.7485)	(1.0516)
		6.9036	14.5562	0.0000	16.9964
	15	4.3592	4.6645	4.1895	4.7624
		(0.9058)	(1.0005)	(0.7210)	(0.9892)
		4.0506	11.3379	0.0000	13.6747
	18	4.1955	4.5674	4.1955	4.5367
		(0.7082)	(0.9706)	(0.7082)	(0.9287)
		0.0000	8.8643	0.0000	8.1325
	21	4.0814	4.4223	4.1565	4.3857
		(0.6763)	(0.9190)	(0.6825)	(0.8969)
		0.0000	8.3525	1.8401	7.4565
	24	4.0656	4.3648	4.1654	4.3394
		(0.6647)	(0.8561)	(0.6594)	(0.8709)
		0.0000	7.3839	2.4547	6.7350
	27	4.0394	4.2829	4.1124	4.2361
		(0.6520)	(0.8275)	(0.6197)	(0.8290)
		0.0000	6.0281	1.8072	4.8695
	30	4.0197	4.2046	4.0885	4.1789
		(0.6182)	(0.7673)	(0.5965)	(0.7994)
		0.0000	4.5998	1.7116	3.9605

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

<u>ตารางที่ 4.9</u> การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 8

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
8	12	8.5543	8.9978	8.2334	9.1512
		(1.6724)	(1.9282)	(1.4359)	(1.9254)
		3.8975	9.2841	0.0000	11.1473
	15	8.2930	8.8532	8.1788	8.9891
		(1.5906)	(1.8078)	(1.3397)	(1.8634)
		1.3963	8.2457	0.0000	9.9073
	18	8.1563	8.7930	8.1563	8.7640
		(1.3172)	(1.7261)	(1.3172)	(1.6792)
		0.0000	7.8062	0.0000	7.4507
	21	8.0580	8.5883	8.1230	8.4966
		(1.2595)	(1.6498)	(1.2818)	(1.5863)
		0.0000	6.5810	0.8067	5.4430
	24	7.9862	8.4724	8.1026	8.3763
		(1.2474)	(1.6233)	(1.2324)	(1.5345)
		0.0000	6.0880	1.4575	4.8847
	27	7.6671	8.1123	7.8095	8.0245
		(1.1662)	(1.5365)	(1.1597)	(1.4556)
		0.0000	5.8066	1.8573	4.6615
	30	7.5946	8.0072	7.6733	7.9542
		(1.1255)	(1.4701)	(1.0827)	(1.4023)
		0.0000	5.4328	1.0363	4.7349

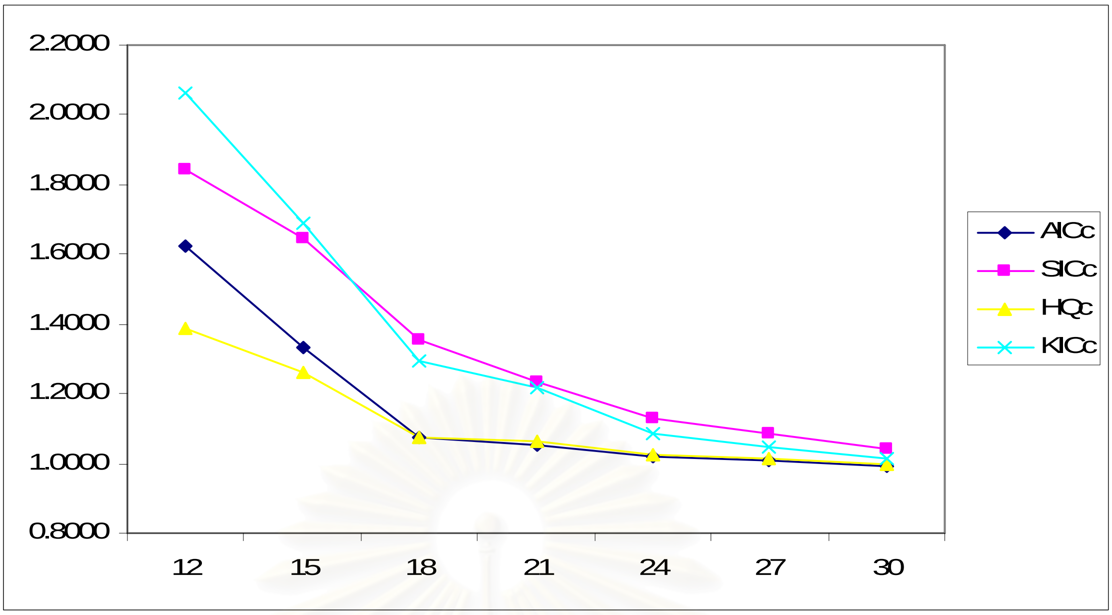
<u>หมายเหตุ</u> ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่
1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.10 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 16

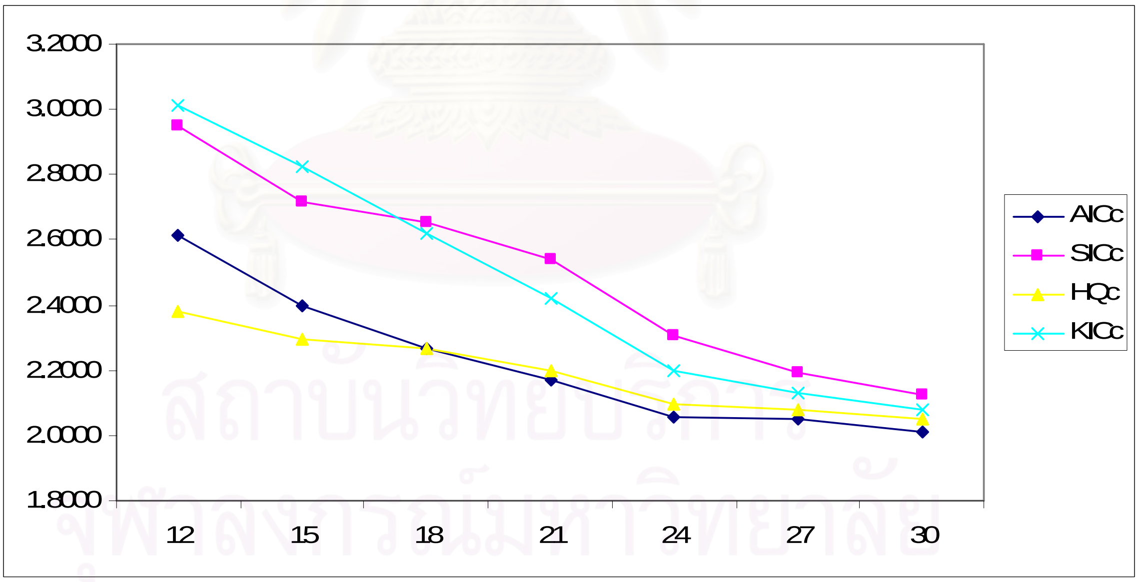
σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
16	12	15.9813	16.3446	15.7364	16.6071
		(3.1243)	(3.5026)	(2.7444)	(3.4941)
		1.5563	3.8649	0.0000	5.5330
	15	15.8454	16.2415	15.6843	16.4609
		(3.0319)	(3.3165)	(2.5691)	(3.4123)
		1.0271	3.5526	0.0000	4.9514
	18	15.5785	16.1855	15.5785	16.1726
		(2.5159)	(3.1772)	(2.5159)	(3.0987)
		0.0000	3.8964	0.0000	3.8136
	21	15.4250	15.8875	15.5020	15.7982
		(2.4109)	(3.0520)	(2.4462)	(2.9495)
		0.0000	2.9984	0.4992	2.4194
	24	15.3182	15.7546	15.4640	15.6523
		(2.3927)	(3.0186)	(2.3521)	(2.8675)
		0.0000	2.8489	0.9518	2.1811
	27	15.0320	15.5575	15.2749	15.4732
		(2.2864)	(2.9466)	(2.2683)	(2.8068)
		0.0000	3.4959	1.6159	2.9351
	30	14.7655	15.3752	14.9858	15.2416
		(2.1882)	(2.8229)	(2.1145)	(2.6871)
		0.0000	4.1292	1.4920	3.2244

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

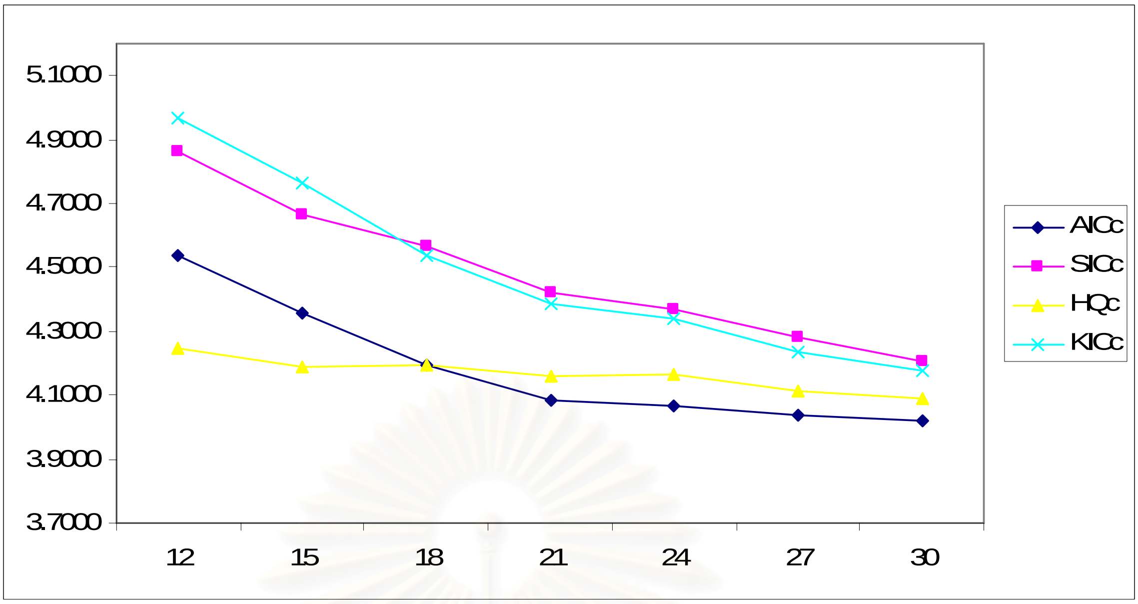
1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE



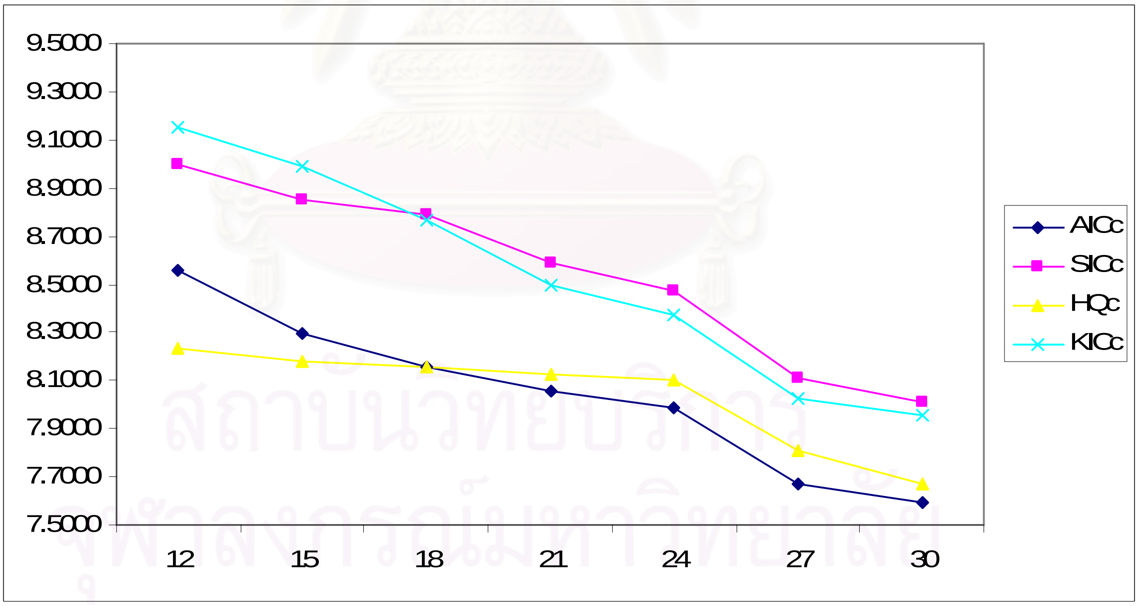
รูปที่ 4.6 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 1



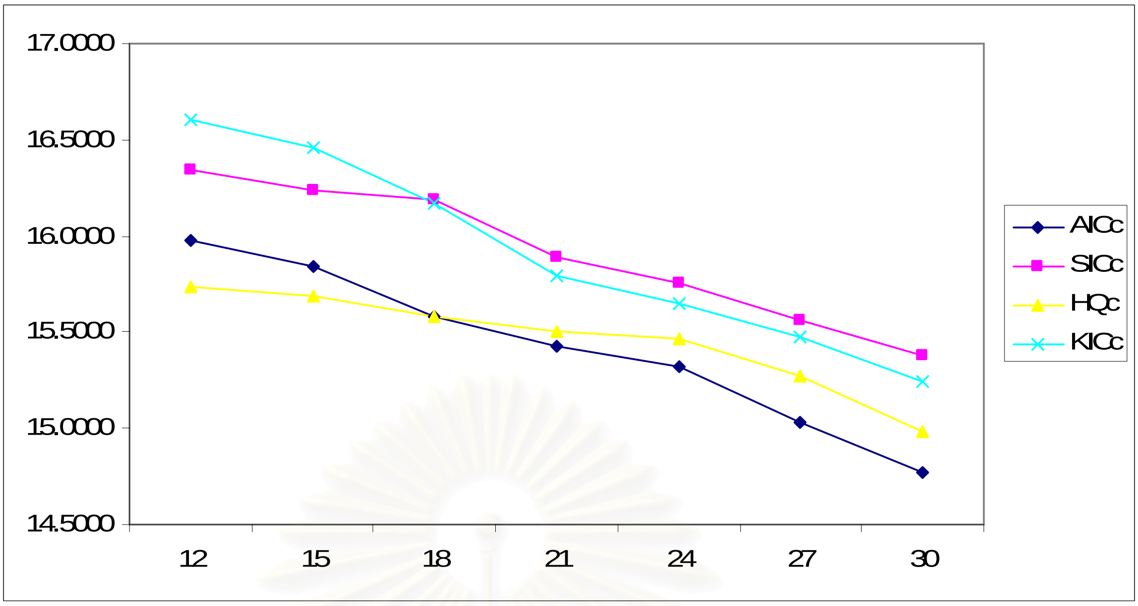
รูปที่ 4.7 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 2



<u>รูปที่ 4.8</u> แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 4



<u>รูปที่ 4.9</u> แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 8



รูปที่ 4.9 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 16

<u>จากผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 5 ตัวแปร</u> เมื่อความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เท่ากับ 1, 2, 4, 8 และ16 (ตารางที่ 4.6 – 4.10 และรูปที่ 4.6 – 4.10) พบว่า <u>ค่า AMSE ของแต่ละวิธีเมื่อเรียงลำดับจากน้อยไปมากขึ้นอยู่กับขนาดตัวอย่าง</u> ดังรายละเอียดต่อไปนี้

<u>เมื่อขนาดตัวอย่างเท่ากับ 12</u> พบว่า ค่าAMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ HQ_C , AIC_C ,SIC_C และ KIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

<u>เมื่อขนาดตัวอย่างเท่ากับ 15</u> พบว่า ค่าAMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ HQ_C , AIC_C ,SIC_C และ KIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

<u>เมื่อขนาดตัวอย่างเท่ากับ 18</u> พบว่า ค่าAMSE ของเกณฑ์ HQ_C และเกณฑ์ AIC_C มีค่าเท่ากันและเป็นค่าที่น้อยที่สุด สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) การที่เกณฑ์ HQ_C และเกณฑ์ AIC_C มีค่าเท่ากันเนื่องจากทั้งสองเกณฑ์มี penalty function ที่คล้ายกันทำให้ประสิทธิภาพในการคัดเลือกตัวแบบใกล้เคียงกันและเมื่อขนาดตัวอย่างเท่ากับ 18 จะเป็นตำแหน่งที่ค่าใกล้เคียงกันมากที่สุดจึงทำให้การคัดเลือกได้ตัวแบบเดียวกันหรือใกล้เคียงกันมาก ส่วนค่า AMSE ของเกณฑ์ KIC_C จะน้อยกว่าค่า AMSE ของเกณฑ์ SIC_C เพียงเล็กน้อยสำหรับทุกความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) และเมื่อพิจารณาค่า AMSE ของแต่ละเกณฑ์พบว่า มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

<u>เมื่อขนาดตัวอย่างเท่ากับ 21</u> พบว่า ค่าAMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

<u>เมื่อขนาดตัวอย่างเท่ากับ 24</u> พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

<u>เมื่อขนาดตัวอย่างเท่ากับ 27</u> พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

<u>เมื่อขนาดตัวอย่างเท่ากับ 30</u> พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

สรุปคือ <u>เมื่อขนาดตัวอย่างเท่ากับ 12 ถึง 15</u> เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพมากที่สุดคือ เกณฑ์ HQ_C รองลงมาคือ เกณฑ์ AIC_C ,SIC_C และ KIC_C ตามลำดับ <u>เมื่อขนาดตัวอย่างเท่ากับ 18</u> เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพดีที่สุดมีสองเกณฑ์คือ เกณฑ์ HQ_C และเกณฑ์ AIC_C และรองลงมาคือเกณฑ์ KIC_C และ SIC_C ซึ่งมีประสิทธิภาพใกล้เคียงกัน <u>เมื่อขนาดตัวอย่างเท่ากับ 21 ถึง 30</u> เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพมากที่สุดคือ เกณฑ์ AIC_C รองลงมาคือ เกณฑ์ HQ_C , KIC_C และ SIC_C ตามลำดับ

