บทที่ 2

ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้สิ่งที่สนใจศึกษาคือ เปรียบเทียบเกณฑ์การคัดเลือกตัวแบบ 5 เกณฑ์ คือ การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาไคเคะที่ปรับค่าความเสี่ยง โดยใช้ตัวประมาณของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) หรือ AIC_{ri} การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_C ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) หรือ KIC_C การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_I ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) หรือ KIC_I ,การคัดเลือกตัวแบบ Cross-Validation (C_p) ที่ใช้สำหรับตัวอย่างขนาดเล็ก หรือ G_n^{CV} ที่ใช้ $C_n^{(R)}$ การคัดเลือกตัวแบบโดยใช้สถิติทดสอบเอฟบางส่วน (The Partial F-test Statistic) ด้วยวิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method) โดยมีเกณฑ์ที่ใช้ในการเปรียบเทียบ คือ ค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Average of Mean Square Error: AMSE) ซึ่งนำค่าอัตราส่วนผลต่างของค่าเฉลี่ยของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Ratio of Different Average Mean Square Error: RDAMSE) มาใช้ประกอบการตัดสินใจ โดยศึกษาจากการจำลองข้อมูล (Simulation Data) และใช้โปรแกรม S-plus 2000 ในการประมวลผล

<u>เอกสารและรายงานวิจัยที่เกี่ยวข้อง</u>

ในการวิจัยเกี่ยวกับการคัดเลือกตัวแบบ สมการถดถอยเชิงเส้นพหุคูณ ผู้วิจัยได้ศึกษาผลงานและบทความที่เกี่ยวข้องซึ่งสรุปได้ดังต่อไปนี้

อาไคเคะ (Akaike, 1973) ได้เสนอเกณฑ์การคัดเลือกตัวแบบคือ เกณฑ์ AIC ซึ่งสร้างจากการประมาณความแปรปรวนของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) ระหว่าง ตัวแบบที่แท้จริง (True Model) กับ ตัวแบบที่เหมาะสม (Fitted Approximating Model) ที่มีคุณสมบัติไม่เอนเอียง เมื่อขนาดตัวอย่างใหญ่ โดยเกณฑ์คัดเลือกตัวแบบ AIC เลือกตัวแบบที่ให้ค่า AIC ต่ำสุด เป็นตัวแบบที่ดีที่สุด และเกณฑ์ AIC คัดเลือกตัวแบบได้ดีเมื่อตัวอย่างมีขนาดใหญ่เท่านั้น เนื่องจากเกิดความผิดพลาดในการคัดเลือกตัวแบบสูง เมื่อขนาดตัวอย่างไม่ใหญ่พอ

มอลโลว์ (Mollow, 1973) ได้เสนอเกณฑ์คัดเลือกตัวแบบโดยพิจารณาจากข้อมูล n ตัวอย่าง มาแบ่งออกเป็น 2 กลุ่ม ถ้าให้ข้อมูลกลุ่มที่ 1 มีข้อมูล n-1 ตัวอย่าง และ ข้อมูลกลุ่มที่ 2 มีข้อมูล 1 ตัวอย่าง โดยนำข้อมูลกลุ่มที่ 1 มาสร้างสมการการถดถอยหรือตัวแบบ และใช้ข้อมูลกลุ่มที่ 2 คำนวณความคลาดเคลื่อนพยากรณ์ตัวแบบที่ได้จากข้อมูลกลุ่มที่ 1 ทำซ้ำจนครบทุกเซตข้อมูล ซึ่งในที่นี้คือ n ครั้ง แล้วนำค่าคลาดเคลื่อนพยากรณ์ที่ได้มาเฉลี่ย และทำขั้นตอนข้างต้นทุกตัวแบบที่ต้องการคัดเลือก โดยตัวแบบที่ให้ค่าเฉลี่ยความคลาดเคลื่อนพยากรณ์ต่ำสุด เป็นตัวแบบที่ดีที่สุด เรียกเกณฑ์คัดเลือกตัวแบบนี้ว่า Cross-Validation (C_p) และเป็นเกณฑ์ที่มีโอกาสสูงที่ตัวแบบที่คัดเลือกได้เป็นตัวแบบที่มีจำนวนตัวแปรอิสระมากกว่าความเป็นจริง (Overfitting)

เฮอร์วิช และ ไซ (Hurvich and Tsai, 1989) ได้เสนอเกณฑ์คัดเลือกตัวแบบที่พัฒนาต่อเนื่องจากเกณฑ์ AIC ในการวิเคราะห์สมการถดถอยเชิงเส้นพหุคูณ โดยลดคุณสมบัติเอนเอียงของตัวประมาณ AIC เมื่อขนาดตัวอย่างเล็ก เพื่อเพิ่มประสิทธิภาพในการคัดเลือกตัวแบบที่ถูกต้อง เรียกเกณฑ์การคัดเลือกตัวแบบนี้ว่าเกณฑ์ AIC_C

เฮอร์วิช ชัมเว และ ไซ (Hurvich, Shumway and Tsai, 1990) ได้เสนอเกณฑ์คัดเลือกตัวแบบที่พัฒนาต่อเนื่องจากเกณฑ์ AIC_C โดยนำการจำลองข้อมูลมาใช้ปรับความเอนเอียงของตัวประมาณค่า AIC เมื่อขนาดตัวอย่างเล็ก เรียกการคัดเลือกตัวแบบนี้ว่าเกณฑ์ AIC_I และทำการเปรียบเทียบพบว่าเกณฑ์ AIC_I คัดเลือกตัวแบบได้ดีกว่าเกณฑ์ AIC และเกณฑ์ AIC_C เมื่อขนาดตัวอย่างเล็ก

เชา (Jun Shao, 1993) ได้เสนอการคัดเลือกตัวแบบที่นำ Cross-Validation (C_p) ของมอลโลว์ (Mallow, 1973) มาปรับปรุงให้มีประสิทธิภาพมากขึ้นในการคัดเลือกตัวแบบ โดยการเพิ่มฟังก์ชัน C_n ในเกณฑ์คัดเลือกตัวแบบ Cross-Validation (C_p) ของมอลโลว์ (Mallow, 1973) เพื่อให้เกณฑ์คัดเลือกตัวแบบ Cross-Validation (C_p) มีคุณสมบัติคงเส้นคงวา เมื่อขนาดตัวอย่างใหญ่ เพื่อลดโอกาสที่ตัวแบบที่คัดเลือกได้เป็นตัวแบบที่มีจำนวนตัวแปรอิสระมากกว่าความเป็นจริง (Overfitting)

เซง และ โล (Zheng and Wei-Yin Loh, 1995) ได้เสนอเกณฑ์คัดเลือกตัวแบบ Cross-Validation (C_p) ที่พัฒนาต่อเนื่องจาก เชา (Jun Shao, 1993) โดยปรับปรุงฟังก์ชัน C_n เพื่อให้เกณฑ์คัดเลือกตัวแบบ Cross-Validation (C_p) มีประสิทธิภาพในการคัดเลือกตัวแบบที่ถูกต้องเพิ่มขึ้น และเสนอให้ใช้สถิติทดสอบ t ทดสอบสมมติฐานของสัมประสิทธิ์การถดถอย (β_i) ของตัวแบบ เพื่อลดขั้นตอนการคำนวณ และได้เปรียบเทียบกับเกณฑ์ AIC และ Cross-Validation (C_p)

เดิม ปรากฏว่า Cross-Validation ที่พัฒนาใหม่ ให้การคัดเลือกตัวแบบที่ถูกต้องแม่นยำกว่า เมื่อขนาดตัวอย่างใหญ่ และลดขั้นตอนการทดสอบสมมติฐานด้วย

