การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่

วสันต์ เดือนแจ้ง

ดุษฎีนิพนธ์นี้เป็นส่วนของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา
วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา
สิงหาคม 2559

คณะกรรมการควบคุมดุษฎีนิพนธ์และคณะกรรมการสอบดุษฎีนิพนธ์ ได้พิจารณาดุษฎีนิพนธ์ของ วสันต์ เดือนแจ้ง ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา ของมหาวิทยาลัยบูรพาได้

คณะกรรมการควบคุมดุษฎีนิพนธ์

....อาจารย์ที่ปรึกษาหลัก

(รองศาสตราจารย์ ดร.เสรี ชัดแช้ม)

....อาจารย์ที่ปรึกษาร่วม

(ดร.พัชรี วงษ์เกษม)

คณะกรรมการสอบดุษฎีนิพนธ์

....ประธาน

(ผู้ช่วยศาสตราจารย์ ดร.กุศยา ปลั่งพงษ์พันธ์)

....กรรมการ

(รองศาสตราจารย์ ดร.เสรี ชัดแช้ม)

....กรรมการ

(ดร.พัชรี วงษ์เกษม)

....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.พูลพงศ์ สุขสว่าง)

วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา อนุมัติให้รับดุษฎีนิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต สาขาวิชาการวิจัยและสถิติทางวิทยาการปัญญา ของมหาวิทยาลัยบูรพา

....คณบดีวิทยาลัยวิทยาการวิจัย

(ผู้ช่วยศาสตราจารย์ ดร.สุชาดา กรเพชรปาณี) และวิทยาการปัญญา

วันที่ 1 เดือน สิงหาคม พ.ศ. 2559

ประกาศคุณูปการ

ดุษฎีนิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดีด้วยความกรุณาจากคณะกรรมการควบคุมดุษฎีนิพนธ์ รองศาสตราจารย์ ดร.เสรี ชัดแช้ม อาจารย์ที่ปรึกษาหลัก และ ดร.พัชรี วงษ์เกษม อาจารย์ที่ปรึกษาร่วมที่กรุณาให้คำแนะนำ คำปรึกษา ตลอดจนช่วยแก้ไขข้อบกพร่องต่าง ๆ เป็นอย่างดียิ่ง จนดุษฎีนิพนธ์ฉบับนี้เสร็จสมบูรณ์ ผู้วิจัยขอกราบพระคุณด้วยความรู้สึกซาบซึ้งและสำนึกในพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ผู้วิจัยขอขอบคุณ วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา ที่สนับสนุนการศึกษาต่อในระดับปริญญาเอก

ผู้วิจัยขอกราบขอบพระคุณครูอาจารย์ทุกท่าน ที่ได้ให้โอกาสทางการศึกษาและประสิทธิ์ประสาทวิชาความรู้ให้แก่ผู้วิจัยจนกระทั่งสำเร็จการศึกษา

ผู้วิจัยขอระลึกถึงพระคุณ คุณพ่อ คุณแม่ พี่ น้องและเพื่อน ๆ ทุกคน รวมทั้งครอบครัว ที่สนับสนุนผู้วิจัยทั้งด้านวิชาการและเป็นกำลังใจให้ผู้วิจัยเสมอมาจนกระทั่งสำเร็จการศึกษา

ท้ายนี้คุณค่าและประโยชน์ของดุษฎีนิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นกตัญญูกตเวทีแด่บุพการี บูรพาจารย์ และผู้มีพระคุณทุกท่านทั้งในอดีตและปัจจุบัน ที่ทำให้ข้าพเจ้าเป็นผู้มีการศึกษาและประสบความสำเร็จมาจนถึงทุกวันนี้

วสันต์ เดือนแจ้ง

51810448: สาขาวิชา: การวิจัยและสถิติทางวิทยาการปัญญา;
ปร.ด. (การวิจัยและสถิติทางวิทยาการปัญญา)
คำสำคัญ: ฟังก์ชันเคอร์เนล/ วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ/ การแจกแจงแบบโคชี/ การแจกแจงแบบปกติ/ ราคาข้าว

วสันต์ เดือนแจ้ง: การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ (USING AN ADJUSTED REPRODUCING KERNEL HILBERT SPACE TO FORECAST THE PRICE OF THAI RICE) คณะกรรมการควบคุมดุษฎีนิพนธ์: เสรี ชัดแช้ม, ค.ด., พัชรี วงษ์เกษม, ปร.ด. 248 หน้า. ปี พ.ศ. 2559.