เมื่อพิจารณาค่า RDAMSE เพื่อบ่งบอกประสิทธิภาพของเกณฑ์แต่ละเกณฑ์พบว่า

<u>เมื่อขนาดตัวอย่างเท่ากับ 12</u> เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C และ SIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 5.53% - 48.72% และเกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 3.86% - 32.87% ส่วนเกณฑ์เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่า AIC_C ไม่มากนักคือประมาณ 1.55% - 16.87% โดยประสิทธิภาพของเกณฑ์ HQ_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 15</u> เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C และ SIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 4.95% -

34.39% และเกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 3.55% - 30.78% ส่วนเกณฑ์เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่า AIC_C ไม่มากนักคือประมาณ 1.02% - 6% โดยประสิทธิภาพของเกณฑ์ HQ_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 18</u> เกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 3.89% - 26.32% และเกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 3.81% - 20.82% โดยประสิทธิภาพของเกณฑ์ HQ_C และ AIC_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 21</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 3% - 17.33% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 2.41% - 15.8% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.5% - 1.84% โดยประสิทธิภาพของเกณฑ์ AIC_C จะแตกต่างจากเกณฑ์อื่นๆมากที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 24</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C พอสมควร กล่าวคือ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 2.84% - 10.81% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 2.18% - 6.56% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.73% - 2.45% โดยประสิทธิภาพของเกณฑ์ AIC_C จะแตกต่างจากเกณฑ์อื่นๆมากที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

<u>เมื่อขนาดตัวอย่างเท่ากับ 27</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C น้อยลง กล่าวคือเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 3.49% - 7.68% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 2.93% - 4.86% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยมากคือประมาณ 0.21% - 1.8%

<u>เมื่อขนาดตัวอย่างเท่ากับ 30</u> เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C , KIC_C และ HQ_C น้อยลงดังนี้ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 4.12% - 5.62% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 2.23% - 4.73% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.2% - 1.7%

สรุปคือ <u>ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบจะแบ่งออกเป็นสองกลุ่มอย่างเห็นได้ชัด คือ กลุ่มที่ 1 ได้แก่ เกณฑ์ HQ_C และเกณฑ์ AIC_C</u> จะมี

ประสิทธิภาพจะใกล้เคียงกันและดีที่สุด ส่วนกลุ่มที่ 2 ได้แก่ เกณฑ์ KIC_C และ SIC_C ซึ่งจะมีประสิทธิภาพใกล้เคียงกันแต่ต่ำกว่ากลุ่มแรก โดยประสิทธิภาพในการคัดเลือกตัวแบบของกลุ่มที่ 1 จะดีกว่ากลุ่มที่ 2 ประมาณ 2% - 50 % ซึ่งประสิทธิภาพในการคัดเลือกตัวแบบของทั้งสองกลุ่มจะแตกต่างกันมากเมื่อขนาดตัวอย่าง 12 ถึง 18 และประสิทธิภาพจะใกล้เคียงกันขึ้นเรื่อยเมื่อขนาดตัวอย่างเพิ่มขึ้น เพราะการเพิ่มขนาดตัวอย่างจะทำให้ทุกเกณฑ์มีความแม่นยำในการคัดเลือกตัวแบบสูงขึ้น นอกจากนี้พบว่าประสิทธิภาพในการคัดเลือกตัวแบบของทั้งสองกลุ่มจะแตกต่างกันมากเมื่อแปรปรวนน้อยคือ ความแปรปรวนเป็น 1 และความแปรปรวนเป็น 2 และประสิทธิภาพจะใกล้เคียงกันมากขึ้นเมื่อความแปรปรวนเพิ่มขึ้นเพราะเมื่อความแปรปรวนมากขึ้น จะส่งผลให้ข้อมูลมีการกระจายสูงทำให้ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์แต่ละเกณฑ์ลดน้อยลง

จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า AMSE ได้แก่ ขนาดตัวอย่างและความแปรปรวนของค่าความคลาดเคลื่อน โดยที่ค่า AMSE จะแปรผกผันกับขนาดตัวอย่าง กล่าวคือการเพิ่มขนาดตัวอย่างสำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) จะส่งผลให้ค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความเบี่ยงเบนที่ไม่ทราบสาเหตุลงได้ แต่ค่า AMSE จะแปรผันตามกับความแปรปรวนของค่าความคลาดเคลื่อน กล่าวคือเมื่อความแปรปรวนเพิ่มขึ้นส่งผลให้ค่า AMSE ของแต่ละเกณฑ์จะมีแนวโน้มเพิ่มขึ้นเนื่องจากเพราะค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

ส่วนอัตราการลดลงของค่า AMSE ของแต่ละเกณฑ์จะไม่แตกต่างกันมากนัก โดยเมื่อพิจารณาจะพบว่า อัตราการลดลงของเกณฑ์ HQ_C และ SIC_C ก็จะเป็นในลักษณะเดียวกันเนื่องจากเป็นเกณฑ์ที่มีคุณสมบัติความคงเส้นคงวาเหมือนกัน ส่วนอัตราการลดลงของเกณฑ์ AIC_C และ KIC_C จะเป็นในลักษณะเดียวกันเนื่องจากทั้งสองเกณฑ์ไม่มีคุณสมบัติความคงเส้นคงวา และเมื่อพิจารณาอัตราการลดลงในแต่ละระดับความแปรปรวนพบว่าเมื่อความแปรปรวนมีค่าน้อยคือที่ระดับความแปรปรวนเท่ากับ 1 และความแปรปรวนเท่ากับ 2 อัตราการลดลงของค่า AMSE ของทุกวิธีในช่วงขนาดตัวอย่าง 12 ถึง 18 จะค่อนข้างมาก และอัตราการลดลงของค่า AMSE ของทุกวิธีในช่วงขนาดตัวอย่าง 21 จนถึง 30 จะค่อนข้างน้อยจนค่าจะใกล้เคียงกันมากเมื่อขนาดตัวอย่างเพิ่มขึ้น ส่วนเมื่อความแปรปรวนปานกลางและมากคือที่ระดับความแปรปรวนเท่ากับ 4 , 8 และ 16 พบว่าอัตราการลดลงของค่า AMSE ของทุกวิธีในทุกช่วงขนาดตัวอย่างจะค่อนข้างน้อย และสม่ำเสมอแล้วมีแนวโน้มลู่เข้าสู่กันเมื่อเพิ่มขนาดตัวอย่างให้สูงขึ้นเพราะทุกเกณฑ์เป็นตัวประมาณไม่เอนเอียงเมื่อใกล้อนันต์

การเพิ่มจำนวนตัวแปรอิสระจาก 3 ตัวแปรเป็น 5 ตัวแปร จะทำให้ค่า AMSE ของทุกเกณฑ์มีแนวโน้มสูงขึ้น เนื่องจากในการวิจัยครั้งนี้ได้กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูปเมื่อมี

จำนวนตัวแปรอิสระเพิ่มขึ้นจำนวนตัวแบบที่เป็นไปได้ทั้งหมดที่นำมาพิจารณาจะมีมากขึ้นทำให้โอกาสที่จะได้ตัวแบบที่ไม่เหมาะสมก็จะมีมากขึ้นทำให้ค่า AMSE ของทุกเกณฑ์มีแนวโน้มสูงขึ้น นอกจากนี้พบว่าการเพิ่มจำนวนตัวแปรอิสระจะทำให้ประสิทธิภาพของเกณฑ์ HQ_C และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า เกณฑ์ KIC_C และ SIC_C ชัดเจนมากขึ้น

ตารางที่ 4.11 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 1

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
1	12	2.6830	3.1480	2.2120	3.5700
		(0.6367)	(0.7577)	(0.4406)	(0.8329)
		21.2929	42.3146	0.0000	61.3924
	15	2.5017	2.9953	2.1815	3.1142
		(0.5666)	(0.7087)	(0.4293)	(0.7250)
		14.6780	37.3046	0.0000	42.7550
	18	1.6867	2.2856	1.6867	2.2184
		(0.3228)	(0.5156)	(0.3228)	(0.4941)
		0.0000	35.5072	0.0000	31.5231
	21	1.0852	1.4544	1.1152	1.3436
		(0.2027)	(0.3253)	(0.2093)	(0.2926)
		0.0000	34.0214	2.7645	23.8113
	24	1.0404	1.2870	1.0586	1.1874
		(0.1803)	(0.2846)	(0.1902)	(0.2558)
		0.0000	23.7024	1.7493	14.1292
	27	1.0141	1.1746	1.0265	1.0973
		(0.1672)	(0.2485)	(0.1799)	(0.2309)
		0.0000	15.8268	1.2228	8.2043
	30	1.0113	1.0943	1.0145	1.0545
		(0.1645)	(0.2241)	(0.1707)	(0.2082)
		0.0000	8.2073	0.3164	4.2717

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.12 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 2

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
2	12	3.0990	3.4746	2.7751	3.6684
		(0.6641)	(0.7703)	(0.5289)	(0.8283)
		11.6717	25.2063	0.0000	32.1898
	15	2.8595	3.1864	2.6595	3.3869
		(0.6099)	(0.6975)	(0.5010)	(0.7580)
		7.5202	19.8120	0.0000	27.3510
	18	2.4850	2.9516	2.4850	2.9170
		(0.4595)	(0.6355)	(0.4595)	(0.6377)
		0.0000	18.7767	0.0000	17.3843
	21	2.3633	2.7215	2.4240	2.6376
		(0.4642)	(0.5895)	(0.4446)	(0.5571)
		0.0000	15.1568	2.5684	11.6067
	24	2.1765	2.4684	2.2276	2.3125
		(0.3713)	(0.5302)	(0.3941)	(0.4861)
		0.0000	13.4114	2.3478	6.2486
	27	2.0710	2.2219	2.1275	2.1869
		(0.3421)	(0.4602)	(0.3778)	(0.4472)
		0.0000	7.2863	2.7282	5.5963
	30	2.0282	2.1754	2.0775	2.1226
		(0.3136)	(0.4349)	(0.3488)	(0.4188)
		0.0000	7.2577	2.4307	4.6544

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.13 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 4

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
4	12	5.0175	5.8427	4.7653	6.2993
		(1.0115)	(1.2673)	(0.8401)	(1.3336)
		5.2924	22.6093	0.0000	32.1910
	15	4.7035	5.1516	4.6276	5.2868
		(0.9252)	(1.1050)	(0.7964)	(1.0981)
		1.6402	11.3234	0.0000	14.2450
	18	4.5762	5.0870	4.5762	5.0259
		(0.7725)	(1.0810)	(0.7725)	(1.0288)
		0.0000	11.1621	0.0000	9.8269
	21	4.4590	4.8665	4.4990	4.8396
		(0.7389)	(1.0113)	(0.7387)	(0.9897)
		0.0000	9.1388	0.8971	8.5355
	24	4.3592	4.7463	4.4185	4.6612
		(0.7127)	(0.9307)	(0.6994)	(0.9355)
		0.0000	8.8801	1.3603	6.9279
	27	4.1394	4.4849	4.2438	4.3862
		(0.6681)	(0.8665)	(0.6395)	(0.8584)
		0.0000	8.3466	2.5221	5.9722
	30	4.0560	4.3785	4.1481	4.2438
		(0.6238)	(0.7991)	(0.6052)	(0.8118)
		0.0000	7.9512	2.2707	4.6302

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.14 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 8

σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
8	12	8.6794	9.5483	8.2824	9.6865
		(1.6968)	(2.0462)	(1.4445)	(2.0380)
		4.7933	15.2842	0.0000	16.9528
	15	8.4363	8.9687	8.2108	9.1245
		(1.6181)	(1.8314)	(1.3449)	(1.8915)
		2.7464	9.2305	0.0000	11.1280
	18	8.1845	8.8384	8.1845	8.7953
		(1.3218)	(1.7350)	(1.3218)	(1.6852)
		0.0000	7.9895	0.0000	7.4629
	21	8.0836	8.6794	8.1494	8.5688
		(1.2635)	(1.6673)	(1.2860)	(1.5998)
		0.0000	7.3705	0.8140	6.0023
	24	8.0172	8.5194	8.1053	8.4274
		(1.2523)	(1.6323)	(1.2328)	(1.5439)
		0.0000	6.2640	1.0989	5.1165
	27	7.8125	8.2985	7.8967	8.1734
		(1.1883)	(1.5717)	(1.1727)	(1.4827)
		0.0000	6.2208	1.0778	4.6195
	30	7.6686	8.0872	7.7481	7.9994
		(1.1365)	(1.4848)	(1.0933)	(1.4103)
		0.0000	5.4586	1.0367	4.3137

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

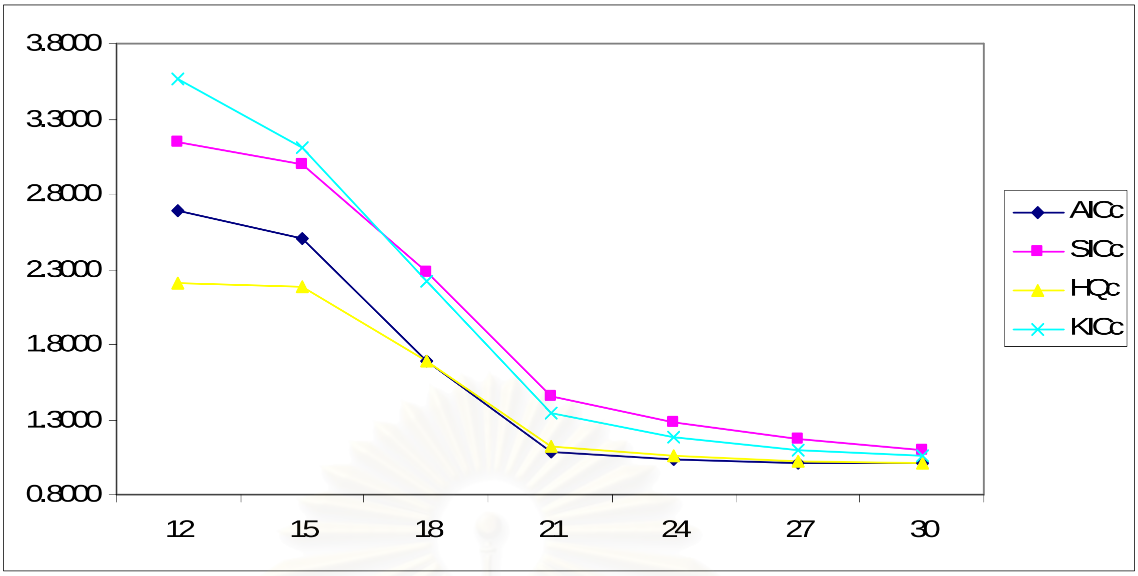
1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE

ตารางที่ 4.15 การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 16

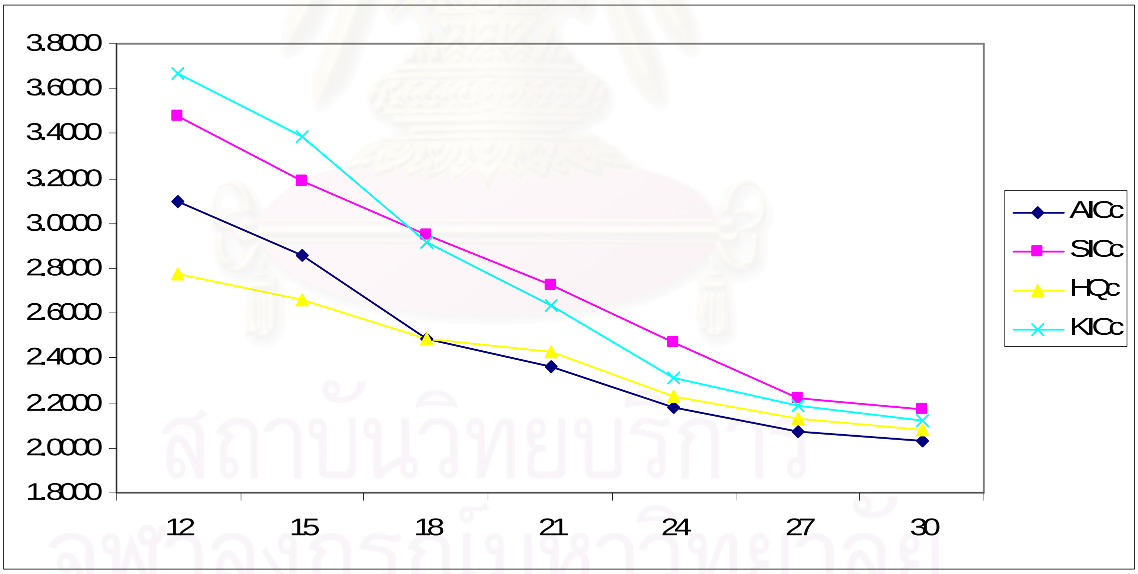
σ^2	n	วิธีการ			
		AIC_C	SIC_C	HQ_C	KIC_C
16	12	16.1542	16.9537	15.8172	17.1876
		(3.1581)	(3.6332)	(2.7585)	(3.6163)
		2.1306	7.1852	0.0000	8.6640
	15	15.9441	16.5160	15.7265	16.7775
		(3.0581)	(3.3726)	(2.5760)	(3.4780)
		1.3837	5.0202	0.0000	6.6830
	18	15.6413	16.4791	15.6413	16.3863
		(2.5261)	(3.2349)	(2.5261)	(3.1396)
		0.0000	5.3566	0.0000	4.7630
	21	15.4878	16.2633	15.5456	16.0718
		(2.4207)	(3.1242)	(2.4531)	(3.0006)
		0.0000	5.0072	0.3732	3.7707
	24	15.3982	15.9656	15.5178	15.7855
		(2.4052)	(3.0590)	(2.3603)	(2.8919)
		0.0000	3.6848	0.7767	2.5152
	27	15.1703	15.7652	15.3964	15.5656
		(2.3074)	(2.9859)	(2.2864)	(2.8236)
		0.0000	3.9215	1.4904	2.6057
	30	14.8533	15.4628	15.0572	15.3105
		(2.2013)	(2.8390)	(2.1246)	(2.6992)
		0.0000	4.1035	1.3728	3.0781