ไบ ราว และ วู (Z.D. Bai, C. Radhakrishna Rao and Yuehua Wu, 1999) ได้เสนอเกณฑ์คัดเลือกตัวแบบ โดยใช้คุณสมบัติความคงเส้นคงวา พัฒนาต่อเนื่องจาก ราว และ วู (C. Radhakrishna Rao and Yuehua Wu, 1989) โดยมีความเห็นว่า ในบางครั้ง ฟังก์ชัน C_n อาจไม่ใช่ค่าที่เหมาะสมสำหรับข้อมูล เนื่องจากเป็นฟังก์ชันที่ขึ้นกับขนาดตัวอย่างอย่างเดียว ดังนั้น ไบ ราว และ วู จึงสร้างฟังก์ชัน C_n โดยหาจากข้อมูลที่นำมาคัดเลือกตัวแบบโดยตรง เมื่อขนาดตัวอย่างเล็ก พบว่าสามารถคัดเลือกตัวแบบได้ดีกว่าเดิม โดยเรียกวิธีการหาฟังก์ชัน C_n ว่า Data-Oriented Penalty

คาวานอฆ์ (Joseph E. Cavanaugh, 1999) เสนอเกณฑ์คัดเลือกตัวแบบในตัวแบบออโตรีเกรซซีพเกาส์เซียน (Gaussian Autoregressive Model) ในการวิเคราะห์อนุกรมเวลา และ ตัวแบบการถดถอยเชิงเส้น (Linear Regression Model) ในการวิเคราะห์การถดถอย โดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลล์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) ไปพัฒนาเกณฑ์ AIC เรียกเกณฑ์การคัดเลือกตัวแบบใหม่นี้ว่าเกณฑ์ KIC และทำการเปรียบเทียบพบว่าเกณฑ์ KIC คัดเลือกตัวแบบได้ดีกว่าเกณฑ์ AIC เมื่อขนาดตัวอย่างใหญ่

คาวานอฆ์ (Joseph E. Cavanaugh, 2001) ได้เสนอเกณฑ์คัดเลือกตัวแบบที่ใช้แนวคิดการแยกออกของความสมมาตรของข้อสนเทศคูลล์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) มาปรับใช้กับเกณฑ์ AIC_C และ AIC_I เพื่อลดโอกาสตัวแบบที่ได้จากการคัดเลือกมีจำนวนตัวแปรอิสระมากกว่าหรือน้อยกว่าความเป็นจริง ในการคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ เรียกเกณฑ์การคัดเลือกตัวแบบใหม่นี้ว่าเกณฑ์ KIC_C และ KIC_I ตามลำดับ

โนดะ เวง ทาคาฮาชิ และ อิโตะ (Kazuo Noda, Jinglong Wang, Rinya Takahashi and Masashi Itoh, 2003) ได้เสนอเกณฑ์คัดเลือกตัวแบบ AIC_{ri} ที่พัฒนาจากเกณฑ์คัดเลือกตัวแบบ AIC ซึ่งสามารถคัดเลือกตัวแบบได้ถูกต้องแม่นยำกว่าเกณฑ์ AIC เมื่อขนาดตัวอย่างเล็ก และให้ค่าความเสี่ยง ที่น้อยกว่าเกณฑ์ AIC

ราว และ วู (C. Radhakrishna Rao and Yuehua Wu, 2005) ได้เสนอเกณฑ์คัดเลือกตัวแบบ Cross-Validation (C_P) ที่มีความแม่นยำในการคัดเลือกตัวแบบที่ถูกต้องมากขึ้น เมื่อตัวอย่างขนาดเล็ก ซึ่งเกณฑ์ใหม่ลดโอกาสตัวแบบที่ได้จากการคัดเลือกมีจำนวนตัวแปรอิสระมากกว่าหรือน้อยกว่าความเป็นจริง ในการคัดเลือกตัวแบบ และสร้างฟังก์ชัน C_n ด้วยวิธี Data-

Oriented Penalty ซึ่งใช้แนวคิดทำนองเดียวกับ ไบ ราว และ วู (Z.D. Bai, C. Radhakrishna Rao and Yuehua Wu, 1999) แต่ปรับลดการคำนวณบางขั้นตอนลง

สถิติที่ใช้ในการวิจัยและทฤษฎีที่เกี่ยวข้อง

งานวิจัยนี้เป็นการศึกษาเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นพหุคูณ ซึ่งมีทฤษฎีที่เกี่ยวข้องในการวิจัยดังนี้

2.1 ตัวแบบการถดถอยเชิงเส้นพหุคูณ (Multiple Linear Regression Model)

มี 2 ตัวแบบ คือ

- ตัวแบบการถดถอยเชิงเส้นพหุคูณที่แท้จริง (True Multiple Linear Regression Model)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}^{+} + \mathbf{e}$$

โดยที่ $\mathbf{Y}$ แทน เวกเตอร์ของตัวแปรตามขนาด $n \times 1$ และ $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}^{+}, \sigma^{+2}I_n)$

$\mathbf{X}$ แทน เมตริกซ์ของตัวแปรอิสระขนาด $n \times k$

$\boldsymbol{\beta}^{+}$ แทน เวกเตอร์ของค่าคงที่การถดถอยและสัมประสิทธิ์การถดถอยที่แท้จริง แทนด้วย $(\beta_0^{+}, \beta_1^{+}, \ldots, \beta_{k-1}^{+})'$

$\mathbf{e}$ แทน เวกเตอร์ของความคลาดเคลื่อนสุ่มขนาด $n \times 1$ และ $e_i \sim N(0, \sigma^{+2})$, $i = 1, 2, \ldots, n$

n แทน ขนาดตัวอย่าง

k แทน จำนวนค่าคงที่การถดถอย กับพารามิเตอร์สัมประสิทธิ์การถดถอยที่แท้จริง

I_n แทน เมตริกซ์เอกลักษณ์ขนาด n

- ตัวแบบการถดถอยเชิงเส้นพหุคูณที่เหมาะสม (Fitted or Candidate Multiple Linear Regression Model)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

โดยที่ $\mathbf{Y}$ แทน เวกเตอร์ของตัวแปรตามขนาด $n \times 1$ และ $\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \sigma^{2}I_n)$

$\mathbf{X}$ แทน เมตริกซ์ของตัวแปรอิสระขนาด $n \times p$

$\boldsymbol{\beta}$ แทน เวกเตอร์ของพารามิเตอร์ แทนด้วย $(\beta_0, \beta_1, \ldots, \beta_{p-1})'$

$\mathbf{e}$ แทน เวกเตอร์ของความคลาดเคลื่อนสุ่มขนาด $n \times 1$ และ $e_i \sim N(0, \sigma^{2})$, $i = 1, 2, \ldots, n$

n แทน ขนาดตัวอย่าง

p แทน จำนวนค่าคงที่การถดถอย กับพารามิเตอร์สัมประสิทธิ์การถดถอย ของตัวแบบที่เหมาะสม

I_n แทน เมตริกซ์เอกลักษณ์ขนาด n

เนื่องจากตัวแบบที่ศึกษาเป็นตัวแบบการถดถอยเชิงเส้นพหุคูณ และค่าคลาดเคลื่อนมีการแจกแจงปกติ ดังนั้น ในการประมาณค่าพารามิเตอร์ของสัมประสิทธิ์การถดถอย ด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) และวิธีการกำลังสองน้อยสุด (Least Likelihood Method) จะได้ตัวประมาณเดียวกัน ซึ่งจะได้ว่า

ตัวประมาณค่าความควรจะเป็นสูงสุด (Maximum Likelihood Estimation: $\hat{\beta}$) คือ

$$\hat{\beta} = (X'X)^{-1}X'Y$$

ซึ่งเป็นตัวประมาณที่ใช้สำหรับ เกณฑ์การคัดเลือกตัวแบบทั้ง 5 วิธี ที่นำมาเปรียบเทียบ เนื่องจาก ตัวประมาณค่าความควรจะเป็นสูงสุด (Maximum Likelihood Estimator) มีคุณสมบัติคงเส้นคงวา (Consistent Estimator) ด้วย

2.2 การประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Estimation)