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์ 2) เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ โดยใช้แบนวิดจ์ 2 วิธี คือ วิธีที่ 1 Rules of Thumb และวิธีที่ 2 Silverman's Rules of Thumb ด้วยขนาดตัวอย่าง 6 ขนาด (5, 10, 15, 30, 50, 100) กำหนดค่าอัตสหสัมพันธ์ 9 ระดับ (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) โดยวิธีการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล และ 3) เพื่อพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าด้วยตัวแบบ ARIMA

ผลการวิจัย ปรากฏว่า

1. ตัวประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ คือ

$$\hat{y} = \frac{3}{4nh^3}\sum(h^2-(x-x_i)^2)I + \rho\varepsilon_{i-1} + a_i$$

2. ผลการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) แบนวิดจ์วิธี Rules of Thumb ปรากฏว่า มี 52 สถานการณ์ที่วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)

3. ตัวแบบที่ใช้ในการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทย คือ ARIMA (1, 0, 0) ผลการพยากรณ์ชี้ให้เห็นว่า สัมประสิทธิ์การทำนายสามารถอธิบายความแปรผันได้ร้อยละ 60.30

51810448: MAJOR: RESEARCH AND STATISTICS IN COGNITIVE SCIENCE;
Ph.D. (RESEARCH AND STATISTICS IN COGNITIVE SCIENCES)
KEYWORDS: KERNEL FUNCTION/ REPRODUCING KERNEL HILBERT SPACE/ CAUCHY DISTRIBUTION/ NORMAL DISTRIBUTION/ GRAIN PRICE

WASAN DUEANCHAENG: USING AN ADJUSTED REPRODUCING KERNEL HILBERT SPACE TO FORECAST THE PRICE OF THAI RICE. ADVISORY COMMITTEE: SEREE CHADCHAM, Ph.D., PATCHAREE WONGKASEM, Ph.D. 248 P. 2016.

The objectives of this study were 1) to develop the parameter estimation in regression using the adjusted Reproducing Kernel Hilbert Space in the case of random error with autocorrelation; 2) to compare the efficiency of the adjusted Reproducing Kernel Hilbert Space and the Reproducing Kernel Hilbert Space by Ferraty (2007) under 108 situations, including two bandwidth methods: (2a) rules of thumb, and (2b) Silverman's rules of thumb with sample sizes of 5, 10, 15, 30, 50, and 100, and the 9-level autocorrelations as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, and 0.9 using Monte Carlo simulation; and 3), to forecast the price of 15% rice grain in Thailand using the ARIMA model.

The results were as follows:

1. The parameter estimate generated by the adjusted Reproducing Kernel Hilbert Space was:

$$\hat{y} = \frac{3}{4nh^3} \sum (h^2 - (x - x_i)^2) I + \rho \varepsilon_{i-1} + a_i$$

2. The comparative results between the adjusted Reproducing Kernel Hilbert Space and the Reproducing Kernel Hilbert Space by Ferraty (2007) with the bandwidth method of the rules of thumb revealed that there were 51 situations where the adjusted Reproducing Kernel Hilbert Space had lower mean square error (MSE) and lower standard error than the Reproducing Kernel Hilbert Space by Ferraty (2007).

3. The model used to forecast the price of 15% rice grain in Thailand was ARIMA (1, 0, 0). The result suggested that the forecasting coefficient could explain up to 60.30 % of the variability.

สารบัญ

	หน้า
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
สารบัญ	ฉ
สารบัญตาราง	ฌ
สารบัญภาพ	ฎ
บทที่	
1 บทนำ	1
ความเป็นมาและความสำคัญของปัญหา	1
วัตถุประสงค์ของการวิจัย	7
กรอบแนวคิดในการวิจัย	8
สมมติฐานของการวิจัย	10
ขอบเขตของการวิจัย	11
ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย	14
นิยามศัพท์เฉพาะ	14
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	16
ตอนที่ 1 การประมาณค่าของฟังก์ชันเคอร์เนล	17
ตอนที่ 2 การประมาณค่าของฟังก์ชัน Reproducing Kernel Hilbert Space	36
ตอนที่ 3 การพยากรณ์ด้วยอนุกรมเวลาและการสร้างตัวแบบ ARIMA	42
ตอนที่ 4 การตรวจสอบความเหมาะสมของอนุกรมเวลา	51
ตอนที่ 5 อัตตสหสัมพันธ์ในข้อมูลอนุกรมเวลา	57
ตอนที่ 6 การจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล	66
ตอนที่ 7 งานวิจัยที่เกี่ยวข้องกับ Reproducing Kernel Hilbert Space	72
ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับการพยากรณ์ราคาข้าวเปลือกเจ้า	80
3 วิธีดำเนินการวิจัย	84
ตอนที่ 1 การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์	84