หมายเหตุ ค่าที่แสดงในแต่ละกรณีจะแสดงตัวเลขทศนิยม 4 ตำแหน่งเรียงลงมาได้แก่

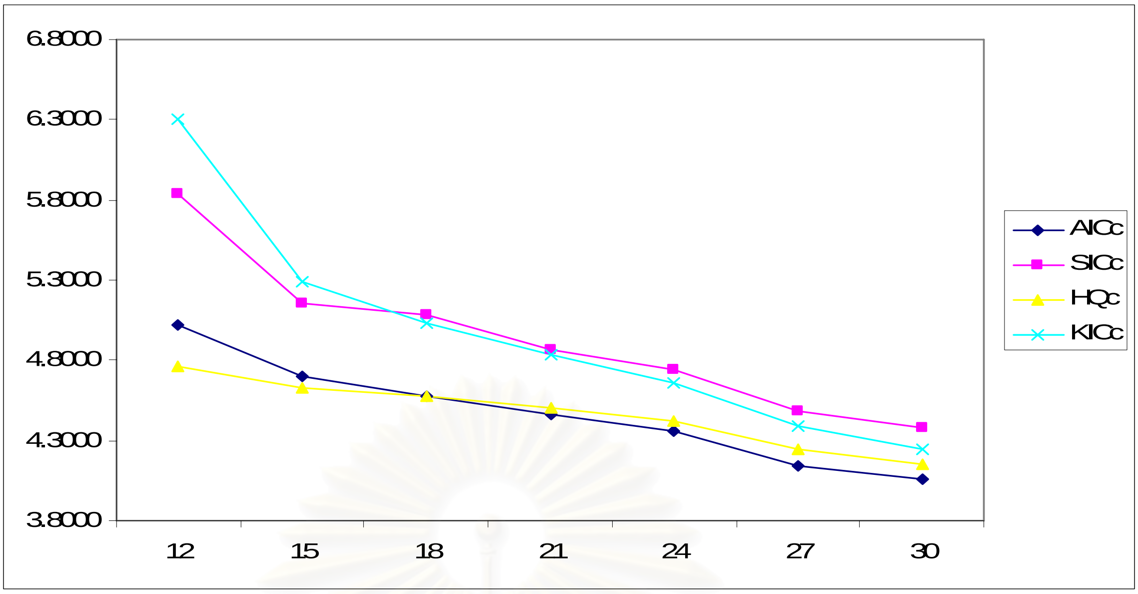
1. ค่า AMSE 2. ค่าส่วนเบี่ยงเบนมาตรฐานของค่า AMSE แสดงในวงเล็บ 3. ค่า RDAMSE



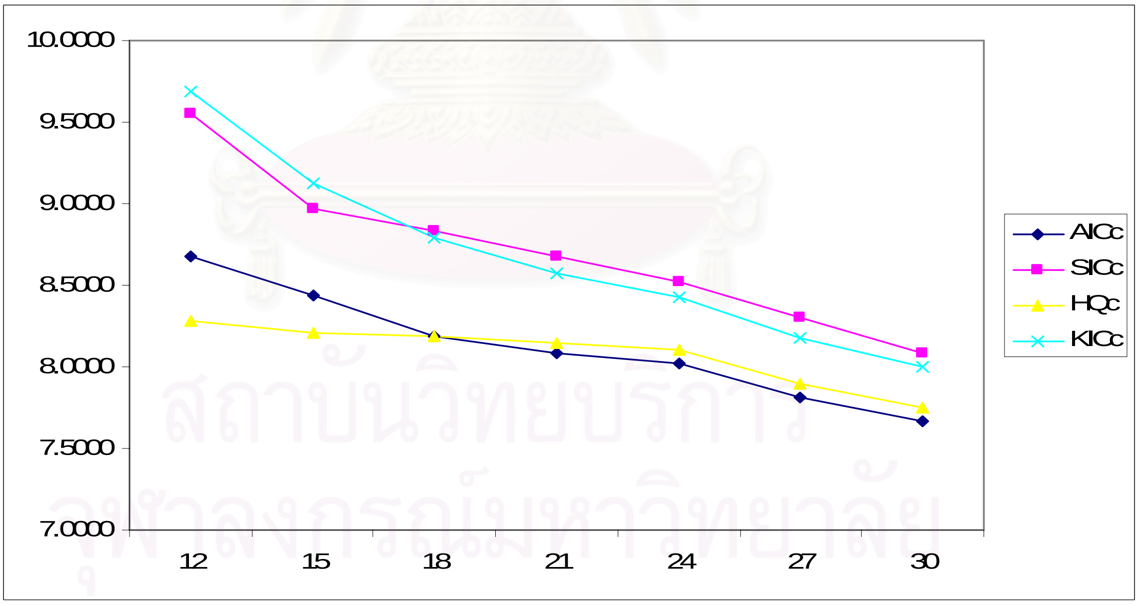
รูปที่ 4.11 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 1



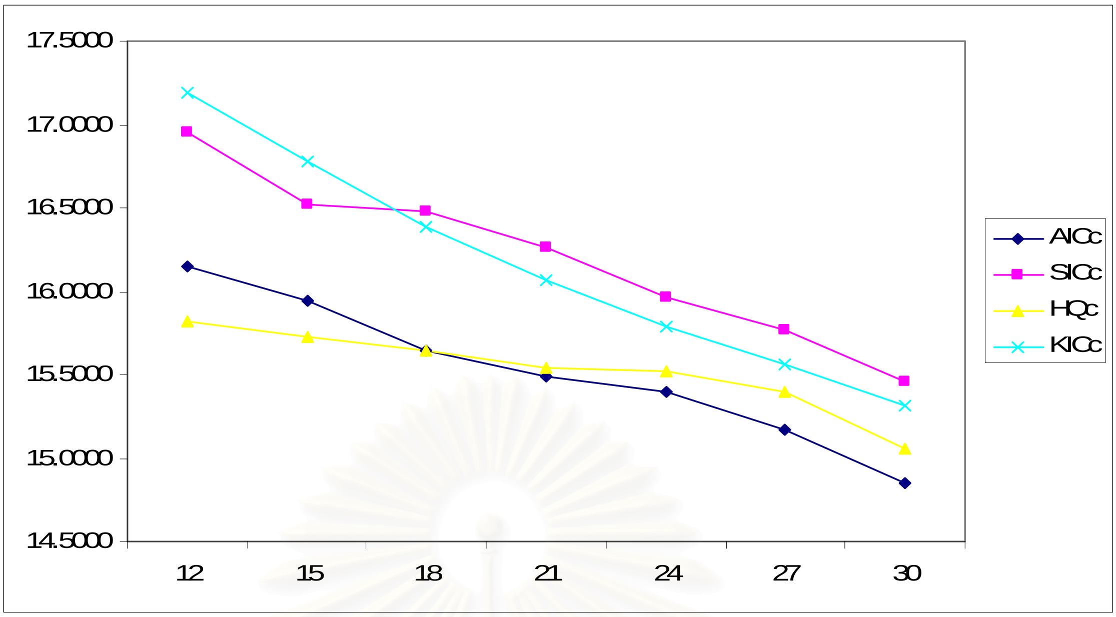
รูปที่ 4.12 แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 2



<u>รูปที่ 4.13</u> แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 4



<u>รูปที่ 4.14</u> แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 8



<u>รูปที่ 4.15</u> แสดงการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณด้วยค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2) เท่ากับ 16

จากผลการวิจัยของการเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณสำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 7 ตัวแปร เมื่อความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เท่ากับ 1, 2, 4, 8 และ16 (ตารางที่ 4.11 – 4.15 และรูปที่ 4.11 – 4.15) พบว่าค่า AMSE ของแต่ละวิธีเมื่อเรียงลำดับจากน้อยไปมากขึ้นกับอยู่กับขนาดตัวอย่าง ดังรายละเอียดต่อไปนี้

เมื่อขนาดตัวอย่างเท่ากับ 12 พบว่าค่าAMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ HQ_C , AIC_C ,SIC_C และ KIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 15 พบว่าค่าAMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ HQ_C , AIC_C ,SIC_C และ KIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 18 พบว่าค่าAMSE ของเกณฑ์ HQ_C และเกณฑ์ AIC_C มีค่าเท่ากันและเป็นค่าที่น้อยที่สุด สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) การที่เกณฑ์ HQ_C และเกณฑ์ AIC_C มีค่าเท่ากันเนื่องจากทั้งสองเกณฑ์มี penalty function ที่คล้ายกันทำให้ประสิทธิภาพในการคัดเลือกตัวแบบใกล้เคียงกันและเมื่อขนาดตัวอย่างเท่ากับ 18 จะเป็นตำแหน่งที่ค่าใกล้เคียงกันมากที่สุดจึงทำให้การคัดเลือกได้ตัวแบบเดียวกันหรือใกล้เคียงกันมาก ส่วนค่า AMSE ของเกณฑ์ KIC_C จะน้อยกว่าค่า AMSE ของเกณฑ์ SIC_C เพียงเล็กน้อยสำหรับทุกความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) และเมื่อพิจารณาค่า AMSE ของแต่ละเกณฑ์พบว่า มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 21 พบว่าค่าAMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 24 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 27 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

เมื่อขนาดตัวอย่างเท่ากับ 30 พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ได้แก่ เกณฑ์ AIC_C , HQ_C , KIC_C และ SIC_C ตามลำดับ สำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) โดยค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มเพิ่มสูงขึ้นเมื่อค่าความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)เพิ่มขึ้น ทั้งนี้เนื่องจากค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

สรุปคือ เมื่อขนาดตัวอย่างเท่ากับ 12 ถึง 15 เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพมากที่สุดคือ เกณฑ์ HQ_C รองลงมาคือ เกณฑ์ AIC_C ,SIC_C และ KIC_C ตามลำดับ เมื่อขนาดตัวอย่างเท่ากับ 18 เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพดีที่สุดมีสองเกณฑ์คือ เกณฑ์ HQ_C และเกณฑ์ AIC_C และรองลงมาคือเกณฑ์ KIC_C และ SIC_C ซึ่งมีประสิทธิภาพใกล้เคียงกัน เมื่อขนาดตัวอย่างเท่ากับ 21 ถึง 30 เกณฑ์การคัดเลือกตัวแบบที่มีประสิทธิภาพมากที่สุดคือ เกณฑ์ AIC_C รองลงมาคือ เกณฑ์ HQ_C , KIC_C และ SIC_C ตามลำดับ

เมื่อพิจารณาค่า RDAMSE เพื่อบ่งบอกประสิทธิภาพของเกณฑ์แต่ละเกณฑ์พบว่า

เมื่อขนาดตัวอย่างเท่ากับ 12 เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C และ SIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 8.66% - 61.4% และเกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 7.18% - 42.31% ส่วนเกณฑ์เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่า AIC_C มากพอสมควรคือประมาณ 2.13% - 21.29% โดยประสิทธิภาพของเกณฑ์ HQ_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

เมื่อขนาดตัวอย่างเท่ากับ 15 เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C และ SIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 6.68% -

42.75% และเกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 5.02% - 37.3% ส่วนเกณฑ์เกณฑ์ HQ_C จะมีประสิทธิภาพดีกว่า AIC_C มากพอสมควรคือประมาณ 1.38% - 14.67% โดยประสิทธิภาพของเกณฑ์ HQ_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

เมื่อขนาดตัวอย่างเท่ากับ 18 เกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 5.35% - 35.5% และเกณฑ์ HQ_C และ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 4.76% - 31.52% โดยประสิทธิภาพของเกณฑ์ HQ_C และ AIC_C จะโดดเด่นที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

เมื่อขนาดตัวอย่างเท่ากับ 21 เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 5% - 34% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 3.77% - 23.81% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.37% - 2.76% โดยประสิทธิภาพของเกณฑ์ AIC_C จะแตกต่างจากเกณฑ์อื่นๆมากที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

เมื่อขนาดตัวอย่างเท่ากับ 24 เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C อย่างเห็นได้ชัด กล่าวคือ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 3.68% - 23.7% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 2.51% - 14.12% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.77% - 2.34% โดยประสิทธิภาพของเกณฑ์ AIC_C จะแตกต่างจากเกณฑ์อื่นๆมากที่สุดเมื่อความแปรปรวนมีค่าเป็น 1 และจะลดลงเรื่อยๆ เมื่อความแปรปรวนเพิ่มขึ้น

เมื่อขนาดตัวอย่างเท่ากับ 27 เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C และ KIC_C พอสมควร กล่าวคือเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 3.92% - 15.82% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 2.6% - 8.2% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยมากคือประมาณ 1% - 2.7%

เมื่อขนาดตัวอย่างเท่ากับ 30 เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C , KIC_C น้อยลงดังนี้ เกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ SIC_C ประมาณ 4.1% - 8.2% และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่าเกณฑ์ KIC_C ประมาณ 3% - 4.65% ส่วนเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า HQ_C น้อยคือประมาณ 0.3% - 2.4%

สรุปคือ ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบจะแบ่งออกเป็นสองกลุ่มอย่างเห็นได้ชัด คือ กลุ่มที่ 1 ได้แก่ เกณฑ์ HQ_C และเกณฑ์ AIC_C มีประสิทธิภาพ

ใกล้เคียงกันและดีที่สุด ส่วนกลุ่มที่ 2 ได้แก่ เกณฑ์ KIC_C และ SIC_C ซึ่งมีประสิทธิภาพใกล้เคียงกันแต่ต่ำกว่ากลุ่มแรก โดยประสิทธิภาพในการคัดเลือกตัวแบบของกลุ่มที่ 1 จะดีกว่ากลุ่มที่ 2 ประมาณ 3% - 60 % ซึ่งประสิทธิภาพในการคัดเลือกตัวแบบของทั้งสองกลุ่มจะแตกต่างกันมากเมื่อขนาดตัวอย่าง 12 ถึง 18 และประสิทธิภาพจะใกล้เคียงกันขึ้นเรื่อยเมื่อขนาดตัวอย่างเพิ่มขึ้นเพราะการเพิ่มขนาดตัวอย่างจะทำให้ทุกเกณฑ์มีความแม่นยำในการคัดเลือกตัวแบบสูงขึ้น นอกจากนี้พบว่าประสิทธิภาพในการคัดเลือกตัวแบบของทั้งสองกลุ่มจะแตกต่างกันมากเมื่อแปรปรวนน้อยคือ ความแปรปรวนเป็น 1 และความแปรปรวนเป็น 2 และประสิทธิภาพจะใกล้เคียงกันมากขึ้นเมื่อความแปรปรวนเพิ่มขึ้นเพราะเมื่อความแปรปรวนมากขึ้น จะส่งผลให้ข้อมูลมีการกระจายสูงทำให้ประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์แต่ละเกณฑ์ลดน้อยลง

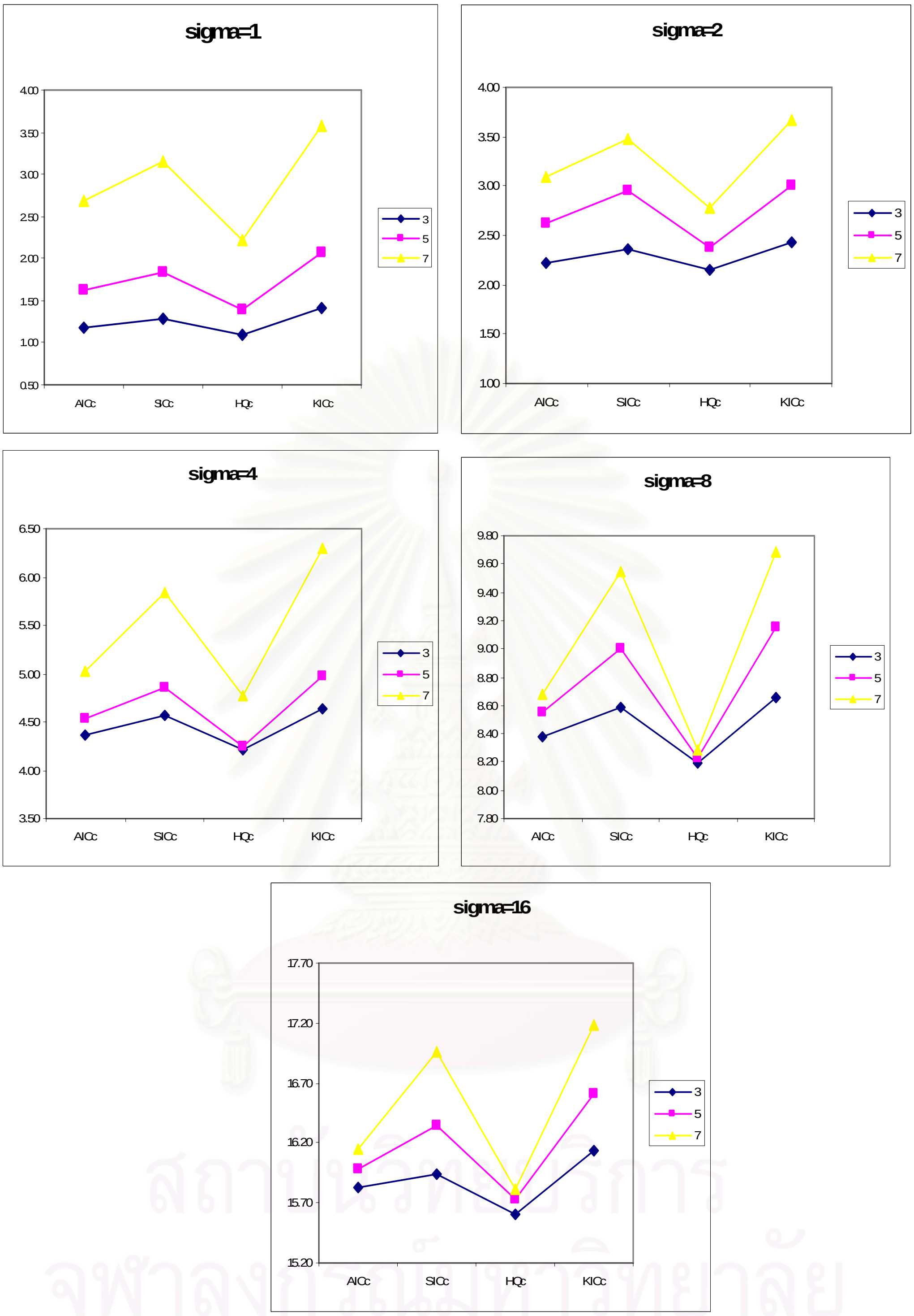
จากข้างต้นสรุปได้ว่าปัจจัยที่ส่งผลต่อค่า AMSE ได้แก่ ขนาดตัวอย่างและความแปรปรวนของค่าความคลาดเคลื่อน โดยที่ค่า AMSE จะแปรผกผันกับขนาดตัวอย่าง กล่าวคือการเพิ่มขนาดตัวอย่างสำหรับทุกๆระดับของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2) จะส่งผลให้ค่า AMSE ของแต่ละเกณฑ์มีแนวโน้มลดลงเพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความเบี่ยงเบนที่ไม่ทราบสาเหตุลงได้ แต่ค่า AMSE จะแปรผันตามกับความแปรปรวนของค่าความคลาดเคลื่อน กล่าวคือเมื่อความแปรปรวนเพิ่มขึ้นส่งผลให้ค่า AMSE ของแต่ละเกณฑ์จะมีแนวโน้มเพิ่มขึ้นเนื่องจากเพราะค่า AMSE เป็นค่าประมาณของความแปรปรวนของค่าคลาดเคลื่อนสุ่ม(σ^2)