ตัวประมาณค่าของ $\mathbf{Y}$ ที่มีการแจกแจง $N(\mathbf{X}\boldsymbol{\beta}, \sigma^2 I_n)$ มีพารามิเตอร์ คือ $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_{p-1})'$ และ σ^2 ดังนั้น

$$L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})}$$

$$\begin{aligned} \ln L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}[\mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}] \end{aligned}$$

$$\frac{\partial}{\partial \boldsymbol{\beta}} \ln L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = -\frac{1}{2\sigma^2}[-2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta}] = 0$$

ดังนั้น $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$

$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$ เป็นตัวประมาณค่าวิธีความควรจะเป็นสูงสุดของ $\boldsymbol{\beta}$

$$\frac{\partial}{\partial \sigma^2} \ln L(\boldsymbol{\beta}, \sigma^2; \mathbf{Y}) = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) = 0$$

ดังนั้น $\hat{\sigma}^2 = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})/n$ เป็นตัวประมาณค่าวิธีความควรจะเป็นสูงสุดของ σ^2

2.3 ฟังก์ชันความเสี่ยง (Risk Function)

ฟังก์ชันความเสี่ยงเป็นฟังก์ชันที่นำมาใช้พัฒนาเกณฑ์คัดเลือกตัวแบบ AIC ในการลดความเอนเอียงและความแปรปรวนของการประมาณข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) ฟังก์ชันความเสี่ยง คือ

$$R(\theta^+, IC) \quad = \quad Var_{\theta^+}(IC) + B^2(\theta^+, IC)$$

เมื่อ θ^+ แทน พารามิเตอร์แท้จริง

IC แทน เกณฑ์การคัดเลือกตัวแบบ

$Var_{\theta^+}(IC)$ แทน ความแปรปรวนของเกณฑ์คัดเลือกตัวแบบ

$B(\theta^+, IC)$ แทน ความเอนเอียงของเกณฑ์การคัดเลือกตัวแบบ

2.4 เกณฑ์การคัดเลือกตัวแบบโดยใช้แนวคิดของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information)

ให้ $f\ (\mathbf{Y} | \theta_k)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นของ $\mathbf{Y}$ ของตัวแบบที่แท้จริง (True Model) เมื่อพารามิเตอร์ที่แท้จริง (True Value) คือ $\theta_k = (\beta^+, \sigma^{+^2})'$ และ $f\ (\mathbf{Y} | \theta_p)$ แทน ฟังก์ชันความหนาแน่นของความน่าจะเป็นของ $\mathbf{Y}$ ของตัวแบบที่เหมาะสม (Candidate Model) เมื่อพารามิเตอร์ของตัวแบบที่เหมาะสมคือ $\theta_p = (\beta, \sigma^2)'$ ดังนั้น ข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ คือ

$$I(\theta_k, \theta_p) = E_{\theta_k} \{ \ln [f\ (\mathbf{Y} | \theta_k) / f\ (\mathbf{Y} | \theta_p)] \} \qquad (2.1)$$

ต่อมา ลินนาด และ ซูชินิ (Linhart and Zucchini, 1986) เสนอ ค่าวัดความแตกต่างระหว่าง $f\ (\mathbf{Y} | \theta_k)$ และ $f\ (\mathbf{Y} | \theta_p)$ ซึ่งเรียกว่า Kullback Leibler Discrepancy ดังนี้

$$d(\theta_k, \theta_p) = E_{\theta_k} [-2 \ln f\ (\mathbf{Y} | \theta_p)] \qquad (2.2)$$

ดังนั้น สามารถเขียน ข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ ในรูป Kullback Leibler Discrepancy ได้ดังนี้

$$2I(\theta_k, \theta_p) = d(\theta_k, \theta_p) - d(\theta_k, \theta_k) \qquad (2.3)$$

สามารถตัดเทอม $d(\theta_k, \theta_k)$ เนื่องจากเป็นค่าคงที่ และประมาณค่า θ_p ด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) แทนด้วย $\hat{\theta}_p$ ดังนั้น

$$d(\theta_k, \hat{\theta}_p) = d(\theta_k, \theta_p)\Big|_{\theta_p = \hat{\theta}_p}$$

$$= E_{\theta_k} [\ -2 \ln f\ (\mathbf{Y} | \theta_p)\ ]_{\theta_p = \hat{\theta}_p} \qquad (2.4)$$

อาไคเคะ (Akaike, 1993) เสนอแนวคิดประมาณ $d(\theta_k,\hat{\theta}_p)$ โดยใช้ $-2\ln f(\mathbf{Y}|\hat{\theta}_p)$ ดังนั้น ความเอนเอียงของการประมาณ คือ

$$B_I(p,\theta_k) = E_{\theta_k}[d(\theta_k,\hat{\theta}_p)] - E_{\theta_k}[-2\ln f(\mathbf{Y}|\hat{\theta}_p)] \qquad (2.5)$$

และ อาไคเคะ (Akaike) ได้แสดงให้เห็นว่าเมื่อขนาดตัวอย่างใหญ่ ทำให้ความเอนเอียง (2.5) มีค่าเข้าใกล้ 2 เท่าของจำนวนพารามิเตอร์ที่ไม่ทราบค่าของตัวแบบ

$$B_I(p,\theta_k) \simeq 2M \qquad (2.6)$$

เมื่อ M แทนจำนวนพารามิเตอร์ที่ไม่ทราบค่าของตัวแบบซึ่งในที่นี้คือ p+1 (ผลรวมของค่าคงที่การถดถอยและจำนวนพารามิเตอร์สัมประสิทธิ์การถดถอยเชิงเส้นและความแปรปรวนของประชากร) ดังนั้น เกณฑ์การคัดเลือกตัวแบบโดยข้อสนเทศของอาไคเคะ (Kullback Leibler Information) คือ

$$AIC = -2\ln f(\mathbf{Y}|\hat{\theta}_p) + 2(p+1) \qquad (2.7)$$

จากแนวคิดของอาไคเคะจะเห็นว่า กรณีที่ขนาดตัวอย่างเล็ก ทำให้ความคลาดเคลื่อนใน (2.5) มีค่ามาก

ซูกูรา (Sugiura, 1978) และ เฮอร์วิช และ ไช (Hurvich and Tsai, 1989) ได้เสนอแนวคิดในการปรับความคลาดเคลื่อน (2.5) เพื่อลดความผิดพลาดในการคัดเลือกตัวแบบโดยใช้เกณฑ์การคัดเลือกตัวแบบของอาไคเคะ เมื่อขนาดตัวอย่างเล็ก ซึ่งได้แสดงให้เห็นว่าสมการ (2.5) คือ

$$B_I(p,\theta_k) = \frac{2n(p+1)}{n-p-2} \qquad (2.8)$$

เมื่อ p แทนจำนวนค่าคงที่การถดถอยและพารามิเตอร์ของสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า

และ n แทนขนาดตัวอย่าง

ดังนั้น เกณฑ์การคัดเลือกตัวแบบ คือ

$$AIC_C = -2\ln f(\mathbf{Y}|\hat{\theta}_p) + \frac{2n(p+1)}{n-p-2} \qquad (2.9)$$

เกณฑ์ AIC_C จะเป็นเกณฑ์การคัดเลือกตัวแบบที่ดีกว่าเกณฑ์ AIC เมื่อขนาดตัวอย่างเล็ก

เฮอร์วิช ชัมเว และ ไช (Hurvich, Shumway and Tsai, 1990) ได้เสนอเกณฑ์ AIC_I ของตัวแบบออโตรีเกรซซีพ (Autoregressive Model) ในการวิเคราะห์อนุกรมเวลา (Time Series Analysis) ซึ่งเกณฑ์ดังกล่าวสามารถนำมาประยุกต์ใช้ในสมการถดถอยเชิงเส้นพหุคูณได้ โดยเสนอแนวคิดในการปรับความเอนเอียง (2.5) จากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ดังนี้

$$AIC_I = -2\ln f(\mathbf{Y}|\hat{\theta}_p) + \frac{1}{R}\sum_{j=1}^{R}[d(\theta_k,\hat{\theta}_p(j)) - [-2\ln f(\mathbf{Y}|\hat{\theta}_p(j))]] \qquad (2.10)$$