สารบัญ (ต่อ)

บทที่ หน้า

ตอนที่ 2 การเปรียบเทียบค่าความคลาดเคลื่อนจากการพยากรณ์ข้อมูลล่วงหน้าโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007).. 88

ตอนที่ 3 การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA........................ 93

4 ผลการวิจัย.. 97

ตอนที่ 1 ผลการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์ 98

ตอนที่ 2 ผลการเปรียบเทียบประสิทธิภาพของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007).. 100

ตอนที่ 3 ผลการพยากรณ์ราคาข้าวเปลือกเจ้าของไทยล่วงหน้าด้วยตัวแบบ ARIMA.... 133

5 สรุปและอภิปรายผล.. 140

สรุปผลการวิจัย.. 140

อภิปรายผล.. 142

ข้อเสนอแนะ.. 146

บรรณานุกรม.. 147

ภาคผนวก.. 153

ภาคผนวก ก Code โปรแกรม R ในการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน (Gaussian) ในกรณีการแจกแจงแบบปกติ.. 154

ภาคผนวก ข Code โปรแกรม R ในการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานีชนิคอฟ (Epanechnikov) ในกรณีการแจกแจงแบบโคชี.. 184

สารบัญ (ต่อ)

บทที่ หน้า

ภาคผนวก ค ผลการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) ด้วยการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ การแจกแจงแบบโคชี และวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) การแจกแจงแบบปกติ.................................. 213

ภาคผนวก ง ข้อมูลราคาข้าวที่ปรับโดยวิธีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ และ OUTPUT จากโปรแกรมสำเร็จรูป... 222

ประวัติย่อของผู้วิจัย... 247

สารบัญตาราง

ตารางที่ หน้า

2-1 รูปแบบของฟังก์ชันเคอร์เนล.. 23
2-2 ค่าประสิทธิภาพสัมพัทธ์ของฟังก์ชันเคอร์เนล.. 26
4-1 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=5 ที่มีแบนวิดจ์แบบ Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 100
4-2 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=10 ที่มีแบนวิดจ์แบบ Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 103
4-3 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=15 ที่มีแบนวิดจ์แบบ Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 105
4-4 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=30 ที่มีแบนวิดจ์แบบ Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 108
4-5 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=50 ที่มีแบนวิดจ์แบบ Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 110
4-6 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=100 ที่มีแบนวิดจ์แบบ Rule of thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 113
4-7 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=5 ที่มีแบนวิดจ์แบบ Silverman’s Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 117
4-8 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=10 ที่มีแบนวิดจ์แบบ Silverman’s Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 119

สารบัญตาราง (ต่อ)

ตารางที่ หน้า

4-9 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=15 ที่มีแบนวิดจ์แบบ Silverman’s Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 119

4-10 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=30 ที่มีแบนวิดจ์แบบ Silverman’s Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 124

4-11 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=50 ที่มีแบนวิดจ์แบบ Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 127

4-12 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (*MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (*Std. error*) เมื่อ *n*=100 ที่มีแบนวิดจ์แบบ Rule of Thumb จำแนกตามการแจกแจงแบบโคชีและการแจกแจงแบบปกติ.. 129

4-13 ราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ. 2552-2556 กับค่าพยากรณ์ราคาโดยตัวแบบ ARIMA (1, 0, 0).. 135

4-14 ราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ. 2557 กับค่าพยากรณ์ราคาโดยตัวแบบ ARIMA (1, 0, 0).. 136

4-15 ตัวแปร Estimate ค่าความคลาดเคลื่อนมาตรฐาน *t* และ p-value................................. 138

4-16 Model Fit Statistic และ Ljung-box... 138

สารบัญภาพ

ภาพที่ หน้า

1-1 กรอบแนวคิดในการวิจัย........ 10

2-1 การประมาณค่าแบบฮิตโตแกรม กับค่าประมาณการความหนาแน่นแบบเคอร์เนล........ 21

2-2 ฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงโคซี........ 30

3-1 แผนผังขั้นตอนวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่........ 86

3-2 แผนผังการพัฒนาวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS)........ 87

3-3 แผนผังของการพัฒนาวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) ของ Ferraty (2007)........ 90

3-4 แผนผังการพัฒนาวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่........ 91

3-5 แผนผังการเปรียบเทียบวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)........ 92

3-6 แผนผังการพัฒนาวิธีการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยด้วยตัวแบบ ARIMA..... 93

3-7 แผนผังขั้นตอนการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ.2557........ 96