ส่วนอัตราการลดลงของค่า AMSE ของแต่ละเกณฑ์จะไม่แตกต่างกันมากนัก โดยเมื่อพิจารณาจะพบว่า อัตราการลดลงของเกณฑ์ HQ_C และ SIC_C ก็จะเป็นในลักษณะเดียวกันเนื่องจากเป็นเกณฑ์ที่มีคุณสมบัติความคงเส้นคงวาเหมือนกัน ส่วนอัตราการลดลงของเกณฑ์ AIC_C และ KIC_C จะเป็นในลักษณะเดียวกันเนื่องจากทั้งสองเกณฑ์ไม่มีคุณสมบัติความคงเส้นคงวา และเมื่อพิจารณาอัตราการลดลงในแต่ละระดับความแปรปรวนพบว่าเมื่อความแปรปรวนมีค่าน้อยคือที่ระดับความแปรปรวนเท่ากับ 1 และความแปรปรวนเท่ากับ 2 อัตราการลดลงของค่า AMSE ของทุกวิธีในช่วงขนาดตัวอย่าง 12 ถึง 18 จะค่อนข้างมาก และอัตราการลดลงของค่า AMSE ของทุกวิธีในช่วงขนาดตัวอย่าง 21 จนถึง 30 จะค่อนข้างน้อยจนค่าจะใกล้เคียงกันมากเมื่อขนาดตัวอย่างเพิ่มขึ้น ส่วนเมื่อความแปรปรวนปานกลางและมากคือที่ระดับความแปรปรวนเท่ากับ 4 , 8 และ 16 พบว่าอัตราการลดลงของค่า AMSE ของทุกวิธีในทุกช่วงขนาดตัวอย่างจะค่อนข้างน้อย และสม่ำเสมอแล้วมีแนวโน้มลู่เข้าสู่กันเมื่อเพิ่มขนาดตัวอย่างให้สูงขึ้นเพราะทุกเกณฑ์เป็นตัวประมาณไม่เอนเอียงเมื่อใกล้อนันต์

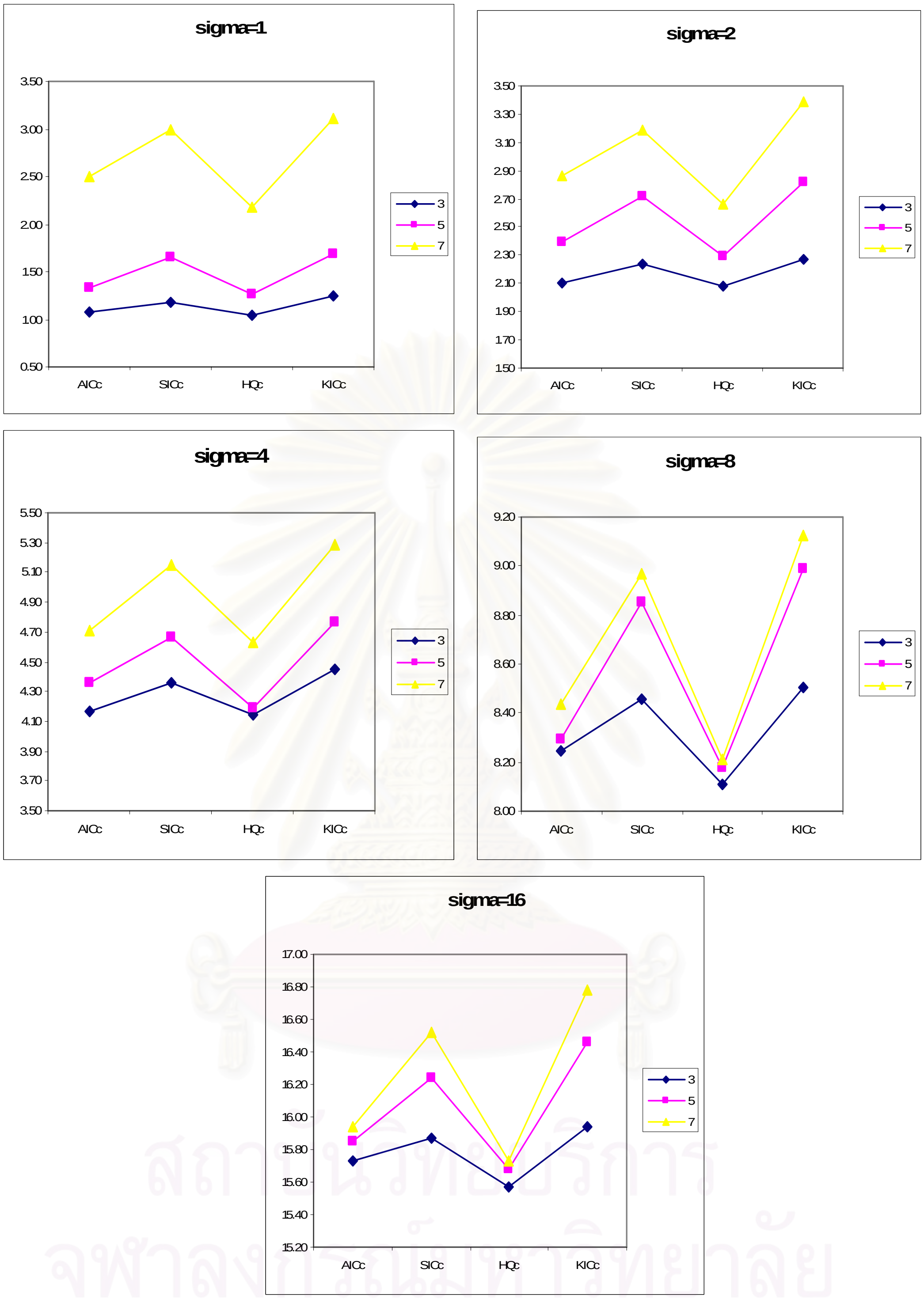
การเพิ่มจำนวนตัวแปรอิสระจาก 5 ตัวแปรเป็น 7 ตัวแปร จะทำให้ค่า AMSE ของทุกเกณฑ์มีแนวโน้มสูงขึ้น เนื่องจากในการวิจัยครั้งนี้ได้กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูปเมื่อมี

จำนวนตัวแปรอิสระเพิ่มขึ้นจำนวนตัวแบบที่เป็นไปได้ทั้งหมดที่นำมาพิจารณาจะมีมากขึ้นทำให้โอกาสที่จะได้ตัวแบบที่ไม่เหมาะสมก็จะมีมากขึ้นทำให้ค่า AMSE ของทุกเกณฑ์มีแนวโน้มสูงขึ้น นอกจากนี้พบว่าการเพิ่มจำนวนตัวแปรอิสระจะทำให้ประสิทธิภาพของเกณฑ์ HQ_C และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า เกณฑ์ KIC_C และ SIC_C ชัดเจนมากขึ้น

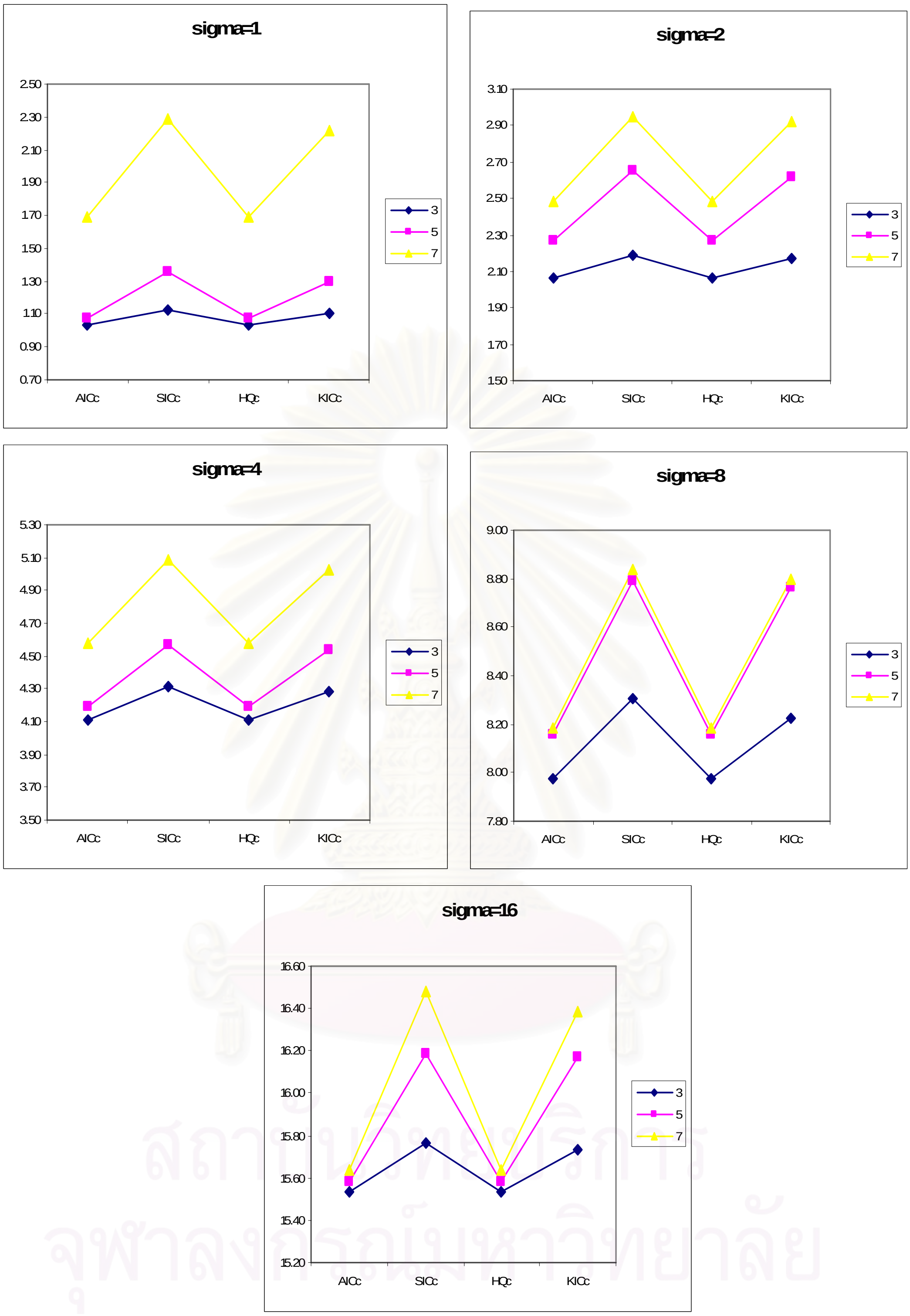
การเพิ่มขนาดตัวอย่างจาก 3 ตัวแปรเป็น 5 ตัวแปรและจาก 5 ตัวแปรเป็น 7 ตัวแปร พบว่าจะทำให้ค่า AMSE ของทุกวิธีมีแนวโน้มสูงขึ้นและประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์ HQ_C และเกณฑ์ AIC_C จะมีประสิทธิภาพดีกว่า เกณฑ์ KIC_C และ SIC_C เพิ่มมากขึ้นตามลำดับ ดังแสดงในรูปที่ 4.16 – 4.22 ในหน้าถัดไป



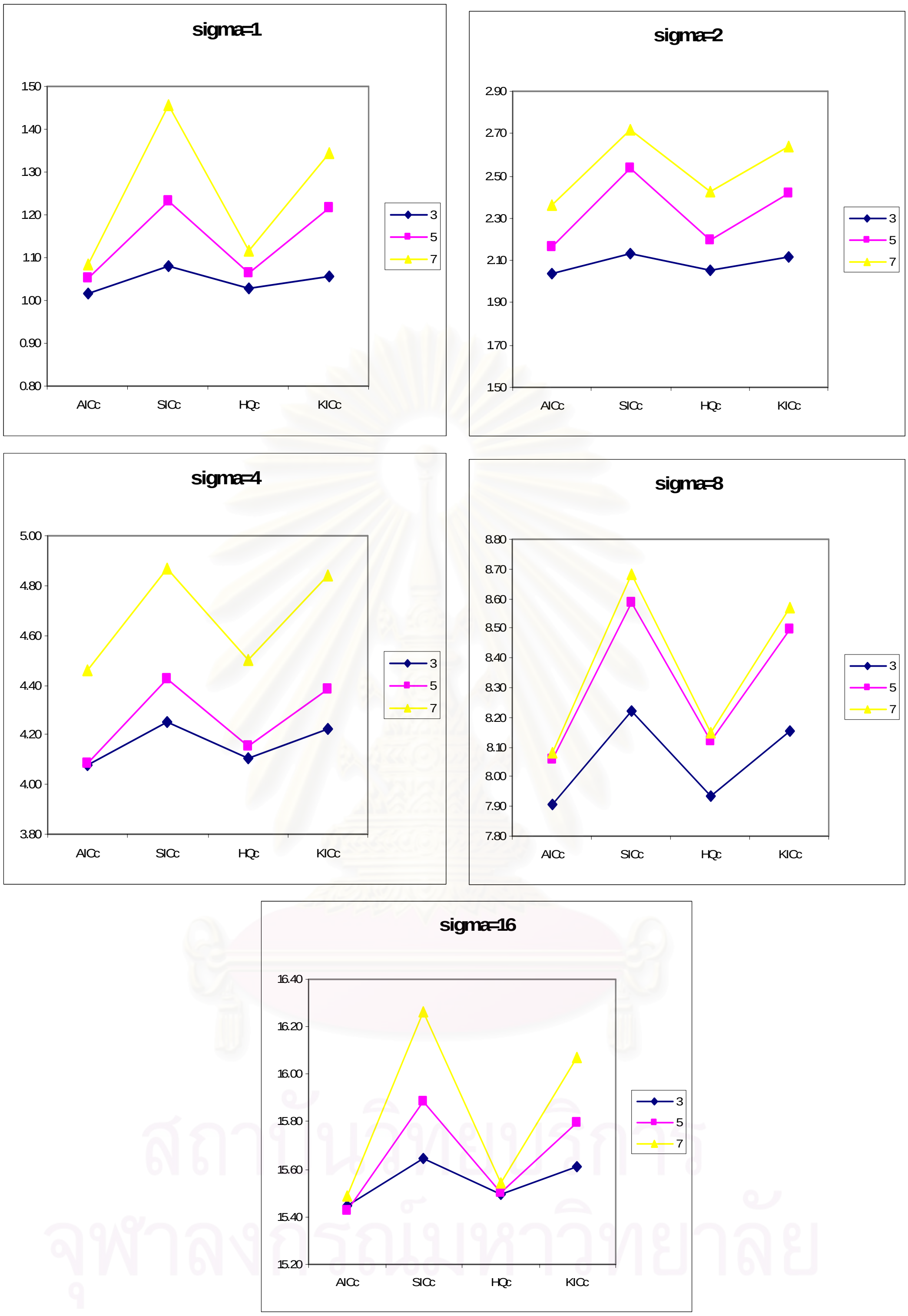
<ins>รูปที่ 4.16</ins> แสดงการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 , 5 และ 7 ตัวแปร เมื่อขนาดตัวอย่างเท่ากับ 12