โดยที่ θ_k แทนเซตของค่าคงที่ ในพารามิเตอร์ที่แท้จริง

R แทนจำนวนการทำซ้ำในการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

เฮอร์วิช ชัมเว และ ไซ (Hurvich, Shumway and Tsai) สรุปว่าเกณฑ์ AIC_I จะเป็นเกณฑ์การคัดเลือกตัวแบบที่ดีกว่าเกณฑ์ AIC_C เมื่อขนาดตัวอย่างเล็ก เมื่อพารามิเตอร์ในเกณฑ์ AIC_C ประมาณโดยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) และ ความคลาดเคลื่อนสุ่มไม่เป็นการแจกแจงปกติ หรือ ฟังก์ชันค่าเฉลี่ยของตัวแปรตามไม่เป็นสมการเชิงเส้นเกณฑ์ AIC_I จะเป็นเกณฑ์การคัดเลือกตัวแบบที่ดีกว่าเกณฑ์ AIC_C และเกณฑ์ AIC

2.5 เกณฑ์การคัดเลือกตัวแบบโดยใช้แนวคิดการแยกออกของความสมมาตรของข้อสนเทศคูลล์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence)

ให้ $f\,(\mathbf{Y}\,|\,\theta_k)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นของ $\mathbf{Y}$ ของตัวแบบที่เหมาะสม เมื่อพารามิเตอร์ของตัวแบบที่เหมาะสม $\theta_k = (\beta^+, \sigma^{+2})'$, $f\,(\mathbf{Y}\,|\,\theta_p)$ แทนฟังก์ชันความหนาแน่นของความน่าจะเป็นของ $\mathbf{Y}$ ของตัวแบบที่เหมาะสม เมื่อพารามิเตอร์ของตัวแบบที่เหมาะสม $\theta_p = (\beta, \sigma^2)'$ และ ข้อสนเทศคูลแบล็ค-ไลท์เบอร์ ใน (2.1)

คาวานอช์ (Cavanaugh, 1999) ได้เสนอเกณฑ์การคัดเลือกตัวแบบ โดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลล์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) หรือ J-Divergence แทนด้วย $J(\theta_k, \theta_p)$

$$J(\theta_k,\theta_p) = E_{\theta_k}\{\ln[f\,(\mathbf{Y}\,|\,\theta_k)/f\,(\mathbf{Y}\,|\,\theta_p)]\} + E_{\theta_p}\{\ln[f\,(\mathbf{Y}\,|\,\theta_p)/f\,(\mathbf{Y}\,|\,\theta_k)]\} \qquad (2.11)$$

เขียนสมการ (2.11) ในรูป Kullback Leibler Discrepancy ได้ดังนี้

$$2J(\theta_k,\theta_p) = \left\{d(\theta_k,\theta_p) - d(\theta_k,\theta_k)\right\} + \left\{d(\theta_p,\theta_k) - d(\theta_p,\theta_p)\right\} \qquad (2.12)$$

เนื่องจาก $d(\theta_k, \theta_k)$ เป็นค่าคงที่ ดังนั้น

$$KS(\theta_k,\theta_p) = d(\theta_k,\theta_p) + \left\{d(\theta_p,\theta_k) - d(\theta_p,\theta_p)\right\} \qquad (2.13)$$

คาวานอช์ (Cavanaugh) เสนอแนวคิดเกณฑ์การคัดเลือกตัวแบบโดยพิจารณาจาก $KS(\theta_k, \theta_p)$ และ พารามิเตอร์ θ_p ประมาณด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) แทนด้วย $\hat{\theta}_p$ ดังนั้น

$$KS(\theta_k,\hat{\theta}_p) = d(\theta_k,\hat{\theta}_p) + \left\{d(\hat{\theta}_p,\theta_k) - d(\hat{\theta}_p,\hat{\theta}_p)\right\} \qquad (2.14)$$

และประมาณ $KS(\theta_k, \theta_p)$ ด้วย $-2\ln f\,(\mathbf{Y}\,|\,\hat{\theta}_p)$ ดังนั้น ความเอนเอียงจากการประมาณ คือ

$$E_{\theta_k}[KS(\theta_k,\hat{\theta}_p)] - E_{\theta_k}[-2\ln f\,(\mathbf{Y}\,|\,\hat{\theta}_p)] = E_{\theta_k}[d(\theta_k,\hat{\theta}_p)] - E_{\theta_k}[-2\ln f\,(\mathbf{Y}\,|\,\theta_p)] + E_{\theta_k}[d(\hat{\theta}_p,\theta_k)] - E_{\theta_k}[d(\hat{\theta}_p,\hat{\theta}_p)] \qquad (2.15)$$

เนื่องจาก (2.5) ดังนั้น กำหนดให้

$$B_2(p,\theta_k) = E_{\theta_k}[d(\hat{\theta}_p,\theta_k)] - E_{\theta_k}[d(\hat{\theta}_p,\hat{\theta}_p)] \tag{2.16}$$

จะเห็นว่าเกณฑ์ AIC , AIC_C และ AIC_I ปรับค่าความเอนเอียงโดยพิจารณาจาก $B_1(p,\theta_k)$ เท่านั้น แต่เกณฑ์ KIC , KIC_C และ KIC_I ปรับค่าความเอนเอียงโดยพิจารณาจาก $B_1(p,\theta_k)$ และ $B_2(p,\theta_k)$ คาวานอธ์ (Cavanaugh) แสดงให้เห็นว่า ถ้านำแนวคิดของเกณฑ์ AIC มาปรับใช้ โดยเพิ่ม $B_2(p,\theta_k)$ และ

$$B_2(p,\theta_k) \approx M \tag{2.17}$$

โดยที่ M แทนจำนวนพารามิเตอร์ที่ไม่ทราบค่าของตัวแบบซึ่งในที่นี้คือ p+1
ดังนั้น เรียกเกณฑ์การคัดเลือกตัวแบบปรับจาก AIC นี้ว่า KIC คือ

$$KIC = -2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p) + 3(p+1) \tag{2.18}$$

เกณฑ์การคัดเลือกตัวแบบ KIC คัดเลือกตัวแบบได้ดีกว่า เมื่อขนาดตัวอย่างเล็ก
ทำนองเดียวกัน เรียกเกณฑ์การคัดเลือกตัวแบบปรับจาก AIC_C นี้ว่า KIC_C คือ

$$KIC_C = -2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p) + \frac{2n(p+1)}{(n-p-2)} + n\ln\left(\frac{n}{n-p}\right) + \frac{n}{n-p} \tag{2.19}$$

และ เรียกเกณฑ์การคัดเลือกตัวแบบปรับจาก AIC_I นี้ว่า KIC_I คือ

$$\begin{aligned} KIC_I = {} & -2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p) + \frac{1}{R}\sum_{j=1}^{R}\left\{\left[n\ln\sigma^{+^2} - n\ln\hat{\sigma}^2(j) + \frac{n\sigma^{+^2}}{\hat{\sigma}^2(j)}\right]\right. \\ & + [(\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\,/\,\hat{\sigma}^2(j)] \\ & \left. + [n\hat{\sigma}^2(j)/\sigma^{+^2} + (\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\,/\,\sigma^{+^2}] - 2n\right\} \end{aligned} \tag{2.20}$$

2.6 เกณฑ์การคัดเลือกตัวแบบ

แบ่งเป็น 5 เกณฑ์ ดังนี้

2.6.1 การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาไคเคะที่ปรับค่าความเสี่ยง โดยใช้ตัวประมาณของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) หรือ AIC_{ri}

2.6.2 การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_C ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบลค-ไลท์เบอร์ (Kullback's Symmetric Divergence)หรือ KIC_C

2.6.3 การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_I ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบลค-ไลท์เบอร์ (Kullback's Symmetric Divergence)หรือ KIC_I