4-1 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี n=5........ 102

4-2 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี n=5........ 102

4-3 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี n=10........ 104

4-4 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี n=10........ 105

4-5 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี n=15........ 107

4-6 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี n=15........ 107

สารบัญภาพ (ต่อ)

ภาพที่ หน้า

4-7 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=30...................................... 109

4-8 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=30.............................. 110

4-9 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=50...................................... 112

4-10 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=50.............................. 112

4-11 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=100.................................. 114

4-12 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=100........................... 115

4-13 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=5....................................... 118

4-14 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=5............................... 119

4-15 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=10.................................... 121

4-16 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=10.............................. 121

4-17 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=15.................................... 123

4-18 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=15.............................. 124

สารบัญภาพ (ต่อ)

ภาพที่ หน้า

4-19 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=30.................................... 126

4-20 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=30............................... 126

4-21 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=50..................................... 128

4-22 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=50............................... 129

4-23 การเปรียบเทียบค่า *MSE* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี *n*=100................................... 131

4-24 การเปรียบเทียบค่า *Std. error* กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี n=100............................ 131

4-25 การเปรียบเทียบราคาข้าวเปลือกเจ้า 15% ของไทย ปี พ.ศ. 2552-2556 กับค่าพยากรณ์ราคาข้าวจากตัวแบบ ARIMA (1,0,0).. 136

4-26 การเปรียบเทียบราคาข้าวเปลือกเจ้า 15% ของไทย ปี พ.ศ. 2557 กับค่าพยากรณ์ราคาข้าวจากตัวแบบ ARIMA (1,0,0).. 137

บทที่ 1
บทนำ

ความเป็นมาและความสำคัญของปัญหา

การพยากรณ์ราคาผลผลิตสินค้าเกษตร เป็นการพยากรณ์เชิงปริมาณโดยการสร้างแบบจำลองทางเศรษฐมิติ (Econometric Model) ด้วยวิธีวิเคราะห์การถดถอย (Regression Analysis) เป็นส่วนใหญ่ทั้งพืชและปศุสัตว์ ซึ่งจะแสดงความสัมพันธ์เชิงเศรษฐศาสตร์ระหว่างตัวแปรที่เกี่ยวข้องกันแบบจำลองที่สร้างขึ้นมีหลายรูปแบบ มีทั้งแบบจำลองที่เป็นเชิงเส้นและไม่เป็นเชิงเส้น มีบางส่วนในการพยากรณ์ระดับประเทศใช้การวิเคราะห์ข้อมูลแบบอนุกรมเวลา เช่น Exponential Smoothing Method ความเคลื่อนไหวของราคาข้าว และพยากรณ์ราคาข้าวแบบรายเดือน โดยใช้เทคนิคการพยากรณ์อนุกรมเวลาด้วยวิธีแยกส่วนประกอบ (Decomposition Method) การศึกษาความเคลื่อนไหวของราคาข้าว และพยากรณ์ราคาข้าวแบบรายเดือน โดยใช้เทคนิคการพยากรณ์อนุกรมเวลาด้วยวิธีการพยากรณ์โดยการปรับเรียบ (Smoothing Method) ได้แก่ วิธี Single Exponential Smoothing, วิธี Holt’s Linear Method: Double Exponential Smoothing และวิธี Winters’ Method: Triple Exponential Smoothing เป็นต้น การจัดทำแบบจำลองเพื่อการพยากรณ์ให้ได้ผล และได้ประสิทธิภาพ จำเป็นต้องศึกษาว่ารูปแบบใดที่มีความเหมาะสม หรือมีความสัมพันธ์เป็นไปตามข้อกำหนดทางสถิติ ซึ่งโดยทั่วไปแบบจำลองที่จัดทำขึ้นเป็นแบบจำลองที่นักเศรษฐศาสตร์ได้ศึกษา และมีการอธิบายเหตุผลสนับสนุนในทุกรูปแบบ