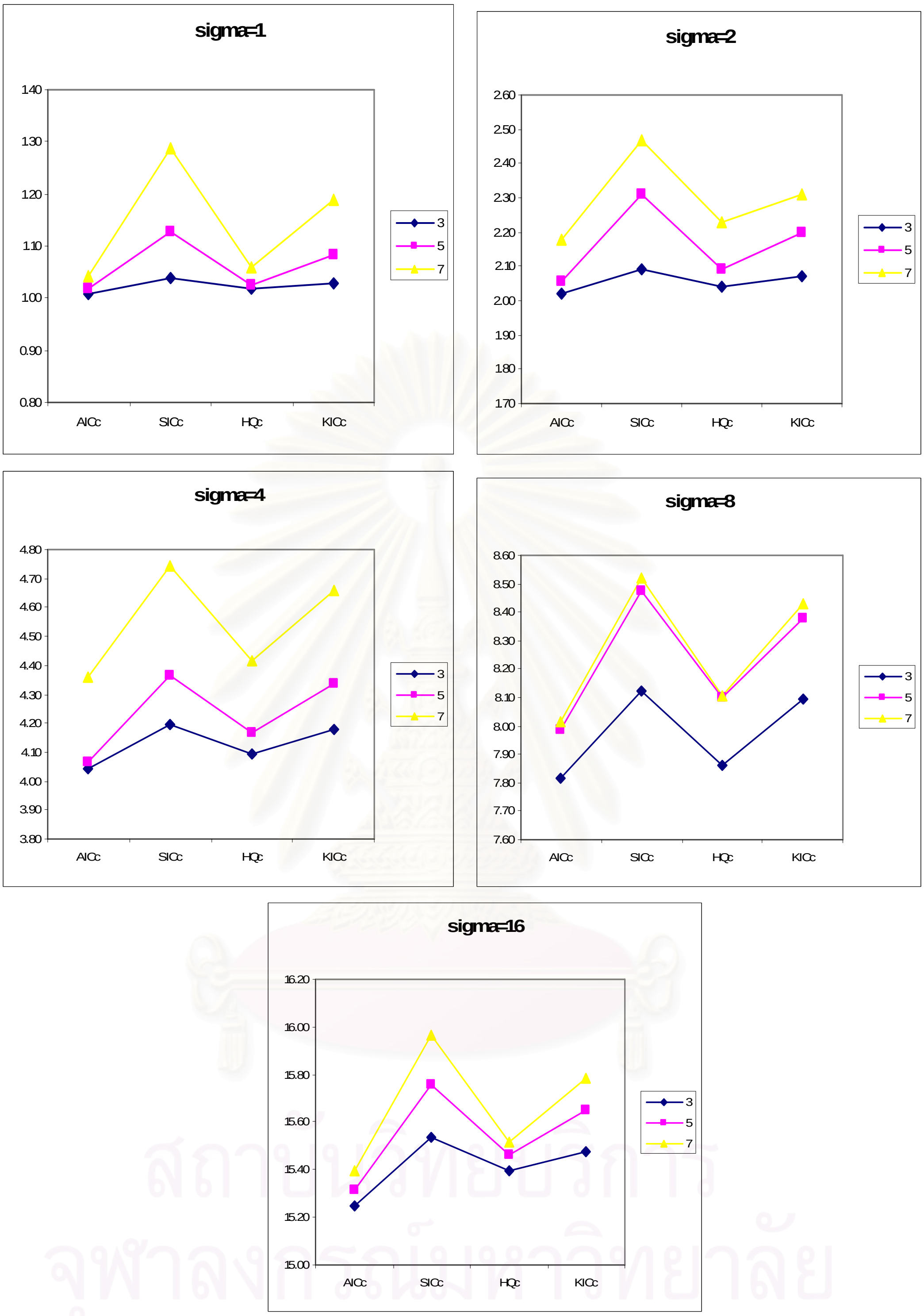
<u>รูปที่ 4.17</u>   แสดงการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 , 5 และ 7 ตัวแปร เมื่อขนาดตัวอย่างเท่ากับ 15



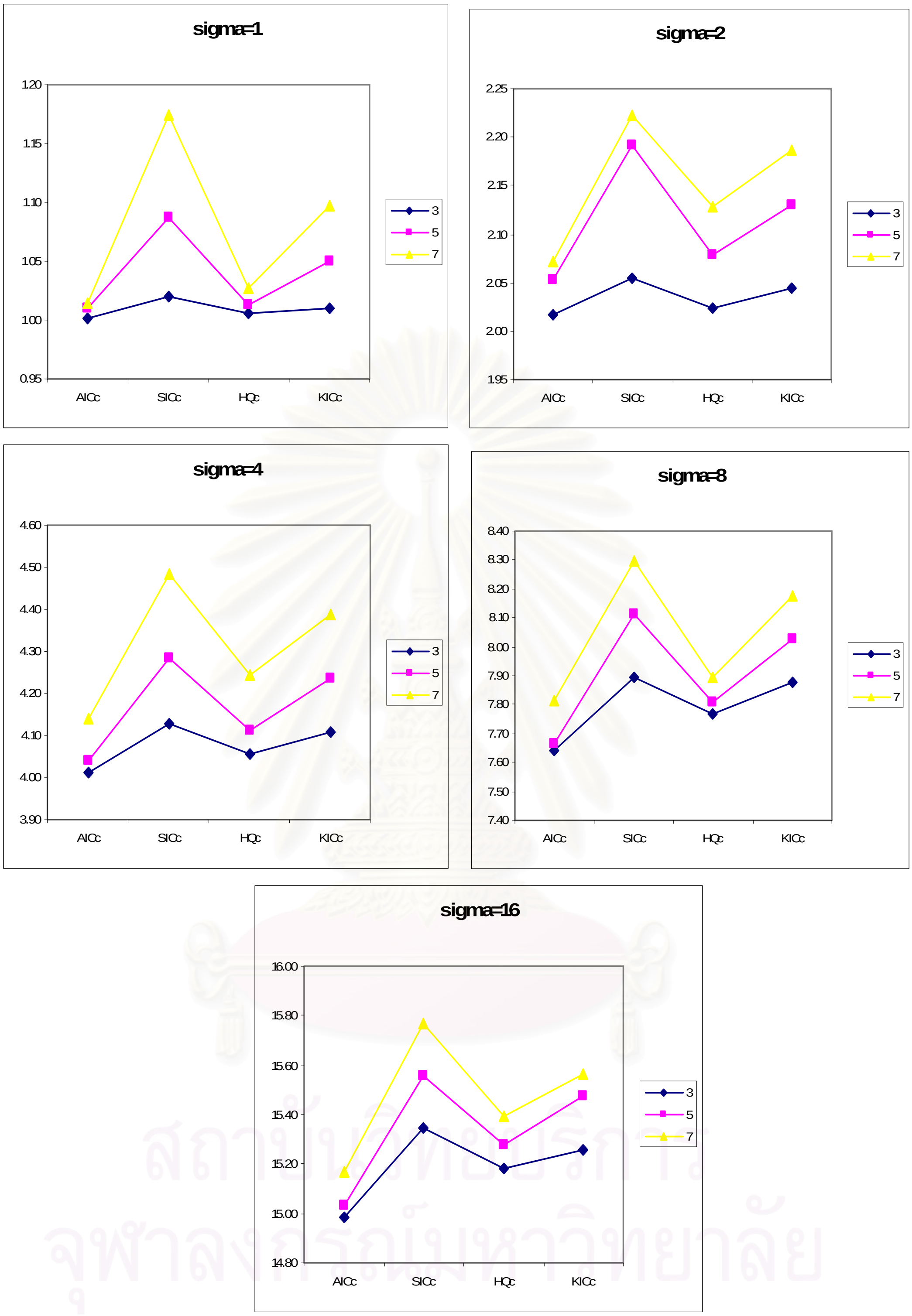
รูปที่ 4.18 แสดงการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 , 5 และ 7 ตัวแปร เมื่อขนาดตัวอย่างเท่ากับ 18



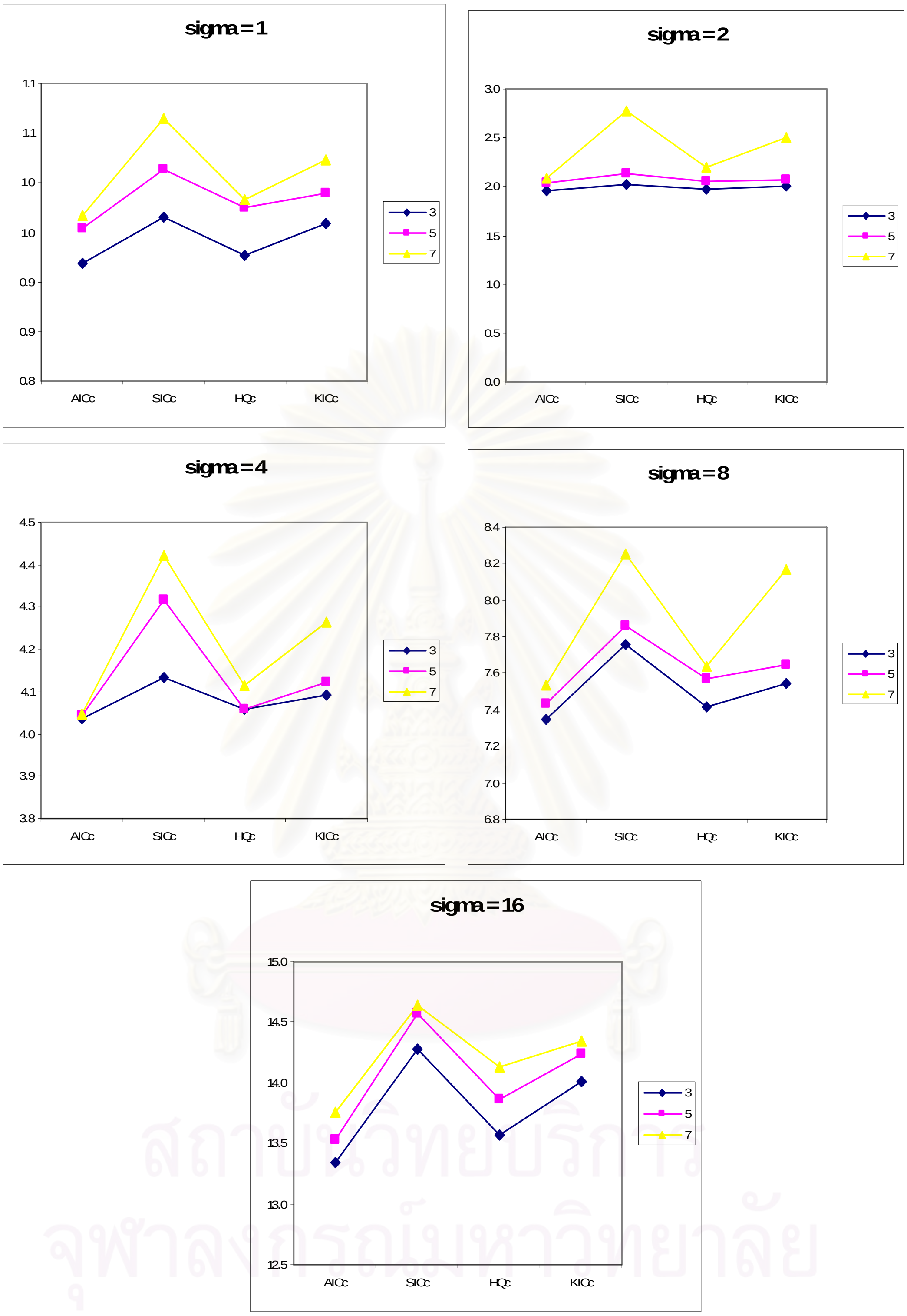
<u>รูปที่ 4.19</u> แสดงการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 , 5 และ 7 ตัวแปร เมื่อขนาดตัวอย่างเท่ากับ 21



<u>รูปที่ 4.20</u> แสดงการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 , 5 และ 7 ตัวแปร เมื่อขนาดตัวอย่างเท่ากับ 24



<u>รูปที่ 4.21</u> แสดงการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 , 5 และ 7 ตัวแปร เมื่อขนาดตัวอย่างเท่ากับ 27



<u>รูปที่ 4.22</u> แสดงการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (AMSE) สำหรับตัวแบบการถดถอยที่ประกอบด้วยตัวแปรอิสระ 3 , 5 และ 7 ตัวแปร เมื่อขนาดตัวอย่างเท่ากับ 30

บทที่ 5

สรุปผลการวิจัยและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบความถูกต้องของการพยากรณ์ที่ได้จากเกณฑ์การคัดเลือกตัวแบบการถดถอยในการวิเคราะห์ความถดถอยเชิงเส้นพหุคูณเมื่อตัวอย่างมีขนาดเล็ก เพราะเมื่อตัวอย่างมีขนาดเล็กความถูกต้องของการพยากรณ์ที่ได้จากเกณฑ์การคัดเลือกตัวแบบที่นิยมใช้โดยทั่วไป เช่น AIC หรือ BIC จะมีความผิดพลาดสูง ดังนั้นจึงควรที่จะเลือกใช้เกณฑ์การคัดเลือกตัวแบบที่เหมาะสมยิ่งขึ้น ซึ่งเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณที่ปรับแก้เพื่อใช้ในกรณีที่ขนาดตัวอย่างเล็กที่นำมาพิจารณามี 4 วิธีดังนี้

1) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของอาไคเคะที่ปรับแก้ (AIC_C)
2) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของชวาร์ซที่ปรับแก้ (SIC_C)
3) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของแฮนแนนและควินน์ที่ปรับแก้(HQ_C)
4) เกณฑ์การคัดเลือกตัวแบบโดยใช้ข้อสนเทศของคูลส์แบล็คที่ปรับแก้ (KIC_C)

ในการศึกษาและเปรียบเทียบว่าเกณฑ์การคัดเลือกตัวแบบใดจะให้ค่าพยากรณ์ที่ถูกต้องและเหมาะสมที่สุดสำหรับตัวอย่างขนาดเล็ก ได้มีการกำหนดสถานการณ์ต่างๆ ในการวิจัยครั้งนี้ไว้ดังนี้

1) การแจกแจงของความคลาดเคลื่อนสุ่ม (ε) เป็นแบบปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 , 2 , 4 , 8 และ 16
2) ขนาดตัวอย่างที่ศึกษาคือ 12 , 15 , 18 , 21 , 24 , 27 และ 30
3) จำนวนตัวแปรอิสระที่ศึกษา คือ 3 , 5 และ 7 ตัวแปร

เกณฑ์ที่ใช้ในการตัดสินใจว่าวิธีการคัดเลือกตัวแบบการถดถอยวิธีใดจะมีความถูกต้องและแม่นยำมากที่สุดพิจารณาจากเกณฑ์ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย(Average of Mean Squares Error (AMSE)) ซึ่งวิธีการคัดเลือกตัวแบบใดให้ค่าคลาดเคลื่อนกำลังสองเฉลี่ยต่ำสุด จะเป็นวิธีการคัดเลือกตัวแบบที่ดีที่สุด และใช้อัตราส่วนผลต่างของค่าเฉลี่ยของค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Squares Error (RDAMSE)) ในการประกอบการตัดสินใจเกี่ยวกับประสิทธิภาพของเกณฑ์ทั้ง 4 เกณฑ์ ผลการวิจัยได้ข้อสรุปดังนี้

5.1 สรุปผลการวิจัย

5.1.1 ผลการเปรียบเทียบค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย

จากการเปรียบเทียบค่า AMSE ของทั้ง 4 เกณฑ์ของการคัดเลือกตัวแบบการถดถอย พบว่าค่า AMSE ของแต่ละเกณฑ์เรียงลำดับจากน้อยไปมาก ขึ้นอยู่กับขนาดตัวอย่างดังนี้

ก) เมื่อขนาดตัวอย่างเท่ากับ 12 ถึง 15 เกณฑ์การคัดเลือกตัวแบบที่ให้ค่า AMSE น้อยที่สุดคือ เกณฑ์ HQ_C รองลงมาคือ เกณฑ์ AIC_C ,SIC_C และ KIC_C ตามลำดับ สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่ม และจำนวนตัวแปรอิสระ

ข) เมื่อขนาดตัวอย่างเท่ากับ 18 เกณฑ์การคัดเลือกตัวแบบที่ค่า AMSE น้อยที่สุดมีสองเกณฑ์คือ เกณฑ์ HQ_C และเกณฑ์ AIC_C และรองลงมาคือเกณฑ์ KIC_C และ SIC_C ซึ่งมีค่า AMSE ใกล้เคียงกัน สำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่ม และจำนวนตัวแปรอิสระ

ค) เมื่อขนาดตัวอย่างเท่ากับ 21 ถึง 30 เกณฑ์การคัดเลือกตัวแบบที่ให้ค่า AMSE น้อยที่สุดคือ เกณฑ์ AIC_C รองลงมาคือ เกณฑ์ HQ_C , KIC_C และ SIC_C ตามลำดับสำหรับทุกระดับความแปรปรวนของความคลาดเคลื่อนสุ่ม และจำนวนตัวแปรอิสระ

และสังเกตได้ว่าประสิทธิภาพในการคัดเลือกตัวแบบของเกณฑ์การคัดเลือกตัวแบบจะแบ่งออกเป็นสองกลุ่มอย่างเห็นได้ชัด คือ กลุ่มที่ 1 ได้แก่ เกณฑ์ HQ_C และเกณฑ์ AIC_C มีประสิทธิภาพใกล้เคียงกันและดีที่สุดสำหรับทุกขนาดตัวอย่าง และความแปรปรวนของความคลาดเคลื่อนสุ่ม ส่วนกลุ่มที่ 2 ได้แก่ เกณฑ์ KIC_C และ SIC_C ซึ่งมีประสิทธิภาพใกล้เคียงกันแต่ต่ำกว่ากลุ่มแรกสำหรับทุกขนาดตัวอย่าง และความแปรปรวนของความคลาดเคลื่อนสุ่ม