2.6.4 การคัดเลือกตัวแบบ Cross-Validation (C_p) ที่ใช้สำหรับขนาดตัวอย่างเล็ก หรือ G_n^{CV} ที่ใช้ $C_n^{(R)}$

2.6.5 การคัดเลือกตัวแบบโดยสถิติทดสอบเอฟบางส่วน (The Partial F-test Statistic) ด้วยวิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method)

มีรายละเอียดดังต่อไปนี้

2.6.1 <u>การคัดเลือกตัวแบบโดยใช้เกณฑ์ข้อสนเทศของอาไคเคะที่ปรับค่าความเสี่ยง โดยใช้ตัวประมาณของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback Leibler Information) หรือ AIC_{ri}</u>

เป็นเกณฑ์คัดเลือกตัวแบบที่พัฒนาขึ้นโดย โนดะ และคณะ (Noda and et.al., 2003) โดยลดความเอนเอียง และความแปรปรวนของตัวประมาณ AIC กับ ข้อสนเทศคูลแบล็ค-ไลท์เบอร์ (Kullback Leibler Information) เกณฑ์ใหม่ที่ได้คือ AIC_{ri} โดยเกณฑ์ AIC_{ri} เป็นเกณฑ์ที่ลดความผิดพลาดในการคัดเลือกตัวแบบ เมื่อขนาดตัวอย่างเล็ก และให้ค่าความเสี่ยง (Risk Function) น้อยกว่าเกณฑ์ AIC

เกณฑ์การคัดเลือกตัวแบบ ที่เสนอโดย โนดะ และ คณะ (Noda and et.al., 2003) คือ ภายใต้เงื่อนไข Noda and et al. (1996)

$$AIC_{ri} = n \ln \hat{\sigma}^2 + \frac{n(n+p)}{n-p-2} - \frac{2np}{(n-p-2)(n-p)} n^{\alpha-\varepsilon}$$

โดยที่ $\hat{\sigma}^2$ แทนค่าความแปรปรวนของตัวแบบที่มีจำนวนค่าคงที่การถดถอยและพารามิเตอร์สัมประสิทธิ์ความถดถอยเท่ากับ p ด้วยวิธีความควรจะเป็นสูงสุด

n แทนจำนวนขนาดตัวอย่าง

p แทนจำนวนค่าคงที่การถดถอยกับพารามิเตอร์สัมประสิทธิ์การถดถอยของตัวแบบ

ε แทนค่าคงที่ใดๆ , $0< \varepsilon <\alpha$ และ α แทนค่าคงที่ใดๆ , $0<\alpha<1$

<u>ขั้นตอนในการคัดเลือกตัวแบบ</u>

1. กำหนดตัวแบบที่เป็นไปได้ทั้งหมด
2. ตัวแบบที่กำหนดจากขั้นตอนที่ 1 ประมาณค่าพารามิเตอร์ $\hat{\boldsymbol{\beta}}$ ด้วยวิธีความควรจะเป็นสูงสุดของแต่ละตัวแบบ ดังนี้

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'Y}$$

3. คำนวณค่าความแปรปรวนของแต่ละตัวแบบ ดังนี้

$$\hat{\sigma}^2 = (\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}})/n$$

4. คำนวณ AIC_{ri} แต่ละตัวแบบ ดังนี้

$$\mathrm{AIC}_{ri} = n\ln\hat{\sigma}^2 + \frac{n(n+p)}{n-p-2} - \frac{2np}{(n-p-2)(n-p)}n^{\alpha-\varepsilon}$$

เมื่อ n แทนจำนวนขนาดตัวอย่าง

p แทนจำนวนค่าคงที่การถดถอยกับพารามิเตอร์สัมประสิทธิ์การถดถอยของตัวแบบ

ε แทนค่าคงที่ใดๆ , $0< \varepsilon <\alpha$ และ α แทนค่าคงที่ใดๆ , $0<\alpha<1$

5. เปรียบเทียบค่า AIC_{ri} แต่ละตัวแบบ โดยเลือกตัวแบบที่ให้ค่า AIC_{ri} ต่ำที่สุด เป็นตัวแบบที่ดีที่สุด

2.6.2 <u>การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ $\mathrm{AIC_C}$ ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบลค-ไลท์เบอร์ (Kullback's Symmetric Divergence)หรือ $\mathrm{KIC_C}$</u>

เป็นเกณฑ์การคัดเลือกตัวแบบที่พัฒนาขึ้นโดย คาวานอซ์ (Joseph E. Cavanaugh, 2001) โดยปรับความเอนเอียงของการประมาณข้อสนเทศคูลแบล็ค-ไลท์เบอร์ โดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) เป็นการพัฒนาเกณฑ์เพื่อลดการเกิด Underfitting และ Overfitting ในการคัดเลือกตัวแบบ ซึ่งลดความผิดพลาดในการคัดเลือกตัวแบบในเชิงความน่าจะเป็น เมื่อขนาดตัวอย่างเล็ก

จากสมการถดถอยเชิงเส้นพหุคูณ เวกเตอร์ตัวแปรตาม ($\mathbf{Y}$) มีการแจกแจงแบบ $N(\mathbf{X}\boldsymbol{\beta},\sigma^2)$ ดังนั้น

ให้ $f(\mathbf{Y}\,|\,\theta_p)$ เป็นฟังก์ชันความหนาแน่นของความน่าจะเป็นของ $\mathbf{Y}$

$$f(\mathbf{Y}\,|\,\theta_p) = \frac{1}{(2\pi\sigma^2)^{n/2}}\,\mathbf{e}^{-\frac{1}{2\sigma^2}(Y-X\beta)'(Y-X\beta)}$$

พิจารณาเกณฑ์การคัดเลือกตัวแบบ $\mathrm{KIC_C}$

$$\mathrm{KIC_C} = -2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p) + B_1(p,\theta_k) + B_2(p,\theta_k)$$

เมื่อ $\hat{\theta}_p = (\hat{\beta}, \hat{\sigma}^2)$ เป็นพารามิเตอร์ที่ประมาณด้วยวิธีความควรจะเป็นสูงสุด

และ $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_{p-1})'$

$$B_1(p,\theta_k) = E_{\theta_k}[d(\theta_k,\hat{\theta}_p)] - E_{\theta_k}[-2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p)]$$
$$= \frac{2n(p+1)}{n-p-2}$$

และ $$B_2(p,\theta_k) = E_{\theta_k}[d(\hat{\theta}_p,\theta_k)] - E_{\theta_k}[d(\hat{\theta}_p,\hat{\theta}_p)]$$
$$= n\ln\frac{n}{n-p} + \frac{n}{n-p}$$

แสดงการพิสูจน์ $B_1(p,\theta_k)$, $B_2(p,\theta_k)$ ในภาคผนวก ค

ดังนั้น

$$\begin{aligned}
\mathrm{KIC_C} &= -2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p) + \left[\frac{2n(p+1)}{n-p-2}\right] + \left[n\ln\frac{n}{n-p} + \frac{n}{n-p}\right] \\
&= n\ln 2\pi\hat{\sigma}^2 + n + \left[\frac{2n(p+1)}{n-p-2}\right] + \left[n\ln\frac{n}{n-p} + \frac{n}{n-p}\right] \\
&= n\ln 2\pi + n\ln\hat{\sigma}^2 + n + \left[\frac{2n(p+1)}{n-p-2}\right] + \left[n\ln\frac{n}{n-p} + \frac{n}{n-p}\right] \\
&= n\ln 2\pi + n\ln\hat{\sigma}^2 + n\ln\frac{n}{n-p} + \frac{n[(n+p)(n-p)+(n-p-2)]}{(n-p-2)(n-p)}
\end{aligned}$$

เนื่องจาก $n\ln 2\pi$ เป็นค่าคงที่ ดังนั้น

เกณฑ์คัดเลือกตัวแบบที่เสนอโดย คาวานอซ์ (Joseph E. Cavanaugh, 2001)