การวิเคราะห์การถดถอย เป็นวิธีการทางสถิติที่ใช้ในการวิเคราะห์ข้อมูลวิธีหนึ่งที่ สามารถนำไปใช้ในการประมาณหรือการพยากรณ์ที่เกี่ยวกับตัวแปร โดยกำหนดให้ตัวแปรตัวหนึ่งที่ต้องการศึกษาเป็นตัวแปรตาม (Dependent Variable) และอาศัยความรู้เกี่ยวกับค่าของตัวแปรอื่น ๆ ที่เกี่ยวข้องซึ่งเป็นตัวแปรที่เหลือเป็นตัวแปรต้น (Independent Variable) เพื่อศึกษาว่าตัวแปรต้นมีอิทธิพลอย่างไรต่อตัวแปรตาม หรือตัวแปรต้นมีส่วนในการอธิบายความผันแปรของตัวแปรตามได้อย่างไร หรือตัวแปรตามขึ้นอยู่กับตัวแปรต้นอย่างไร การอธิบายลักษณะความสัมพันธ์จะอธิบายได้ด้วยรูปแบบการถดถอย (Regression Model) เช่นรูปแบบการถดถอยที่เป็นเส้นตรง (Linear) แบบเส้นโค้ง (Nonlinear) เป็นต้น (ทรงศิริ แต้สมบัติ, 2549, หน้า 74) ในการวิเคราะห์การถดถอยเชิงเส้น (Linear Regression Analysis) ตัวแปรที่เกี่ยวข้องกับตัวแปรที่ต้องการศึกษา หรือตัวแปรอิสระ มีเพียงหนึ่งตัวแปร และลักษณะความสัมพันธ์เป็นเส้นตรง การวิเคราะห์ในลักษณะนี้เรียกว่า การวิเคราะห์การถดถอยอย่างง่าย (Simple Linear Regression) ถ้าในบางกรณีตัวแปรที่ต้องการศึกษามีตั้งแต่สองตัวขึ้นไป การวิเคราะห์ลักษณะนี้เรียกว่า การวิเคราะห์การถดถอยเชิงเส้นแบบพหุ

(Multiple Linear Regression) ซึ่งรูปแบบการวิเคราะห์การถดถอยเชิงเส้นอย่างง่าย มีรูปแบบทั่วไป คือ (Bower man & O'Connell, 1993, p. 235)

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

โดยที่ Y_t คือ ค่าของตัวแปรตาม Y ณ เวลาที่ t

X_t คือ ค่าของตัวแปรอิสระ X ณ เวลาที่ t

β_0, β_1 คือ พารามิเตอร์หรือค่าสัมประสิทธิ์การถดถอยที่ไม่ทราบค่า (Unknown Parameter)

ε_t คือ ความคลาดเคลื่อนสุ่ม (Random Error)

n คือ ขนาดตัวอย่าง

การวิเคราะห์การถดถอยเริ่มจากการสร้างสมการถดถอย ซึ่งจะทำได้โดยการประมาณค่าพารามิเตอร์หรือค่าสัมประสิทธิ์การถดถอย β_1 วิธีการประมาณค่าที่นิยมใช้กันอย่างแพร่หลาย ได้แก่ วิธีกำลังสองน้อยที่สุด (Ordinary Least Squares: OLS) ซึ่งเป็นวิธีที่ให้ตัวประมาณของสัมประสิทธิ์การถดถอยที่มีคุณสมบัติเป็นตัวประมาณเชิงเส้นที่ไม่เอนเอียงที่ดีที่สุด (Best Linear Unbiased Estimator: BLUE) ตามทฤษฎีเกาส์-มาร์คอฟ (Gauss-Markov Theorem) (Wonnacott & Wonnacott, 1970, p. 67) ข้อตกลงเบื้องต้น (Assumption) ของความคลาดเคลื่อน (ε_t) ดังนี้

1. ค่าคาดหวัง (Expected Value) ของ ε_t เป็น 0 หรือ $E(\varepsilon_t) = 0$
2. ความแปรปรวนของ ε_t คงที่ หรือ $\sigma^2_{\varepsilon_t} = \sigma^2$
3. ค่า ε_t และ ε_s ไม่มีสหสัมพันธ์กันหรือความแปรปรวนร่วมเป็น 0 หรือ $E(\varepsilon_t \varepsilon_s) = 0$ สำหรับ $t \neq s$

ในทางปฏิบัติส่วนใหญ่ปรากฏว่า ข้อมูลที่นำมาใช้ในการวิเคราะห์ไม่เป็นไปตามข้อตกลงเบื้องต้นบางข้อ งานวิจัยนี้จะศึกษาเฉพาะกรณีที่ข้อตกลงเบื้องต้นข้อที่ 3 ไม่เป็นจริง กล่าวคือ ค่าความคลาดเคลื่อนมีสหสัมพันธ์กันคือ $E(\varepsilon_t \varepsilon_s) \neq 0$ สำหรับ $t \neq s$ ซึ่งการเกิดเหตุการณ์เช่นนี้เรียกว่า อัตตสหสัมพันธ์ของความคลาดเคลื่อน (Serial Correlation หรือ Autocorrelation) เมื่อเกิดปัญหาอัตตสหสัมพันธ์ของความคลาดเคลื่อน ในการวิเคราะห์ข้อมูลด้วยการวิเคราะห์การถดถอยจะมีผลต่อการอนุมานโดยค่าสัมประสิทธิ์ของสมการถดถอยที่ประมาณได้จะขาดคุณสมบัติของการเป็นตัวประมาณค่าที่ดี และไม่เหมาะที่จะนำไปใช้ในการประมาณค่าพารามิเตอร์และการพยากรณ์ คือจะทำให้การพยากรณ์ด้วยตัวแบบที่มีปัญหาดังกล่าวขาดความแม่นยำและความน่าเชื่อถือ การเกิดอัตตสหสัมพันธ์ของความคลาดเคลื่อนมักจะเกิดในกรณีที่ค่าของตัวแปรตามและตัวแปรต้นมีการเก็บตามช่วงเวลา ที่เรียกว่าอนุกรมเวลา (Box et al., 1994, p. 96) ข้อมูลอนุกรมเวลา (Time Series Data) ส่วนใหญ่ข้อมูลที่นำมาวิเคราะห์จะเป็นข้อมูลทางด้าน