จากผลการวิจัยที่พบว่า เกณฑ์ AIC_C เป็นเกณฑ์การคัดเลือกตัวแบบที่เหมาะสมกับตัวอย่างขนาดเล็กมากที่สุดซึ่งไม่เป็นไปตามสมมติฐานของการวิจัยที่คาดว่า KIC_C น่าจะเป็นเกณฑ์ที่ดีที่สุด ทั้งนี้เพราะ การวิจัยครั้งนี้มีขอบเขตการวิจัยเมื่อตัวแบบที่นำมาพิจารณาเป็นแบบติดกลุ่ม โดยกำหนดให้ตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูป และตัวแบบที่นำมาพิจารณาเป็นตัวแบบลดรูปของตัวแบบเริ่มต้น ซึ่งจะไม่มีโอกาสได้ตัวแบบที่มีจำนวนพารามิเตอร์สูงกว่าตัวแบบจริง ทำให้ผลการวิจัยพบว่าเกณฑ์ AIC_C เป็นเกณฑ์ที่ให้ค่าพยากรณ์ที่ถูกต้องและเหมาะสมที่สุด ทั้งนี้เนื่องจากเกณฑ์ AIC_C เป็นเกณฑ์ที่ไม่มีความเอนเอียงสำหรับการได้จำนวนพารามิเตอร์น้อยกว่าจำนวนพารามิเตอร์จริง (underfitting estimation) ซึ่งได้เปรียบเกณฑ์ KIC_C เป็นเกณฑ์ที่มีความเอนเอียง[1] นี้อยู่มากกว่า และจากการศึกษาพบว่าเกณฑ์ AIC_C เป็นตัวประมาณที่ไม่เอนเอียงซึ่งมีความแปรปรวนต่ำสุดของข้อสนเทศของคลูแบล็ค-ไลท์เบลอร์ ด้วย

---

[1]ความเอนเอียงของเกณฑ์การคัดเลือกตัวแบบการถดถอยคือ ความไม่แม่นยำในการคัดเลือกให้ได้ตัวแบบการถดถอยที่แท้จริงซึ่งความเอนเอียงที่เกิดขึ้นมีสองลักษณะคือ การคัดเลือกตัวแบบการถดถอยที่มีจำนวนพารามิเตอร์มากกว่าจำนวนพารามิเตอร์ในตัวแบบที่แท้จริง(overfitting) และการคัดเลือกตัวแบบการถดถอยที่มีจำนวนพารามิเตอร์น้อยกว่าจำนวนพารามิเตอร์ในตัวแบบที่แท้จริง(underfitting)

5.1.2 ปัจจัยที่มีผลต่อค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ยของแต่ละวิธี

1. ขนาดตัวอย่าง (n)

ค่า AMSE มีความสัมพันธ์แบบแปรผกผันกับขนาดตัวอย่าง นั่นคือ เมื่อขนาดตัวอย่างเพิ่มขึ้น ค่า AMSE ของทุกวิธีมีแนวโน้มลดลง เพราะขนาดตัวอย่างที่เพิ่มขึ้นจะช่วยลดความเบี่ยงเบนที่ไม่ทราบสาเหตุลงได้ โดยอัตราการลดลงของค่า AMSE ของทุกวิธีจะเห็นได้ชัดเมื่อขนาดตัวอย่าง 12 ถึง 18 สำหรับความแปรปรวนที่ระดับ 1 และ 2 และเมื่อขนาดตัวอย่างเพิ่มขึ้นค่า AMSE ของทุกวิธีมีแนวโน้มที่ละลู่เข้าสู่ค่าเดียวกัน เนื่องจากทุกวิธีเป็นตัวประมาณไม่เอนเอียงเมื่อใกล้อนันต์

2. ความแปรปรวนของความคลาดเคลื่อนสุ่ม$\left(\sigma_{\varepsilon}^{2}\right)$

ค่า AMSE มีความสัมพันธ์แบบแปรผันตามกับความแปรปรวนของความคลาดเคลื่อนสุ่ม กล่าวคือ เมื่อความแปรปรวนของคลาดเคลื่อนสุ่มเพิ่มขึ้น ค่า AMSE ของทุกวิธีมีแนวโน้มเพิ่มขึ้นเพราะค่า AMSE เป็นตัวประมาณค่าความแปรปรวนของความคลาดเคลื่อนสุ่ม (σ^2)

3. จำนวนตัวแปรอิสระ

ค่า AMSE มีความสัมพันธ์แบบแปรผันตามจำนวนตัวแปรอิสระ กล่าวคือ เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น จะส่งผลให้ค่า AMSE ของทุกวิธีมีแนวโน้มสูงขึ้น เนื่องจากในการวิจัยครั้งนี้ได้กำหนดตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูปเมื่อมีจำนวนตัวแปรอิสระเพิ่มขึ้นจำนวนตัวแบบที่เป็นไปได้ทั้งหมดที่นำมาพิจารณาจะมีมากขึ้นจึงทำให้โอกาสที่จะได้ตัวแบบที่ไม่เหมาะสมก็จะมีมากขึ้นทำให้ค่า AMSE ของทุกเกณฑ์มีแนวโน้มสูงขึ้น และสังเกตได้ว่า เมื่อจำนวนตัวแปรอิสระเพิ่มขึ้น ค่า AMSE ของเกณฑ์ HQ_C และเกณฑ์ AIC_C จะแตกต่างจากค่า AMSE เกณฑ์ KIC_C และ SIC_C มากยิ่งขึ้น

5.2 ข้อเสนอแนะ

1. การวิจัยครั้งนี้มีขอบเขตการวิจัยเมื่อตัวแบบที่นำมาพิจารณาเป็นแบบติดกลุ่ม โดยกำหนดให้ตัวแบบเริ่มต้นเป็นตัวแบบเต็มรูป และตัวแบบที่นำมาพิจารณาเป็นตัวแบบลดรูปของตัวแบบเริ่มต้น ซึ่งจะไม่มีโอกาสได้ตัวแบบที่มีจำนวนพารามิเตอร์สูงกว่าตัวแบบจริง ทำให้ผลการวิจัยพบว่าเกณฑ์ AIC_C เป็นเกณฑ์ที่ให้ค่าพยากรณ์ที่ถูกต้องและเหมาะสมที่สุด ทั้งนี้เนื่องจากเกณฑ์ AIC_C เป็นเกณฑ์ที่ไม่มีความเอนเอียงสำหรับการได้จำนวนพารามิเตอร์น้อยกว่าจำนวนพารามิเตอร์จริง (underfitting estimation) ซึ่งได้เปรียบเกณฑ์อื่น ๆ ที่มีความเอนเอียงนี้อยู่ ดังนั้นในการวิจัยครั้งต่อไปควรมีการกำหนดตัวแบบที่นำมาพิจารณาอาจจะไม่ใช่ตัวแบบลดรูปของตัวแบบเริ่มต้นหรือไม่ติดกลุ่มกับตัวแบบเริ่มต้น ทั้งนี้เพื่อศึกษาว่าจะให้ผลสรุปเหมือนหรือแตกต่างกันอย่างไร

2. การวิจัยครั้งนี้ศึกษาเฉพาะการวิเคราะห์การถดถอยเชิงเส้นพหุคูณ ไม่รวมถึงตัวแบบการถดถอยพหุนาม ซึ่งเป็นกรณีเฉพาะและพบได้บ่อยครั้งในการวิเคราะห์การถดถอย ดังนั้นในการวิจัยครั้งต่อไปควรมีการศึกษาเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยสำหรับตัวแบบการถดถอยพหุนามด้วย ทั้งนี้เพื่อศึกษาว่าการนำเกณฑ์การคัดเลือกตัวแบบทั้ง 4 วิธีเมื่อนำไปใช้กับตัวแบบพหุนามจะให้ผลสรุปเหมือนหรือแตกต่างกันอย่างไร

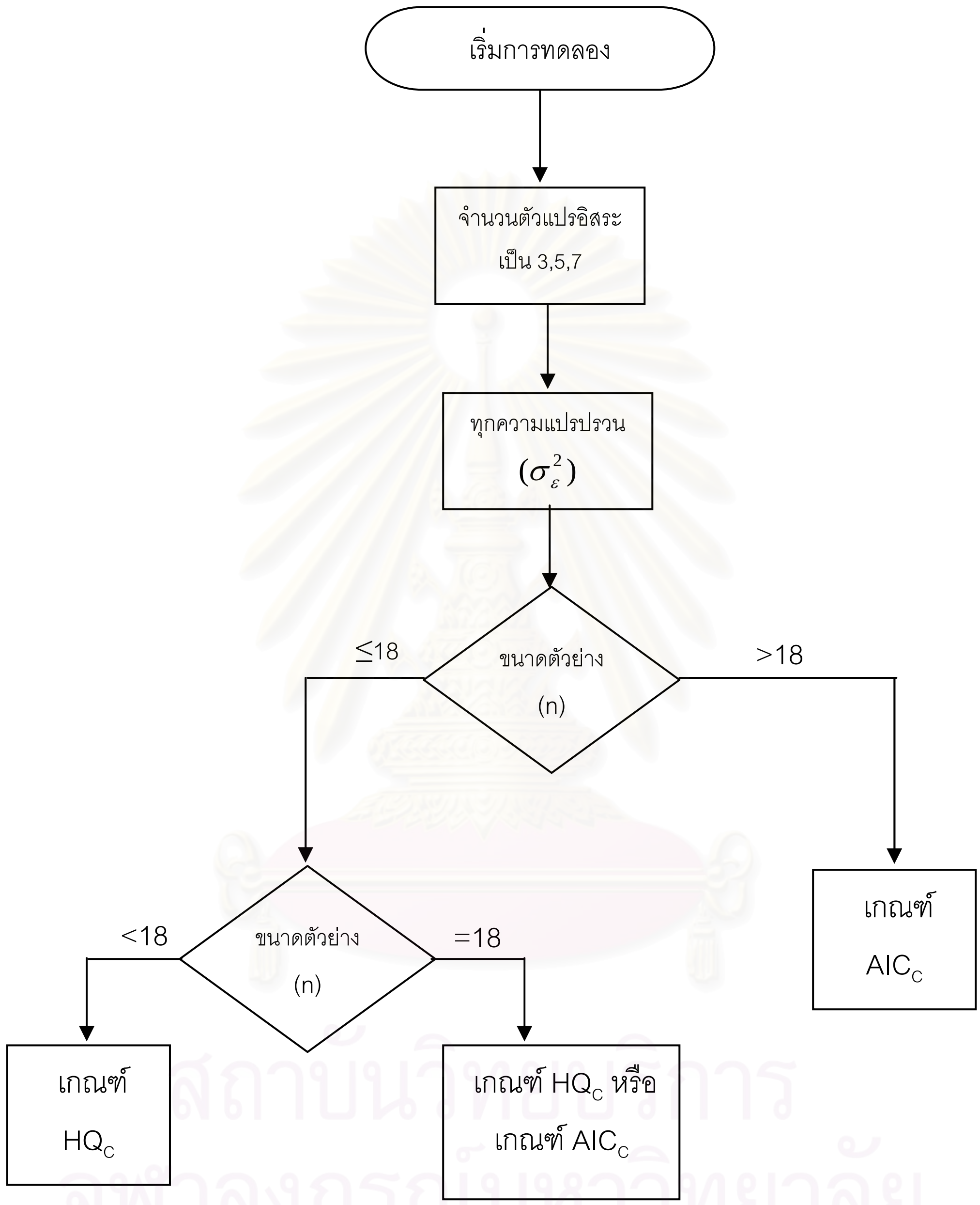
3. การวิจัยครั้งนี้มีขอบเขตในการวิจัยโดยกำหนดให้ การแจกแจงของความคลาดเคลื่อนสุ่มเป็นแบบปกติ และเป็นไปตามข้อตกลงเบื้องต้น ซึ่งไม่ได้พิจารณากรณีที่การแจกแจงของความคลาดเคลื่อนสุ่มเป็นแบบอื่น ดังนั้นในการวิจัยครั้งต่อไปควรจะพิจารณาการแจกแจงของความคลาดเคลื่อนสุ่มเป็นแบบอื่นๆ อาจเป็น เบ้ขวา เช่น การแจกแจงแบบเอ็กโปเนนเชียล การแจกแจงแบบปัวซอง การแจกแจงเบตาและแกมมา เป็นต้น ซึ่งจะพบได้บ่อยในการวิเคราะห์ด้านเศรษฐศาสตร์และอุตสาหกรรม เช่น ระยะเวลาในการรอรับบริการของลูกค้า จำนวนสินค้าที่ต้องตรวจสอบจนกว่าจะพบของเสีย ฯลฯ เพราะในการคัดเลือกตัวแบบของแต่ละเกณฑ์ต้องใช้ฟังก์ชันความควรจะเป็น ซึ่งขึ้นอยู่กับลักษณะการแจกแจงของข้อมูล ดังนั้นจึงอาจส่งผลให้ได้ข้อสรุปที่แตกต่างกันไปก็ได้

4. การวิจัยครั้งนี้มีขอบเขตการวิจัยเป็นตัวแบบติดกลุ่ม (nested model) เท่านั้น ดังนั้นในการวิจัยครั้งต่อไป ควรพิจารณาเกณฑ์การคัดเลือกตัวแบบการถดถอยเมื่อตัวแบบเป็นตัวแบบไม่ติดกลุ่ม ทั้งนี้จากการศึกษาพบว่าสามารถประยุกต์เกณฑ์การคัดเลือกตัวแบบไปใช้เมื่อตัวแบบไม่ติดกลุ่มได้แต่ต้องระมัดระวังในการเปรียบเทียบค่าที่ได้จากเกณฑ์(criterion value) เพราะความแตกต่างของค่าที่ได้จากเกณฑ์จะอธิบายได้ดีน้อยกว่าตัวแบบติดกลุ่ม[2]

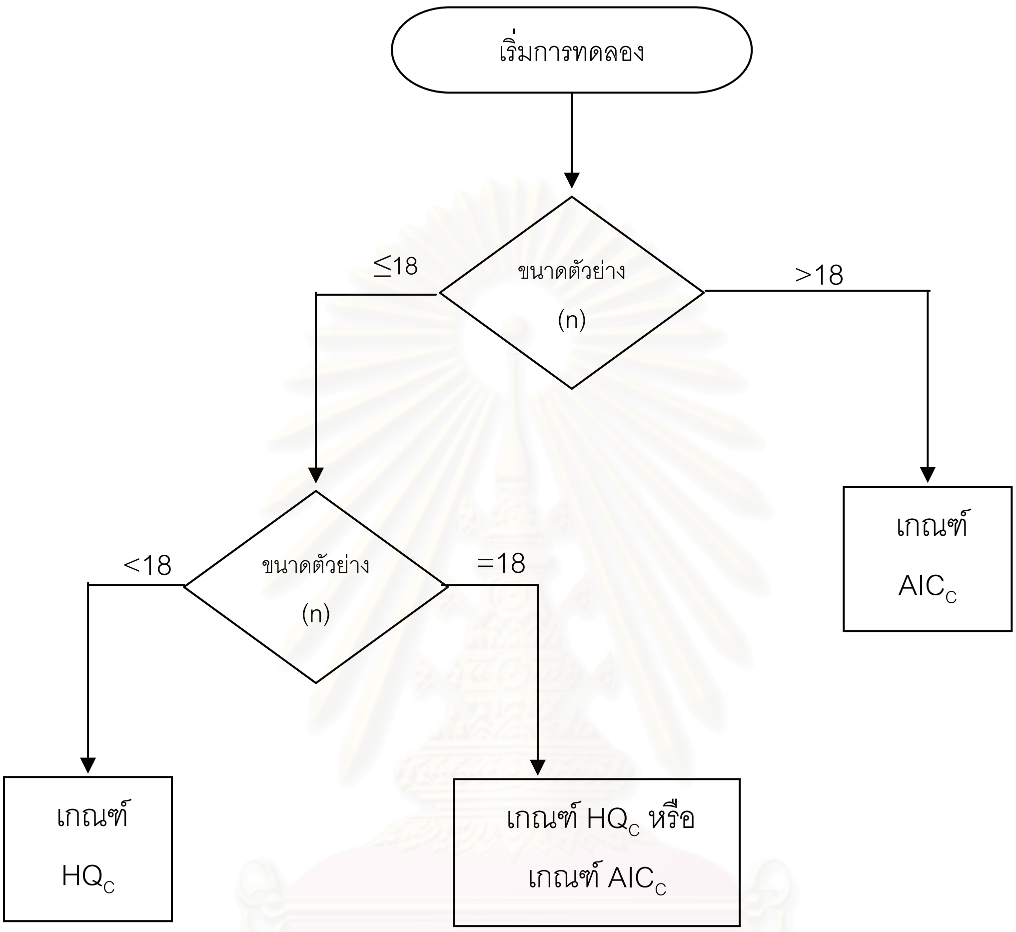
---

[2]Joe Cavanaugh.Advanced Biostatistics Seminar:Model selection.Department of Biostatistics,College of Pubplic Health,The University of Iowa.2005

แผนผังผลสรุปการเลือกเกณฑ์การคัดเลือกตัวแบบในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณสำหรับตัวอย่างขนาดเล็ก สำหรับใช้ในทางทฤษฎี



แผนผังผลสรุปการเลือกเกณฑ์การคัดเลือกตัวแบบในการวิเคราะห์การถดถอยเชิงเส้นพหุคูณสำหรับตัวอย่างขนาดเล็ก สำหรับใช้ในทางปฏิบัติ



รายการอ้างอิง

ภาษาไทย

ธีระพร วีระถาวร. การประมาณอนุมานเชิงสถิติขั้นกลาง:โครงสร้างและความหมาย. กรุงเทพมหานคร:สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย. 2536.

ธีระพร วีระถาวร. ตัวแบบเชิงเส้น ทฤษฎีและการประยุกต์.กรุงเทพมหานคร: พิมพ์ดี . 2541.

นิทัสน์ สุขสุวรรณ. การเปรียบเทียบวิธีการคัดเลือกสมการการถดถอยที่ดีที่สุดภายใต้แนวทางของเบส์. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2545.

บุญจิรา มากอ้น. การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยแบบไม่ติดกลุ่ม. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2545.

ประชุม สุวัตถี. ทฤษฎีการอนุมานเชิงสถิติ. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2545.