คือ

$$\mathrm{KIC_C} = n\ln\hat{\sigma}^2 + n\ln\frac{n}{n-p} + \frac{n[(n+p)(n-p)+(n-p-2)]}{(n-p-2)(n-p)}$$

โดยที่ $\hat{\sigma}^2$ แทนค่าความแปรปรวนของตัวแบบที่มีจำนวนค่าคงที่การถดถอยและพารามิเตอร์สัมประสิทธิ์ความถดถอยเท่ากับ p ด้วยวิธีความควรจะเป็นสูงสุด

n แทนจำนวนขนาดตัวอย่าง

p แทนจำนวนค่าคงที่การถดถอยกับพารามิเตอร์สัมประสิทธิ์การถดถอยของตัวแบบ

ขั้นตอนในการคัดเลือกตัวแบบ

1. กำหนดตัวแบบที่เป็นไปได้ทั้งหมด

2. ตัวแบบที่กำหนดจากขั้นตอนที่ 1 ทำการประมาณค่าพารามิเตอร์ $\hat{\beta}$ ด้วยวิธีความควรจะเป็นสูงสุด ของแต่ละตัวแบบ ดังนี้

$$\hat{\beta} = (X'X)^{-1}X'Y$$

3. คำนวณค่าความแปรปรวนของแต่ละตัวแบบ ดังนี้

$$\hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/n$$

4. คำนวณ KIC_C แต่ละตัวแบบ ดังนี้

$$KIC_C = n\ln\hat{\sigma}^2 + n\ln\frac{n}{n-p} + \frac{n[(n+p)(n-p)+(n-p-2)]}{(n-p-2)(n-p)}$$

เมื่อ n แทนจำนวนขนาดตัวอย่าง

p แทนจำนวนค่าคงที่การถดถอยกับพารามิเตอร์สัมประสิทธิ์การถดถอยของตัวแบบ

5. เปรียบเทียบค่า KIC_C แต่ละตัวแบบ โดยเลือกตัวแบบที่ให้ค่า KIC_C ต่ำที่สุด เป็นตัวแบบที่ดีที่สุด

2.6.3 การคัดเลือกตัวแบบที่พัฒนาเกณฑ์ AIC_I ที่ลดความเอนเอียงโดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบลค-ไลท์เบอร์ (Kullback's Symmetric Divergence)หรือ KIC_C

เป็นเกณฑ์การคัดเลือกตัวแบบที่พัฒนาขึ้นโดย คาวานอซ์ (Joseph E. Cavanaugh, 2001) โดยปรับความเอนเอียงของการประมาณข้อสนเทศคูลแบล็ค-ไลท์เบอร์ โดยใช้การแยกออกของความสมมาตรของข้อสนเทศคูลส์แบล็ค-ไลท์เบอร์ (Kullback's Symmetric Divergence) เป็นการพัฒนาเกณฑ์เพื่อปรับลดการเกิด Underfitting และ Overfitting ซึ่งลดความผิดพลาดในการคัดเลือกตัวแบบในเชิงการทดลองสุ่มด้วยเทคนิคมอนติคาร์โล เมื่อขนาดตัวอย่างเล็ก

จากสมการการถดถอยเชิงเส้นพหุคูณ เวกเตอร์ตัวแปรตาม (Y) มีการแจกแจงแบบ $N(X\beta, \sigma^2)$ และให้ $f(Y \mid \theta_p)$ เป็นฟังก์ชันความหนาแน่นความน่าจะเป็นของ Y ดังนั้น เกณฑ์การคัดเลือกตัวแบบ KIC_I คือ

$$KIC_I = -2\ln f(Y \mid \hat{\theta}_p) + B_1(p, \theta_k) + B_2(p, \theta_k)$$

โดยที่ $\hat{\theta}_p = (\hat{\beta}, \hat{\sigma}^2)$ เป็นพารามิเตอร์ที่ประมาณด้วยวิธีความควรจะเป็นสูงสุด

$$B_1(p, \theta_k) = E_{\theta_k}[d(\theta_k, \hat{\theta}_p)] - E_{\theta_k}[-2\ln f(Y \mid \hat{\theta}_p)]$$

$$B_2(p,\theta_k) = E_{\theta_k}[d(\hat{\theta}_p,\theta_k)] - E_{\theta_k}[d(\hat{\theta}_p,\hat{\theta}_p)]$$

และ $B_1(p,\theta_k), B_2(p,\theta_k)$ ได้จากการทดลองสุ่มด้วยเทคนิคมอนติคาร์โล

ดังนั้น

$$KIC_I = -2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p)$$
$$+\frac{1}{R}\sum_{j=1}^{R}[d(\theta_k,\hat{\theta}_p(j)+d(\hat{\theta}_p(j),\theta_k)-d(\hat{\theta}_p(j),\hat{\theta}_p(j))-[-2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p)(j)]]$$

โดยที่ $d(\theta_k,\hat{\theta}_p(j)) = E_{\theta_k}\{-2\ln f(\mathbf{Y}\,|\,\theta_p(j)\}_{\theta_p(j)=\hat{\theta}_p(j)}$

$$= n\ln 2\pi + n\ln\hat{\sigma}^2(j) + \frac{n\sigma^{+^2}}{\hat{\sigma}^2(j)} + (\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\ /\hat{\sigma}^2(j)$$

$$d(\hat{\theta}_p(j),\theta_k) = E_{\theta_p(j)}\{-2\ln f(\mathbf{Y}\,|\,\theta_k)\}_{\theta_p(j)=\hat{\theta}_p(j)}$$

$$= n\ln 2\pi + n\ln\sigma^{+^2} + \frac{n\hat{\sigma}^2(j)}{\sigma^{+^2}} + (\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\ /\sigma^{+^2}$$

$$d(\hat{\theta}_p(j),\hat{\theta}_p(j)) = E_{\theta_p(j)}\{-2\ln f(\mathbf{Y}\,|\,\theta_p(j)\}_{\theta_p(j)=\hat{\theta}_p(j)}$$

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

$$-2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p(j)) = n\ln 2\pi + n\ln\hat{\sigma}^2(j) + n$$

ดังนั้น

$$KIC_I = -2\ln f(\mathbf{Y}\,|\,\hat{\theta}_p) + \frac{1}{R}\sum_{j=1}^{R}\{[n\ln\sigma^{+^2} - n\ln\hat{\sigma}^2(j) + n\sigma^{+^2}/\hat{\sigma}^2(j)]$$
$$+ [(\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\ /\hat{\sigma}^2(j)]$$
$$+ [n\hat{\sigma}^2(j)/\sigma^{+^2} + (\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\ /\sigma^{+^2}] - 2n\}$$
$$= n\ln 2\pi + n\ln\hat{\sigma}^2 + n + \frac{1}{R}\sum_{j=1}^{R}\{[n\ln\sigma^{+^2} - n\ln\hat{\sigma}^2(j) + n\sigma^{+^2}/\hat{\sigma}^2(j)]$$
$$+ [(\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\ /\hat{\sigma}^2(j)]$$
$$+ [n\hat{\sigma}^2(j)/\sigma^{+^2} + (\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\ /\sigma^{+^2}] - 2n\}$$

เนื่องจาก $n\ln 2\pi$ เป็นค่าคงที่ ซึ่งไม่มีผลในการคำนวณ ดังนั้น

เกณฑ์คัดเลือกตัวแบบที่เสนอโดย คาวานอซ์ (Joseph E. Cavanaugh, 2001) คือ

$$KIC_I = n\ln\hat{\sigma}^2 + n + \frac{1}{R}\sum_{j=1}^{R}\{[n\ln\sigma^{+^2} - n\ln\hat{\sigma}^2(j) + n\sigma^{+^2}/\hat{\sigma}^2(j)]$$
$$+ [(\hat{\beta}(j)-\beta^+)'\mathbf{X}'\mathbf{X}(\hat{\beta}(j)-\beta^+)\ /\hat{\sigma}^2(j)]$$

$$+ [n\hat{\sigma}^2(j)/\sigma^{+^2} + (\hat{\beta}(j)-\beta^+)'X'X(\hat{\beta}(j)-\beta^+)\ /\sigma^{+^2}] - 2n\}$$