เศรษฐศาสตร์และด้านธุรกิจ และลักษณะของอัตตสหสัมพันธ์ของความคลาดเคลื่อนที่พบโดยทั่วไปจะเป็นอัตตสหสัมพันธ์ที่หนึ่งช่วงเวลา หรือ อันดับที่ 1 (Autocorrelationที่ lag 1) หมายถึง ค่าความคลาดเคลื่อน ณ เวลา *t* มีสหสัมพันธ์กับค่าความคลาดเคลื่อน ณ เวลา *t-1* ซึ่งเกิดขึ้นกับข้อมูลอนุกรมเวลารายปี เช่น ราคาข้าวของปีนี้ขึ้นอยู่กับราคาข้าวของปีที่ผ่านมาอันเป็นผลสืบเนื่องมาจากการเกิดสถานการณ์แบบใยแมงมุม (Cobweb Phenomena) ในทางเศรษฐศาสตร์การเกิดสถานการณ์เช่นนี้ คือ การตัดสินใจผลิตในปีที่ *t* ขึ้นอยู่กับราคาในปีที่ *t-1* และถ้าพบว่าราคาของพืชในปีที่ *t* ลดต่ำลงกว่าปีที่ *t-1* เกษตรกรจะตัดสินใจปลูกพืชหรือผลิตพืชในปีที่ *t+1* น้อยกว่าปีที่ *t* จึงเป็นเหตุให้การผลิตในช่วงเวลา *t+1* ลดลง หรือถ้าเกษตรกรผลิตมากเกินไปในปีที่ *t* เกษตรกรจะลดการผลิตในปีที่ *t+1* ลง และอาจจะเพิ่มการผลิตในปีต่อไป ซึ่งนำไปสู่การเกิดสถานการณ์แบบใยแมงมุม ก่อให้เกิดปัญหาอัตตสหสัมพันธ์ของความคลาดเคลื่อนอันดับที่ 1 อย่างไรก็ตามอัตตสหสัมพันธ์ของความคลาดเคลื่อนเกิดขึ้นได้ทั้งในข้อมูลอนุกรมเวลาและข้อมูลภาคตัดขวาง (Cross Section Data) แต่มักเกิดขึ้นกับข้อมูลอนุกรมเวลามากกว่า