พจนา แววสวัสดิ์. การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบความถดถอยพหุนามแบบติดกลุ่ม. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย, 2543.

มานพ วราภักดิ์. การจำลองเบื้องต้น : Introduction to Simulation. กรุงเทพมหานคร : ศูนย์ผลิตตำราเรียน สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ. 2547.

สุพล ดุรงค์วัฒนา. การวิเคราะห์เชิงสถิติ : การวิเคราะห์ความถดถอย. กรุงเทพมหานคร.โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย. 2536.

ภาษาต่างประเทศ

A.D.Maquarrie,C.L.Tsai. Regression and time series model selection. Singapore : Word scienctific Publishing. 1998.

A.D.Maquarrie. A small-sample correction for The Schwarz SIC model selection criterion.Statistics & Probability Letter,44(1999):79-86.

And-Krim Seghouance and Maiza Bekora.A small sample model selection criterion based on Kullback's symmetric divergence. IEEE Transactios on signal processing,52(2004):3314-3323.

C.M.Hurvich and C.L.Tsi.Regression and time series model selection in small samples.Biometrika,76(1989):297-307.

E.J.Hannan and B.G.Quinn.The determination of the order of an autoregression. Journal of Royal statistics associate.serie B,41(1979): 287-307.

H.linhart and W.Zucchini.Model selection.New York :John Wisely & son.1986.

Joseph E. Cavanaugh. Unifying the derivations of the Akaike and corrected Akaike information criteria. Statistics & Probability Letter,33(1997):201-208.

Joseph E. Cavanaugh. A large-sample model selection criterion based on Kullback's symmetric divergence. Statistics & Probability Letter,42(1999):333-343.

Joseph E. Cavanaugh. Criteria for linear model selection based on Kullback's symmetric divergence . Australian and New Zealand Journal of Statistics,46(2004):257-274.

Solomon Kullback. Information Theory and statistiics. New York :John Wisely & son.1959.

ภาคผนวก

ตารางแสดงลักษณะการทำงานของโปรแกรมทั้งหมดที่ใช้การวิจัย

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อยที่เรียกใช้
โปรแกรมหลัก	thesis_code	- สร้างข้อมูลของตัวแปรอิสระ ค่าคลาดเคลื่อนสุ่ม และตัวแปรตาม	creat_x , creat_y , creat_err
		- สร้างตัวแบบทั้งหมด	model_all_out
		- คำนวณค่า MSE ของวิธี AIC_c	get_mse
		- คำนวณค่า MSE ของวิธี SIC_c	
		- คำนวณค่า MSE ของวิธี HQ_c	
		- คำนวณค่า MSE ของวิธี KIC_c	
โปรแกรมย่อย			
1	datanormal	- การสร้างเลขสุ่มให้มีการแจกแจงแบบปกติ	
2	creat_x	- สร้างตัวแปรอิสระ	datanormal
3	creat_err	- สร้างความคลาดเคลื่อนสุ่ม	datanormal
4	creat_y	- สร้างตัวแปรตาม	
5	model_all_out	- สร้างตัวแบบทั้งหมด	
6	beta	- คำนวณค่าสัมประสิทธิ์การถดถอย	
7	xtx	- คำนวณ x ทรานโพส x	
8	xty	- คำนวณ x ทรานโพส y	
9	inv	- คำนวณเมทริกซ์ผกผัน	det
10	xbcross	- คำนวณ $X\hat{\beta}$	

อันดับที่	ชื่อโปรแกรม	การทำงานของโปรแกรม	ชื่อโปรแกรมย่อยที่เรียกใช้
ฟังก์ชัน			
11	det	- คำนวณค่าดีเทอร์มิแนน	
12	aic	- คำนวณค่า AIC_c ของตัวแบบ	
13	sic	- คำนวณค่า SIC_c ของตัวแบบ	
14	hq	- คำนวณค่า HQ_c ของตัวแบบ	
15	kic	- คำนวณค่า KIC_c ของตัวแบบ	
16	get_mse	- คำนวณค่า MSE	xbcross
17	get_sse	- คำนวณค่า SSE	xbcross