โดยที่ $\mathbf{X}$ แทนเมตริกซ์ของตัวแปรอิสระขนาด nxp

β^+ แทนเวกเตอร์ค่าพารามิเตอร์ที่แท้จริง

σ^{+^2} แทนเวกเตอร์ค่าความแปรปรวนของตัวแบบที่แท้จริง

$\hat{\beta}(j)$ แทนค่าประมาณพารามิเตอร์ของตัวแบบจากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลครั้งที่ j

$\hat{\sigma}^2(j)$ แทนเวกเตอร์ค่าประมาณความแปรปรวนของตัวแบบที่ได้จากการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลครั้งที่ j

$\hat{\sigma}^2$ แทนเวกเตอร์ค่าประมาณความแปรปรวนของตัวแบบที่มีจำนวนค่าคงที่การถดถอยกับพารามิเตอร์สัมประสิทธิ์การถดถอยเท่ากับ p ด้วยวิธีความควรจะเป็นสูงสุด

p แทนจำนวนค่าคงที่การถดถอยกับพารามิเตอร์สัมประสิทธิ์การถดถอยของตัวแบบ

n แทนจำนวนขนาดตัวอย่าง

R แทนจำนวนครั้งในการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

และ β^+, σ^{+^2} แทนพารามิเตอร์ของตัวแบบที่กำหนดขึ้นตามความสะดวกของผู้ใช้ นิยมกำหนดให้เป็น $\beta^+ = (0,0,...,0)$ และ $\sigma^{+^2} = 1$

หรือสามารถเขียน KIC_I ในรูป

$$KIC_I = n\ln\hat{\sigma}^2 + n + \frac{1}{R}\sum_{j=1}^{R}\{[-n\ln\hat{\sigma}^2(j) + n/\hat{\sigma}^2(j)]$$
$$+ [(X\hat{\beta}(j))'(X\hat{\beta}(j))/\hat{\sigma}^2(j)] + [n\hat{\sigma}^2(j) + (X\hat{\beta}(j))'(X\hat{\beta}(j))] - 2n\}$$

<u>ขั้นตอนในการคัดเลือกตัวแบบ</u>

1. กำหนดตัวแบบที่เป็นไปได้ทั้งหมด
2. ตัวแบบที่กำหนดจากขั้นตอนที่ 1 ทำการประมาณพารามิเตอร์ β ด้วยวิธีความควรจะเป็นสูงสุด (Maximum Likelihood Method) ของแต่ละตัวแบบ ดังนี้

$$\hat{\beta} = (X'X)^{-1}X'Y$$

3. คำนวณค่าความแปรปรวนของแต่ละตัวแบบ ดังนี้

$$\hat{\sigma}^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})/n$$

4. จำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ของเทอมปรับค่าความเอนเอียง ด้วยการกำหนดค่า $\beta^{+} = (0,0,...,0)$ และ $\sigma^{+^2} = 1$ ดังนั้น เทอมที่นำมาจำลองค่า คือ

เมื่อ R เป็นจำนวนการสุ่มซ้ำ

$$\mathrm{Bias}_{adj} = \frac{1}{R}\sum_{j=1}^{R}\{[-n\ln\hat{\sigma}^2(j) + n/\hat{\sigma}^2(j)] + [(\mathbf{X}\hat{\beta}(j))'(\mathbf{X}\hat{\beta}(j))/\hat{\sigma}^2(j)] + [n\hat{\sigma}^2(j) + (\mathbf{X}\hat{\beta}(j))'(\mathbf{X}\hat{\beta}(j))] - 2n\}$$

4. คำนวณ KIC_I แต่ละตัวแบบ ดังนี้

$$KIC_I = n\ln\hat{\sigma}^2 + n + \mathrm{Bias}_{adj}$$

5. เปรียบเทียบค่า KIC_I แต่ละตัวแบบ โดยเลือกตัวแบบที่ให้ค่า KIC_I ต่ำที่สุด เป็นตัวแบบที่ดีที่สุด

2.6.4 <u>การคัดเลือกตัวแบบ Cross-Validation (C_p) ที่ใช้สำหรับขนาดตัวอย่างเล็ก หรือ G_n^{CV} ที่ใช้ $C_n^{(R)}$</u>

เป็นเกณฑ์การคัดเลือกตัวแบบที่พัฒนาขึ้นโดย ราว และ วู (Rao and Wu, 2005) ใช้แนวคิดเกณฑ์การคัดเลือกตัวแบบมีคุณสมบัติคงเส้นคงวา (Strongly Consistent) เมื่อขนาดตัวอย่างเล็ก ซึ่งนำแนวคิด Cross-Validation เดิม กับ ทฤษฎีคงเส้นคงวาของเกณฑ์การคัดเลือกตัวแบบ (Consistent Information Theoretic Criterion) มาสร้างเกณฑ์ Cross-Validation ใหม่ ซึ่งเป็นเกณฑ์ที่ลดปัญหาการเกิด Overfitting และ Underfitting ในการคัดเลือกตัวแบบ เพราะวิธีการคัดเลือกตัวแบบ Cross-Validation เดิม เป็นเกณฑ์ที่เกิดปัญหา Overfitting และ ทฤษฎีคงเส้นคงวาของเกณฑ์การคัดเลือกตัวแบบ (Consistent Information Theoretic Criterion) เป็นเกณฑ์ที่เกิดปัญหา Underfitting และเกณฑ์ที่เสนอใหม่นี้ได้นำแนวคิดของ ไบ ราว และ วู ในปี ค.ศ. 1999 มาสร้างฟังก์ชัน C_n แต่ลดบางขั้นตอนลง เพื่อลดการคำนวณฟังก์ชัน C_n ทำให้เกณฑ์ Cross-Validation มีคุณสมบัติคงเส้นคงวาที่มีความเหมาะสมสำหรับข้อมูลมากขึ้น

เกณฑ์การคัดเลือกตัวแบบ ที่เสนอโดย ราว และ วู (Rao and Wu, 2005) คือ

$$G_n^{CV} = \sum_{i=1}^{n}(y_i - \hat{y}_{-i})^2 + pC_n^{(R)}$$

เมื่อ y_i แทน ค่าสังเกตของข้อมูลที่ i

$\hat{y}_{-i}$ แทน ค่าพยากรณ์ของตัวแบบของเมตริกซ์ตัวแปรอิสระ $\mathbf{X}$ ที่ตัดแถวที่ i

$C_n^{(R)}$ แทน ฟังก์ชันที่หาโดยวิธี Data-Oriented Penalty

p แทน จำนวนค่าคงที่การถดถอย กับ พารามิเตอร์สัมประสิทธิ์การถดถอยของตัวแบบ

<u>ขั้นตอนในการคัดเลือกตัวแบบ</u>

1. กำหนดตัวแบบที่เป็นไปได้ทั้งหมด

2. คำนวณหา $\hat{y}_{-i}$ เมื่อ i=1,2,…,n แต่ละตัวแบบ

หรือ $\hat{y}_{-i} = X_i\hat{\beta}_{-i}$ และ $\hat{\beta}_{-i} = (X_i'X_i)^{-1}(X_i'y_{-i})$

เมื่อ $\hat{y}_{-i}$ แทน ค่าพยากรณ์ของตัวแบบของเมตริกซ์ตัวแปรอิสระ $\mathbf{X}$ ที่ตัดแถวที่ i

และ $\hat{\beta}_{-i}$ แทน ตัวประมาณกำลังสองน้อยที่สุด จากข้อมูล

$\{ (y_1,X_1), \ldots, (y_{i-1},X_{i-1}), (y_{i+1},X_{i+1}), \ldots, (y_n,X_n) \}$