การประมาณความหนาแน่นแบบนอนพาราเมตริกซ์ ถือเป็นเครื่องมือสำคัญที่จะช่วยในการวิเคราะห์และนำเสนอโครงสร้างของข้อมูลได้อย่างมีประสิทธิภาพ มีข้อตกลงเบื้องต้นน้อยกว่าการประมาณความหนาแน่นแบบอิงพารามิเตอร์ สามารถทำได้หลายวิธี วิธีฮิสโตแกรมถือเป็นวิธีที่ง่ายและเก่าแก่ที่สุด นิยมใช้อย่างกว้างขวางเพื่อแสดงรูปร่างของฟังก์ชันความหนาแน่นของความน่าจะเป็น (Want & Jones, 1995, p. 5) วิธีฮิสโตแกรมจะแสดงรูปร่างของฟังก์ชันความหนาแน่นของความน่าจะเป็น เป็นกราฟแท่งความถี่หลายแท่งเรียงต่อกัน การประมาณความหนาแน่นโดยใช้ฮิสโตแกรมจะต้องคำนึงถึงการเลือกจุดเริ่มต้นและการกำหนดความกว้างของช่วงเนื่องจากส่งผลต่อรูปร่างของฮิสโตแกรมที่จะแตกต่างกันออกไป การใช้ฮิสโตแกรมมักเกิดข้อบกพร่องในการประมาณความหนาแน่น คือ อาจเกิดความไม่ต่อเนื่องของกราฟฟังก์ชันความหนาแน่นของความน่าจะเป็น สาเหตุเกิดจากการมีจำนวนช่วงมากเกินไป หรือกำหนดค่าความกว้างของช่วงน้อยเกินไป ทำให้มีบางช่วงไม่มีค่าสังเกตตกอยู่และรูปร่างของฮิสโตแกรมที่ได้จะขึ้นอยู่กับการเลือกจุดเริ่มต้นของฮิสโตแกรม และช่วงความกว้างของอันตรภาคชั้น จากปัญหาดังกล่าวในเวลาต่อมาได้มีผู้คิดวิธีการต่าง ๆ หลายวิธีเพื่อประมาณความหนาแน่นของความน่าจะเป็น เช่น วิธีตัวประมาณอย่างง่าย (Fix & Hodges, 1951, pp. 233-247) และวิธีประมาณความหนาแน่นแบบเคอร์เนล (Kernel Density Estimation) ซึ่งถูกเสนอครั้งแรกโดย Fix and Hodge (1951, quoted in Silverman, 1986, pp. 1-2) ตามมาด้วยวิธี Silverman (1986) (Rosenblatt (1956) Parzen (1962) วิธีฟังก์ชันถ่วงน้ำหนักทั่วไป (Whittle, 1958) วิธีเคอร์เนลแปรผัน (Bremen, Merisel & Purcell, 1977, pp. 135-144)

การประมาณความหนาแน่นแบบนอนพาราเมตริก ไม่ต้องคำนึงถึงค่าพารามิเตอร์ใด ๆ กล่าวคือ การประมาณความหนาแน่นแบบนอนพาราเมตริกเป็นการประมาณฟังก์ชันความหนาแน่นจากชุดข้อมูลโดยตรง การประมาณความหนาแน่นแบบนอนพาราเมตริกด้วยวิธีเคอร์เนล (Kernel Density Estimation) เป็นวิธีหนึ่งที่สำคัญการประมาณความหนาแน่นแบบเคอร์เนล ถูกเสนอขึ้นอย่างเป็นทางการในปี ค.ศ. 1956 โดย Murray Rosenblatt การประมาณค่าความหนาแน่นแบบเคอร์เนลเป็นวิธีนิยมอย่างแพร่หลายและเป็นวิธีที่มีประสิทธิภาพมาก เพราะคุณสมบัติทางคณิตศาสตร์ที่เกี่ยวข้อง สามารถทำความเข้าใจได้ไม่ยาก ปรับเปลี่ยนรูปแบบได้ง่าย เป็นพื้นฐานสำคัญที่นำไปสู่การศึกษาวิธีการประมาณความหนาแน่นแบบไม่อิงพารามิเตอร์วิธีอื่น การประมาณความหนาแน่นด้วยวิธีเคอร์เนลเป็นการถ่วงน้ำหนักข้อมูลของตัวแปรสุ่มผ่านฟังก์ชันถ่วงน้ำหนักในการวิเคราะห์แบบตัวแปรเดียว (Univariate) ซึ่งถือว่าเป็นวิธีการประมาณค่าความหนาแน่นแบบเคอร์เนลเป็นวิธีการนอนพาราเมตริก เพื่อการประมาณค่าฟังก์ชันความหนาแน่นความน่าจะเป็นของตัวแปรสุ่ม การประมาณค่าความหนาแน่นแบบเคอร์เนล (Kernel Density Estimation) เป็นพื้นฐานของการปรับข้อมูลให้เรียบ เมื่อมีการอ้างอิงสรุปไปถึงประชากร โดยมีตัวปรับแบบเคอร์เนล (Kernel Smoother) ที่เรียกว่า แบนวิดจ์ (Bandwidth) ซึ่งมีรูปแบบทั่วไปดังนี้ (Wolfgang, 1990, p. 20)

$$K_h(\mathrm{x}) = \frac{1}{h} K(\frac{x}{h})$$

โดยที่ h แทน ค่าแบนวิดจ์ (Bandwidth) มีค่าเป็นจำนวนจริงบวก

$K(\mathrm{x})$ แทน ฟังก์ชันเคอร์เนล

ถ้า $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มจากประชากรที่มีฟังก์ชันการแจกแจงความหนาแน่นของความน่าจะเป็น $f(\mathrm{x})$ ที่ไม่ทราบตัวแบบ ตัวประมาณค่าความหนาแน่นแบบเคอร์เนล (Kernel Estimator) ของ $f(\mathrm{x})$ คือ

$$\hat{f}_h(\mathrm{x}) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

โดยที่ h แทน ค่าแบนวิดจ์ (Bandwidth) มีค่าเป็นจำนวนจริงบวก

x แทน ตัวแปรที่ไม่ทราบค่า

X_i แทน ตัวแปรสุ่ม X โดยที่ i มีค่าตั้งแต่ $i = 1, \ldots, n$

K แทน ฟังก์ชันเคอร์เนล (Kernel Function) ในกรณีนี้เรียก K เป็นฟังก์ชันเคอร์เนลลำดับที่ K