โดยรายละเอียดของโปรแกรมมีดังนี้

```
program thesis_code;
{$APPTYPE CONSOLE}
uses
  SysUtils,
  Math;
const
   sample1 = 12; sample2 = 15; sample3 = 18; sample4 = 21; sample5 = 24; sample6 = 27;sample7=30;
   sigma1 = 1; sigma2 = 2; sigma3 =4 ; sigma4 = 8; sigma5 = 16;  iteration = 1000; all_model=128;
   inde1 = 3; inde2 = 5; inde3 = 7; test = 4;
type
   data = array[1..sample7,1..(inde3+1)]of double;   data2 = array[1..(inde3+1),1..(inde3+1)]of double;
   data3 = array[1..sample7]of double;   data4 = array[1..(inde3+1)]of double;
   data5 = array[1..all_model,1..test]of double;   data6 = array[1..all_model,1..sample7,1..(inde3+1)]of double;
   data7 = array[1..all_model,1..sample7]of double;   data8 = array[1..all_model,1..(inde3+1),1..(inde3+1)]of double;
   data9 = array[1..test,1..iteration]of double;   data10 = array[1..(inde3+1)]of integer;
   data11 = array[1..test]of double;   data12 = array[1..test,1..iteration]of integer;
   data13 = array[1..all_model]of double;
var
  total,count,sam,num,para,k_para,ind,q,r,s,i1,j1,varia:integer;
  ch,ch1,ch2,ch3,ch4,ch5,ch6,ch7,mo,mo2,mo4:integer;
  repeat1,move,move1,column,row,num_start,num_end:integer;
  temp_mse,temp_sse,amse_min,min_mse,mse_full,sig:double;
  x_full,x_full1,x_model:data;  xtx_model,inv_xtx_model:data2;
  err_model,y_model:data3;   xty_model,beta_model:data4;
  seclec:data5;  model_x:data6;  mse:data13;  mse_min,count1:data9;
  up,down,para_in:data10;  amse,sd,diff,count2:data11;
procedure datanormal(n:integer;var z:data3);
var i,j:integer; r1,r2:double; z2:data3;
begin
   randomize;
   i := 1;
   for j:= 1 to sample7 do
     z2[j]:=0;
   repeat
     if (i <= n)then
       begin
         r1 := random;  r2 := random;
         z2[i]:=sqrt((-2)*ln(r1))*(cos(2*pi*r2));
       end
     else
         z2[i]:= 0;
```

```
          i := i+1;
        until i > sample7;
  z := z2;
end;
procedure creat_err(n:integer;variance:double;var error:data3);
var i:integer; temp,z1:data3;
begin
  for i:= 1 to sample7 do
  begin
    temp[i]:= 0;
    z1[i]:=0;
  end;
  datanormal(n,z1);
  for i:= 1 to n do
  begin
    temp[i]:=  sqrt(variance)*z1[i];
  end;
  error:=temp;
end;
procedure creat_x(n,p:integer;var x_out:data);
var i,j,k:integer; z1:data3; temp_in:data;
begin
  for i:= 1 to sample7 do
  begin
    for j:= 1 to inde3+1 do
      temp_in[i,j]:= 0;
  end;
  for j:= 1 to p do
  begin
    if (j=1) then
    for i:= 1 to n do
      temp_in[i,j]:= 1
    else
    begin
      for k:= 1 to sample7 do
      begin
        z1[k]:=0;
        temp_in[k,j]:=0;
      end;
      datanormal(n,z1);
      for i:= 1 to n do
```

```
        begin
          temp_in[i,j]:=z1[i];
        end;
      end;
    end;
    x_out:= temp_in;
end;
function fractorial(num1:integer):integer;
begin
   if (num1=0) then
     fractorial:=1
   else
     fractorial:=num1*fractorial(num1-1);
end;
function combination(num,r:integer):double;
var temp:double;
begin
    temp:=fractorial(num)/(fractorial(r)*fractorial(num-r));
    combination:=temp;
end;
procedure model_all_out(n,p:integer;x_input:data;var model_all1:data6);
var i,j,k,model,temp_index,index,index1,index2,index3,index4,index5,index6:integer;  model_all:data6;
begin
  model:= round(power(2,(p-1)));
  index:=2;   index1:=2;   index2:=3;   index3:=4;   index4:=5;   index5:=6;   index6:=7;
  for i:= 1 to all_model do
    for j:= 1 to sample7 do
      for k:= 1 to inde3+1 do
        model_all[i,j,k]:= 0;
  for k:= 1 to model do
  begin
    if (p = 4) then
    begin
      if(k=1)then
      begin
        for i:=1 to n do
          model_all[k,i,k]:= x_input[i,k];
      end
      else if ((2<=k)and(k<=4)) then
      begin
        for j:=1 to 2 do
```

```
    begin
      for i:= 1 to n do
      begin
        if (j=1) then
          model_all[k,i,j]:= x_input[i,j]
        else
          model_all[k,i,j]:= x_input[i,index];
      end;
    end;
    index:=index+1;
  end
  else if ((5<=k)and(k<=7)) then
  begin
    for j:= 1 to 3 do
    begin
      for i:= 1 to n do
      begin
        if(j=1) then
          model_all[k,i,j]:=x_input[i,j]
        else if(j=2) then
        begin
          model_all[k,i,j]:=x_input[i,index1];
        end
        else
        begin
          model_all[k,i,j]:=x_input[i,index2];
        end;
      end;
    end;
    index2:=index2+1;
    if (k=6) then index1:=index1+1;
    if (k=6) then index2:=index2-1;
  end
  else
  begin
    for j:= 1 to 4 do
    begin
      for i:= 1 to n do
      begin
        model_all[k,i,j]:=x_input[i,j]
      end;
```

```
    end;
  end;
end
else if (p=6) then
begin
  if(k=1)then
  begin
    for i:=1 to n do
      model_all[k,i,k]:= x_input[i,k];
  end
  else if ((2<=k)and(k<=6)) then
  begin
    for j:=1 to 2 do
    begin
      for i:= 1 to n do
      begin
        if (j=1) then
          model_all[k,i,j]:= x_input[i,j]
        else
          model_all[k,i,j]:= x_input[i,index];
      end;
    end;
    index:=index+1;
  end
  else if ((7<=k)and(k<=16)) then
  begin
    for j:= 1 to 3 do
    begin
      for i:= 1 to n do
      begin
        if(j=1) then
          model_all[k,i,j]:=x_input[i,j]
        else if(j=2) then
        begin
          model_all[k,i,j]:=x_input[i,index1];
        end
        else
        begin
          model_all[k,i,j]:=x_input[i,index2];
        end;
      end;
```

```
  end;
  index2:=index2+1;
  if (k=10) then index1:=index1+1;
  if (k=13) then index1:=index1+1;
  if (k=15) then index1:=index1+1;
  if (k=10) then index2:=index2-3;
  if (k=13) then index2:=index2-2;
  if (k=15) then index2:=index2-1;
  if (k=16) then
  begin
    index1:= 2;
    index2:= 3;
  end;
end
else if ((17<=k)and(k<=26)) then
begin
  for j:= 1 to 4 do
  begin
    for i:= 1 to n do
    begin
      if(j=1) then
        model_all[k,i,j]:=x_input[i,j]
      else if(j=2) then
      begin
        model_all[k,i,j]:=x_input[i,index1];
      end
      else if(j=3) then
      begin
        model_all[k,i,j]:=x_input[i,index2];
      end
      else
      begin
        model_all[k,i,j]:=x_input[i,index3];
      end;
    end;
  end;
  index3:=index3+1;
  if (k=19) then
  begin
    index2:=index2+1;  index3:=index3-2;
  end;
```

```
  if (k=21) then
  begin
    index2:=index2+1;  index3:=index3-1;
  end;
  if (k=22) then
  begin
    index1:=index1+1;   index2:= index2-1;   index3:= index3-2;
  end;
  if (k=24) then
  begin
    index2:=index2+1;   index3:=index3-1;
  end;
  if (k=25) then
  begin
    index1:=index1+1;   index3:=index3-1;
  end;
  if (k=26) then
  begin
    index1:=2;   index2:=3;   index3:=4;
  end;
end
else if ((27<=k)and(k<=31)) then
begin
 for j:= 1 to 5 do
  begin
    for i:= 1 to n do
    begin
      if(j=1) then
        model_all[k,i,j]:=x_input[i,j]
      else if(j=2) then
      begin
        model_all[k,i,j]:=x_input[i,index1];
      end
      else if(j=3) then
      begin
        model_all[k,i,j]:=x_input[i,index2];
      end
      else if(j=4) then
      begin
        model_all[k,i,j]:=x_input[i,index3];
      end
```

```
          else
          begin
            model_all[k,i,j]:=x_input[i,index4];
          end;
        end;
      end;
      if (k=30) then index1:=index1+1;
      if (k=29) then index2:=index2+1;
      if (k=28) then index3:=index3+1;
      if (k=27) then index4:=index4+1;
    end
    else
    begin
      for j:= 1 to 6 do
      begin
        for i:= 1 to n do
        begin
          model_all[k,i,j]:=x_input[i,j]
        end;
      end;
    end;
  end
  else  { degree 7 }
  begin
    if(k=1)then
    begin
      for i:=1 to n do
        model_all[k,i,k]:= x_input[i,k];
    end
    else if ((2<=k)and(k<=8)) then
    begin
      for j:=1 to 2 do
      begin
        for i:= 1 to n do
        begin
          if (j=1) then
            model_all[k,i,j]:= x_input[i,j]
          else
            model_all[k,i,j]:= x_input[i,index];
        end;
      end;
```

```
  index:=index+1;
end
else if ((9<=k)and(k<=29)) then
begin
  for j:= 1 to 3 do
  begin
    for i:= 1 to n do
    begin
      if(j=1) then
        model_all[k,i,j]:=x_input[i,j]
      else if(j=2) then
      begin
        model_all[k,i,j]:=x_input[i,index1];
      end
      else
      begin
        model_all[k,i,j]:=x_input[i,index2];
      end;
    end;
  end;
  index2:=index2+1;
  if (k=14) then index1:=index1+1;
  if (k=19) then index1:=index1+1;
  if (k=23) then index1:=index1+1;
  if (k=26) then index1:=index1+1;
  if (k=28) then index1:=index1+1;
  if (k=14) then index2:=index2-5;
  if (k=19) then index2:=index2-4;
  if (k=23) then index2:=index2-3;
  if (k=26) then index1:=index2-2;
  if (k=28) then index1:=index2-1;
  if (k=29) then
  begin
    index1:= 2;  index2:= 3;
  end;
end
else if ((30<=k)and(k<=64)) then
begin
  for j:= 1 to 4 do
  begin
    for i:= 1 to n do
```

```
  begin
    if(j=1) then
      model_all[k,i,j]:= x_input[i,j]
    else if(j=2) then
    begin
      model_all[k,i,j]:= x_input[i,index1];
    end
    else if(j=3) then
    begin
      model_all[k,i,j]:= x_input[i,index2];
    end
    else
    begin
      model_all[k,i,j]:= x_input[i,index3];
    end;
  end;
end;
index3:=index3+1;
if (k=34) then
begin
  index2:=index2+1;  index3:=index3-4;
end;
if (k=38) then
begin
  index2:=index2+1;   index3:=index3-3;
end;
if (k=41) then
begin
  index2:=index2+1;  index3:=index3-2;
end;
if (k=43) then
begin
  index2:=index2+1;   index3:=index3-1;
end;
if (k=44) then
begin
  index1:=index1+1;  index2:= index2-3;   index3:= index3-4;
end;
if (k=48) then
begin
  index2:=index2+1;   index3:=index3-3;
```

```
  end;
  if (k=51) then
  begin
    index2:=index2+1;   index3:=index3-2;
  end;
  if (k=53) then
  begin
    index2:=index2+1;   index3:=index3-1;
  end;
  if (k=54) then
  begin
    index1:=index1+1;   index2:= index2-2;   index3:= index3-3;
  end;
  if (k=57) then
  begin
    index2:=index2+1;   index3:=index3-2;
  end;
  if (k=59) then
  begin
    index2:=index2+1;   index3:=index3-1;
  end;
  if (k=60) then
  begin
    index1:=index1+1;   index2:= index2-1;   index3:= index3-2;
  end;
  if (k=62) then
  begin
    index2:=index2+1;   index3:=index3-1;
  end;
  if (k=63) then
  begin
    index1:=index1+1;   index3:= index3-1;
  end;
  if (k=64) then
  begin
    index1:=2;
    index2:=3;
    index3:=4;
  end;
end
else if ((65<=k)and(k<=99)) then
```

```
begin
 for j:= 1 to 5 do
  begin
    for i:= 1 to n do
    begin
      if(j=1) then
        model_all[k,i,j]:=x_input[i,j]
      else if(j=2) then
      begin
        model_all[k,i,j]:=x_input[i,index1];
      end
      else if(j=3) then
      begin
        model_all[k,i,j]:=x_input[i,index2];
      end
      else if(j=4) then
      begin
        model_all[k,i,j]:=x_input[i,index3];
      end
      else
      begin
        model_all[k,i,j]:=x_input[i,index4];
      end;
    end;
  end;
  if (k=68) then
   begin
    index3:=index3+1;   index4:= index4-3;
   end;
  if (k=71) then
   begin
    index3:=index3+1;   index4:= index4-2;
   end;
  if (k=73) then
   begin
    index3:=index3+1;   index4:= index4-1;
   end;
  if (k=74) then
   begin
    index2:=index2+1;  index3:=index3-3;  index4:=index4-3;
   end;
```

```
if (k=68) then
  begin
    index3:=index3+1;  index4:= index4-3;
  end;
if (k=71) then
  begin
    index3:=index3+1;   index4:= index4-2;
  end;
if (k=73) then
  begin
    index3:=index3+1;   index4:= index4-1;
  end;
if (k=74) then
  begin
    index2:=index2+1; index3:=index3-2;   index4:=index4-3;
  end;
if (k=77)then
  begin
    index3:=index3+1;  index4:=index4-2;
  end;
 if (k=79) then
  begin
    index3:=index3+1;  index4:= index4-1;
  end;
if (k=80) then
  begin
    index2:=index2+1;  index3:=index3-1;   index4:=index4-2;
  end;
 if (k=82) then
  begin
    index3:=index3+1;  index4:= index4-1;
  end;
if (k=83) then
  begin
    index2:=index2+1;  index4:=index4-1;
  end;
if (k=84) then
  begin
    index1:=index1+1;   index2:=index2-2;   index3:=index3-2;   index4:=index4-3;
  end;
if (k=87) then
```

```
  begin
    index3:=index3+1;   index4:=index4-2;
  end;
 if (k=89) then
  begin
    index3:=index3+1;   index4:= index4-1;
  end;
 if (k=90) then
  begin
    index2:=index2+1;  index3:=index3-1;  index4:=index4-2;
  end;
 if (k=92) then
  begin
    index3:=index3+1;   index4:= index4-1;
  end;
 if (k=93) then
  begin
    index2:=index2+1;  index4:=index4-1;
  end;
 if (k=94) then
  begin
    index1:=index1+1;  index2:=index2-1;   index3:=index3-1;  index4:=index4-2;
  end;
 if (k=96) then
  begin
    index3:=index3+1;  index4:= index4-1;
  end;
 if (k=97) then
  begin
    index2:=index2+1;  index4:=index4-1;
  end;
if (k=98) then
  begin
    index1:=index1+1;   index4:=index4-1;
  end;
if (k=99) then
 begin
   index1:=2;
   index2:=3;
   index3:=4;
   index4:=5;
```

```
  end;
end
else if ((100<=k)and(k<=120)) then
begin
 for j:= 1 to 6 do
  begin
    for i:= 1 to n do
    begin
      if(j=1) then
        model_all[k,i,j]:=x_input[i,j]
      else if(j=2) then
      begin
        model_all[k,i,j]:=x_input[i,index1];
      end
      else if(j=3) then
      begin
        model_all[k,i,j]:=x_input[i,index2];
      end
      else if(j=4) then
      begin
        model_all[k,i,j]:=x_input[i,index3];
      end
      else if (j=5)then
      begin
        model_all[k,i,j]:=x_input[i,index4];
      end
      else
      begin
        model_all[k,i,j]:=x_input[i,index5];
      end;
    end;
  end;
  if (k=102) then
   begin
    index4:=index4+1;  index5:= index5-2;
   end;
  if (k=104) then
   begin
    index4:=index4+1;  index5:= index5-1;
   end;
  if (k=105) then
```

```
  begin
    index3:=index3+1;  index4:= index4-1;   index5:= index5-2;
  end;
if (k=107) then
  begin
    index4:=index4+1;   index5:=index5-1;
  end;
if (k=108) then
  begin
    index3:=index3+1;  index5:= index5-1;
  end;
if (k=109) then
  begin
    index2:=index2+1;  index3:= index3-1;  index4:=index4-1;  index5:= index5-2;
  end;
if (k=111) then
  begin
    index4:=index4-1;  index5:= index5-1;
  end;
if (k=112) then
  begin
    index3:=index3+1;  index5:= index5-1;
  end;
if (k=113) then
  begin
    index2:=index2+1;  index5:= index5-1;
  end;
if (k=114) then
  begin
    index1:=index1+1;  index2:=index2-1;  index3:= index3-1; index4:=index4-1;  index5:= index5-2;
  end;
  if (k=116) then
  begin
    index4:=index4+1;  index5:= index5-1;
  end;
if (k=117) then
  begin
    index3:=index3+1;  index5:= index5-1;
  end;
if (k=118) then
  begin
```

```
      index2:=index2+1;  index5:= index5-1;
    end;
   if (k=119) then
    begin
      index1:=index1+1;   index5:= index5-1;
    end;
   if (k=120) then
   begin
     index1:=2;
     index2:=3;
     index3:=4;
     index4:=5;
     index5:=6;
   end;
end
else if ((121<=k)and(k<=127)) then
begin
  for j:= 1 to 7 do
   begin
     for i:= 1 to n do
     begin
       if(j=1) then
         model_all[k,i,j]:=x_input[i,j]
       else if(j=2) then
       begin
         model_all[k,i,j]:=x_input[i,index1];
       end
       else if(j=3) then
       begin
         model_all[k,i,j]:=x_input[i,index2];
       end
       else if(j=4) then
       begin
         model_all[k,i,j]:=x_input[i,index3];
       end
       else if(j=5)then
       begin
         model_all[k,i,j]:=x_input[i,index4];
       end
       else if(j=6)then
       begin
```

```
              model_all[k,i,j]:=x_input[i,index5];
            end
            else
            begin
              model_all[k,i,j]:=x_input[i,index6];
            end;
          end;
        end;
        if (k=122) then
          begin
            index5:=index5+1; index6:= index6-1;
          end;
        if (k=123) then
          begin
            index4:=index4+1;  index5:= index5-1;  index6:= index6-1;
          end;
        if (k=124) then
          begin
            index3:=index3+1;  index6:= index6-1;
          end;
        if (k=125) then
          begin
            index2:=index2+1;  index6:=index6-1;
          end;
        if (k=126) then
          begin
            index1:=index1+1;   index6:= index6-1;
          end;
      end
      else
      begin
        for j:= 1 to 8 do
        begin
          for i:= 1 to n do
          begin
            model_all[k,i,j]:=x_input[i,j]
          end;
        end;
      end;
    end;
end;
```

```
  model_all1:= model_all;
end;

procedure creat_y(n,p:integer;x:data;err:data3;var y:data3);
var i,j:integer; b:data4;
begin
  for i:= 1 to sample7 do
    for j:= 1 to p do
    begin
      y[i]:= y[i]+ x[i,j];
    end;
    y[i]:=y[i]+err[i];
  end;
end;
procedure xtx(n,p:integer;x:data;var xtx_out:data2);
var i,j,k:integer; temp:data2;
begin
  for i:= 1 to inde3+1 do
    for j:= 1 to inde3+1 do
      temp[i,j]:=0;
  for i:= 1 to p do
   for j:= 1 to p do
   begin
     temp[i,j]:= 0;
     for k:= 1 to n do
     begin
       temp[i,j]:= temp[i,j] + (x[k,i]*x[k,j]);
     end;
   end;
  xtx_out:=temp;
end;
procedure xty(n,p:integer;x:data;y:data3;var xty_out:data4);
var i,j:integer; temp:data4;
begin
  for i:= 1 to p do
  begin
    temp[i]:=0;
    for j:= 1 to n do
    begin
      temp[i]:=temp[i]+ (x[j,i]*y[j]);
    end;
```

```
  end;
  xty_out:=temp;
end;
procedure beta(p:integer;x:data2;y:data4;var beta_out:data4);
var i,j:integer; temp:data4;
begin
  for i:= 1 to p do
  begin
    temp[i]:=0;
    for j:= 1 to p do
    begin
      temp[i]:=temp[i]+(x[i,j]*y[j]);
    end;
  end;
  beta_out:=temp;
end;
procedure xbcross(x:data;b:data4;n,p:integer;var xb:data3);
var i,j:integer;
begin
  for i:= 1 to n do
  begin
   xb[i]:=0;
   for j:= 1 to p do
     xb[i]:=xb[i]+(x[i,j]*b[j]);
  end;
end;
function get_sse(x:data;y:data3;b:data4; n,p:integer):double;
var tmp_dif,temp:double; i:integer; xb:data3;
begin
  xbcross(x,b,n,p,xb);
  tmp_dif:=0;
  for i:=1 to n do
  begin
    temp:= y[i]- xb[i];
    tmp_dif:= tmp_dif + power(temp,2);
  end;
  tmp_dif:= tmp_dif/n;
  get_sse := tmp_dif;
end;
function get_mse(x:data;y:data3;b:data4; n,p:integer):double;
var tmp_dif,temp:double; i:integer; xb:data3;
```

```
begin
  xbcross(x,b,n,p,xb);
  tmp_dif:=0;
  for i:=1 to n do
  begin
    temp:= y[i]- xb[i];
    tmp_dif:= tmp_dif + power(temp,2);
  end;
  tmp_dif:= tmp_dif/(n-p);
  get_mse := tmp_dif;
end;
function det(x:data2;p:integer):double;
var i,j,k:integer; temp,temp1:data2; sum:double;
begin
  if (p=1) then
    det:= x[1,1]
  else
  if (p = 2) then
    det:= (x[1,1]*x[2,2])-(x[2,1]*x[1,2])
  else
  begin
    sum:= 0;
    for j:= 1 to p-1 do
      for k:= 1 to p do
          temp[j,k]:= x[j+1,k];
    temp1:=temp;
    for i:= 1 to p do
    begin
      if i=1 then
      begin
        for j:= 1 to p-1 do
          for k:= 1 to p-1 do
            temp[k,j]:= temp1[k,j+1]
      end
      else
      begin
        for j:= 1 to p-1 do
          for k:= 1 to p-1 do
            if j < i then  temp[k,j]:= temp1[k,j]  else  temp[k,j]:= temp1[k,j+1];
      end;
        sum:=sum+(power((-1),(i+1))*x[1,i]*det(temp,p-1));
```

```
    end;
    det := sum;
  end;
end;
procedure inv(x:data2; p:integer; var xin:data2);
var i,j,kr,kc,m:integer; d,cal:double; temp,temp1:data2;
begin
  for i:=1 to inde3 do
    for j:=1 to inde3 do
    begin
      xin[i,j]:=0;
      temp[i,j]:=0;
      temp1[i,j]:=0;
    end;
  d:= det(x,p);
  if(d<>0)then
  begin
    for i:= 1 to p do
    begin
      for j:=1 to p do
      begin
        if (i=1) then
        begin
          for kr:= 1 to p-1 do
            for kc:=1 to p do
              temp[kr,kc]:=x[kr+1,kc];
          if (j=1) then
            for kc:= 1 to p-1 do
              for kr:= 1 to p-1 do
                temp1[kr,kc]:=temp[kr,kc+1]
          else
            for kc:= 1 to p-1 do
              for kr:= 1 to p-1 do
                if (kc<j) then
                  temp1[kr,kc]:=temp[kr,kc]
                else
                  temp1[kr,kc]:=temp[kr,kc+1];
        end
        else
        begin
          for kr:= 1 to p-1 do
```

```
          for kc:= 1 to p do
            if (kr<i) then
              temp[kr,kc]:=x[kr,kc]
            else
              temp[kr,kc]:=x[kr+1,kc];
        if (j=1) then
          for kc:= 1 to p-1 do
            for kr:= 1 to p-1 do
              temp1[kr,kc]:=temp[kr,kc+1]
        else
          for kc:= 1 to p-1 do
            for kr:= 1 to p-1 do
              if (kc<j) then
                temp1[kr,kc]:=temp[kr,kc]
              else
                temp1[kr,kc]:=temp[kr,kc+1];
      end;
      xin[j,i]:= power((-1),(i+j))*(1/d)*(det(temp1,p-1));
    end;
  end;
 end
 else
  writeln('Undefined  Invert  of  XTX');
end;
function aic(n1,k1:integer;sigma:double):double;
var temp:double;
begin
 temp:= (n1*ln(sigma))+((2*n1*(k1+1))/(n1-k1-2));
 aic:= temp;
end;
function sic(n1,k1:integer;sigma:double):double;
var temp:double;
begin
 temp:= (ln(sigma))+((k1*ln(n1))/(n1-k1-2));
 sic:=temp;
end;
function hq(n1,k1:integer;sigma:double):double;
var temp:double;
begin
 temp:= (n1*ln(sigma))+((2*n1*k1*ln(ln(n1)))/(n1-k1-2));
 hq:= temp;
```

```
end;
function kic(n1,k1:integer;sigma:double):double;
var temp:double;
begin
  temp:= (n1*ln(sigma))+(n1*ln(n1/(n1-k1)))+((n1*(((n1-k1)*((2*k1)+3))-2))/((n1-k1-2)*(n1-k1)));
  kic:= temp;
end;
begin {Main Program}
 { TODO -oUser -cConsole Main : Insert code here }
  randomize;
  for sam:= 1 to 7 do
  begin
    case sam of
      1 : num:= sample1;  2 : num:= sample2;  3 : num:= sample3;  4 : num:= sample4;
      5 : num:= sample5;   6 : num:= sample6;
      else
        num:= sample7;
      end;
      for para:= 1 to 3 do
      begin
        case para of
          1 : k_para:= inde1+1;   2 : k_para:= inde2+1;   else   k_para:= inde3+1;
          end;
          for varia:= 1 to 4 do
          begin
            case varia of
              1 : sig:= sigma1;   2 : sig:= sigma2;   3 : sig:= sigma3;  4 : sig:= sigma4;   else sig := sigma5;
            end;
            write('sample : ',num,'    ');   write('parameter : ',k_para,'    ');   writeln('variance : ',varia);
            creat_x(num,k_para,x_full);   model_all_out(num,k_para,x_full,model_x);
            total := round(power(2,(k_para-1)));
            for ind:= 1 to k_para do
            begin
              up[ind]:=0;  down[ind]:=0;  if(ind=1)then  up[ind]:= 1  else  up[ind]:= down[ind-1]+1;
              if (ind<k_para) then
                down[ind]:= down[ind-1]+ round(combination((k_para-1),(ind-1)))
              else
                down[ind]:= round(power(2,(k_para-1)));
            end;
            for i1:= 1 to 3 do
              for j1:= 1 to num do
```

```
    x_full1[j1,i1]:= x_full[j1,i1];
for i1:= 1 to 4 do
  count2[i1]:=0;
for repeat1:= 1 to iteration do
begin
  creat_err(num,sig,err_model);
  creat_y(num,k_para,x_full1,err_model,y_model);
  xtx(num,k_para,x_full,xtx_model);
  xty(num,k_para,x_full,y_model,xty_model);
  inv(xtx_model,k_para,inv_xtx_model);
  beta(k_para,inv_xtx_model,xty_model,beta_model);
  mse_full:= 0;
  mse_full:=get_mse(x_full,y_model,beta_model,num,k_para);
  {writeln(' repeat : ',repeat1);}
  for move:= 2 to k_para do
  begin
    num_start:= up[move];
    num_end:= down[move];
    for move1:= num_start to num_end do
    begin
      for column:= 1  to move do
      begin
        for row:= 1 to num do
        begin
          x_model[row,column]:= 0;
          x_model[row,column]:= model_x[move1,row,column];
        end;
      end;
      xtx(num,move,x_model,xtx_model);
      xty(num,move,x_model,y_model,xty_model);
      inv(xtx_model,move,inv_xtx_model);
      beta(move,inv_xtx_model,xty_model,beta_model);
      temp_sse:=get_sse(x_model,y_model,beta_model,num,move);
      temp_mse:=get_mse(x_model,y_model,beta_model,num,move);
      mse[move1]:=temp_mse;
      {writeln('model is : ',move1);}
      for count:= 1 to test do
      begin
        case count of
          1 : seclec[move1,count]:= aic(num,move,temp_sse);
          2 : seclec[move1,count]:= sic(num,move,temp_sse);
```

```
                3 : seclec[move1,count]:= hq(num,move,temp_sse);
                else seclec[move1,count]:= kic(num,move,temp_sse);
              end;
              {writeln(' test ',count,'  ',seclec[move1,count]:7:2);}
            end;
        end;
      end;
      for ch:= 1 to test do
      begin
        for mo:= 2 to total do
        begin
          if (mo = 2)then
            min_mse:= seclec[mo,ch]
          else
          begin
            if(min_mse > seclec[mo,ch])then
              min_mse:= seclec[mo,ch];
          end;
        end;
        for mo:= 2 to total do
        begin
          if (min_mse = seclec[mo,ch])then
          begin
            mse_min[ch,repeat1]:= mse[mo];
          end;
        end;
      end;
    end;
    for ch1:= 1 to test do
    begin
      amse[ch1]:= 0;
      sd[ch1]:= 0;
      diff[ch1]:= 0;
    end;
    for ch2:= 1 to test do
      for mo2:= 1 to iteration do
        amse[ch2]:= mse_min[ch2,mo2]+amse[ch2];
    for ch3:= 1 to test do
    begin
      amse[ch3]:= amse[ch3]/iteration;
    end;
```

```
            for ch4:= 1 to test do
            begin
              for mo4:= 1 to iteration do
              begin
                sd[ch4]:= sd[ch4] + power((mse_min[ch4,mo4]-amse[ch4]),2);
              end;
            end;
            for ch5:= 1 to test do
            begin
              sd[ch5]:= sd[ch5]/iteration;
              sd[ch5]:= sqrt(sd[ch5]);
            end;
            for ch6:= 1 to test do
            begin
              if (ch6=1) then
                amse_min:= amse[ch6]
              else
              begin
                if(amse_min>amse[ch6])then
                  amse_min:=amse[ch6];
              end;
            end;
            for ch7:= 1 to test do
            begin
              diff[ch7]:= (((amse[ch7]-amse_min)/amse_min)*100);
            end;
            writeln;
            for ch7:= 1 to test do
            begin
             write('amse',ch7,' = ',amse[ch7]:5:4);
             write('      sd',ch7,' = ',sd[ch7]:5:4);
             writeln('      diff',ch7,' = ',diff[ch7]:5:4);
            end;
            writeln;
            end;
          readln;
          end;
      end;
   end;
end.
```

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวภัทรสุดา สุดแสน เกิดเมื่อวันที่ 5 ธันวาคม พ.ศ.2522 สำเร็จการศึกษาปริญญาครุศาสตรบัณฑิต เกียรตินิยมอันดับหนึ่ง(ค.บ.) สาขาคณิตศาสตร์ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2544 และเข้าศึกษาต่อในหลักสูตรสถิติศาสตรมหาบัณฑิต ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2546