3. หาฟังก์ชัน C_n โดยวิธี Data-Oriented Penalty ($C_n^{(R)}$)

4. คำนวณ $G_n^{CV} = \sum_{i=1}^{n}(y_i - \hat{y}_{-i})^2 + pC_n^{(R)}$ แต่ละตัวแบบ

5. เปรียบเทียบค่า G_n^{CV} ที่ใช้ $C_n^{(R)}$ แต่ละตัวแบบ โดยเลือกที่ให้ค่า G_n^{CV} ที่ใช้ $C_n^{(R)}$ ต่ำสุด เป็นตัวแบบที่ดีที่สุด

<u>ขั้นตอนการหา ฟังก์ชัน $C_n^{(R)}$ ด้วยวิธี Data-Oriented Penalty</u>

1.คำนวณตัวประมาณสัมประสิทธิ์การถดถอยด้วยวิธีกำลังสองน้อยสุด ของตัวแบบเต็มรูป (Full Model) ซึ่งประกอบไปด้วยตัวแปรอิสระทุกตัว ดังนี้

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}(\mathbf{X'Y})$$

2. คำนวณค่าความคลาดเคลื่อน ดังนี้

$$\hat{\mathbf{e}} = \mathbf{Y} - \mathbf{X}\tilde{\boldsymbol{\beta}}$$

3. สร้างสัมประสิทธิ์การถดถอยของตัวแบบใหม่ คือ

$$\mathbf{u} = \mathbf{X}\bar{\boldsymbol{\beta}} + \hat{\mathbf{e}}$$

โดยที่ $\bar{\boldsymbol{\beta}} = (\bar{\beta}_0, \ldots, \bar{\beta}_{p-1})'$

และ $\bar{\beta}_i = \begin{cases} \tilde{\beta}_i & \text{if } |\tilde{\beta}_i| \geq k \\ k(\text{sign}\tilde{\beta}_i) & \text{if } 0 < |\tilde{\beta}_i| < k \\ k & \text{if } |\tilde{\beta}_i| = 0 \end{cases}$ เมื่อ i = 0,…,p-1

เมื่อ k เป็น ค่าสัมประสิทธิ์การถดถอยที่มีค่าน้อยที่สุด และมีความเหมาะสมสำหรับข้อมูล

4. คำนวณ $D_{n0} = \sum_{i=1}^{n}(u_i - \hat{u}_{-i})^2$, i=1,2,…n

เมื่อ $\hat{u}_{-i}$ เป็นค่าพยากรณ์ของตัวแบบใหม่ จากเมตริกซ์ตัวแปรอิสระ $\mathbf{X}$ ตัดแถวที่ i

5. คำนวณ $D_n(h) = \sum_{i=1}^{n}(u_i - \hat{u}_{-i,-h})^2$ เมื่อ i=1,2,…n , h = 1,2,…,p-1

เมื่อ $\hat{u}_{-i,-h}$ เป็นค่าพยากรณ์ของตัวแบบใหม่ที่ตัดข้อมูลชุดที่ i และ ตัวแปรอิสระตัวที่ h

6. คำนวณ $\Delta_n(h) = D_n(h) - D_{n0}$

7. เลือกฟังก์ชัน C_n จาก $C_n^{(R)}$ ดังนี้

$C_n^{(R)}$ = ค่ามากที่สุดของ {ค่าต่ำสุดของ $\Delta_n(h), h = 1, 2, ..., p-1, \tau$} $/(1+\sqrt{\lfloor 0.0001n \rfloor})$

เมื่อ τ = 10(Q_3 ของ $|\hat{\beta}_i|$, $1 \le i \le p-1$) หรือ เป็นค่าที่กำหนดขึ้นตามความเหมาะสม (กำหนดขึ้นเพื่อลดการเกิด Overfitting และไม่ควรมีค่าน้อยเกินไป เพราะทำให้เกิด Underfitting)

และ $\lfloor 0.0001n \rfloor$ แทนค่าของจำนวนเต็มที่ได้จากการคำนวณ 0.0001n

2.6.5 <u>การคัดเลือกตัวแบบโดยใช้สถิติทดสอบเอฟบางส่วน (The partial F-test statistic) ด้วยวิธีการถดถอยแบบขั้นบันได (Stepwise Regression Method)</u>

วิธีการเลือกตัวแปรอิสระแบบขั้นบันได เป็นวิธีการเลือกตัวแปรอิสระที่นำวิธีการเลือกตัวแปรอิสระแบบไปข้างหน้ากับวิธีการกำจัดตัวแปรอิสระแบบถอยหลังมาผสมกัน ซึ่งขั้นแรกจะคัดเลือกตัวแปรที่มีความสัมพันธ์กับตัวแปรตามสูงสุดเข้าไปในสมการถดถอย แล้วจึงพิจารณาเลือกตัวแปรอิสระที่มีความสัมพันธ์เชิงส่วนกับตัวแปรตามมากที่สุดเข้าในสมการถดถอย และพร้อมกันนั้นก็พิจารณาว่าเมื่อตัวแปรอิสระตัวใหม่เข้าในสมการถดถอยแล้วตัวแปรอิสระเดิมควรจะอยู่ในสมการถดถอยหรือไม่ และพิจารณาเช่นนี้ทุกครั้งที่มีตัวแปรอิสระใหม่เข้าในสมการถดถอย

<u>ขั้นตอนในการคัดเลือกตัวแปร</u>

1. คำนวณค่าสัมประสิทธิ์สหสัมพันธ์ (r_{xy}) ระหว่างตัวแปรตามกับตัวแปรอิสระแต่ละตัว

2. เลือกตัวแปรอิสระที่มีค่าสัมประสิทธิ์สหสัมพันธ์สูงสุด สมมติให้ตัวแปรอิสระที่ถูกเลือกคือ X_j สมการถดถอยที่ได้คือ

$$Y = \beta_0 + \beta_j X_j \,;\, j = 1, 2, ..., p-1$$

3. ทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์การถดถอยของ X_j โดยใช้ Partial F-Test ถ้าค่าสถิติน้อยกว่าค่าวิกฤตซึ่งทำให้ยอมรับสมมติฐาน แสดงว่าไม่มีตัวแปรอิสระใดถูกเลือกเข้าในสมการถดถอย ถ้าสมมติฐานแสดงว่าตัวแปรอิสระ X_j ควรอยู่ในสมการถดถอย แล้วดำเนินการคัดเลือกตัวแปรอิสระต่อไป

4. คำนวณค่าสัมประสิทธิ์สหสัมพันธ์เชิงส่วนระหว่างตัวแปรตามกับตัวแปรอิสระที่เหลือแต่ละตัวโดยกำหนดให้ตัวแปรอิสระที่อยู่ในสมการมีค่าคงที่

5. เลือกตัวแปรอิสระที่มีค่าสัมประสิทธิ์สหสัมพันธ์เชิงส่วนสูงสุดเข้าในสมการถดถอย สมมติให้ตัวแปรอิสระที่ถูกเลือกคือ X_k

6. ทดสอบสมมติฐานเกี่ยวกับสัมประสิทธิ์การถดถอยของ X_j กับ X_k โดยใช้ Partial F-Test ถ้าค่าสถิติมากกว่าค่าวิกฤตตัวแปรอิสระนั้นจะถูกคัดออกจากสมการถดถอย

7. เลือกตัวแปรอิสระตัวต่อไปโดยทำตามขั้นตอนที่ 4-6 โดยในขั้นตอนที่ 6 จะต้องพิจารณาค่า Partial F-Test ของตัวแปรอิสระทุกตัวที่อยู่ในสมการถดถอยนั้น การคัดเลือกตัวแปรอิสระจะดำเนินการต่อไปจนกระทั่งไม่มีตัวแปรอิสระใดถูกนำเข้าหรือคัดออกจากสมการถดถอยอีก จึงหยุดการคัดเลือกตัวแปรอิสระซึ่งแสดงว่าได้สมการถดถอยที่เหมาะสม