การวิเคราะห์แบบรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (Reproducing Kernel Hilbert Space: RKHS) ช่วยให้สามารถสร้างแบบจำลองทางสถิติได้ดี การวิเคราะห์ข้อมูลที่มีความเจาะจงหลายรูปแบบ มีกรอบแนวคิดสำหรับการประมาณค่าฟังก์ชันที่ดี มีความยืดหยุ่นในการสร้างตัวแบบ

ทางสถิติที่มีข้อมูลทั้งโดยตรงและโดยอ้อม ตามแนวการวิเคราะห์ฟังก์ชันแบบรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ สามารถจัดการกับปัญหาการกระจายของข้อมูลที่หลากหลายโดยใช้ทฤษฎีต่าง ๆ เช่น ทฤษฎีของเกาส์ใช้สกุลเลขยกกำลังทั่วไป (General Exponential Families) การประมาณค่าที่มีค่าความแกร่ง (Robust Estimation) การสังเกตแบบช่วง (Interval Observations) ในแบบจำลองที่มีข้อจำกัดต่าง ๆ เช่น ความเป็นบวก (Positivity), การนูนออก (Convexity) ข้อจำกัดความไม่เสมอภาคเชิงเส้นอื่น (Other Inequality Constraint) อีกทั้งยังสามารถจัดการกับค่าจากการสังเกตที่ซับซ้อนที่เกี่ยวกับอนุพันธ์ (Derivatives) ปริพันธ์ (Integrals) และฟังก์ชันเชิงเส้นอื่นที่มีกรอบแนวคิดสำหรับการรวมข้อมูลประเภทต่าง ๆ เช่น ค่าจากการสังเกตเฉลี่ย ณ บริเวณ หรือช่วงเวลาที่ผิดปกติ และไม่สอดคล้องกัน ประกอบกับสามารถประมาณค่าปริพันธ์และอนุพันธ์ของฟังก์ชันได้

จากงานวิจัยของ Heng Lian (2007) ที่ได้ศึกษาและพัฒนาเรื่องโมเดลการทำงานแบบไม่เป็นเชิงเส้นสำหรับการตอบสนองการทำงานในพื้นที่ทำซ้ำของรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) ผลการศึกษาปรากฏว่า ค่าความคลาดเคลื่อนของวิธีการประมาณค่าโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) ได้รับการปรับเข้าสู่ปกติเป็น 1.00 และความคลาดเคลื่อนของค่าประมาณแบบ RKHS ที่ปรับแล้วที่ประมาณด้วยวิธี RKHS สำหรับตัวแบบเชิงเส้นนั้น การใช้วิธีการประมาณค่าแบบเชิงเส้น ค่าประมาณเชิงเส้นที่ได้มีค่าความคลาดเคลื่อนของการประมาณค่าด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) ที่ได้มีค่าต่ำกว่า 1.00 ส่วนตัวแบบไม่เป็นเชิงเส้น การใช้วิธีการประมาณค่าแบบเชิงเส้นในการประมาณค่าความคลาดเคลื่อนทำให้ได้ประสิทธิภาพไม่ดีมากนักเมื่อเทียบกับค่าประมาณอื่น ๆ ประสิทธิภาพของการประมาณแบบ RKHS และแบบ RKHS ที่ปรับแล้วมีประสิทธิภาพไม่ต่างกัน แต่การประมาณแบบเคอร์เนลโดยใช้วิธีการแบบ เชิงเส้นให้ประสิทธิภาพไม่ดีเท่ากับการใช้วิธีประมาณค่าแบบ RKHS

ด้วยเหตุที่ข้อมูลเป็นโมเดลไม่เป็นเชิงเส้น การใช้วิธีการแบบพาราเมตริกไม่ได้ให้ผลน่าพอใจ ดังนั้นภายใต้กรอบของวิธี RKHS และการพัฒนาของตัวแบบฟังก์ชันการถดถอยแบบไม่เป็นเชิงเส้น (Nonlinear Functional Regression Model) เป็นการดำเนินการโดยการใช้การประมาณค่าแบบเคอร์เนลฮิลเบอร์ทสเปซ (RKHS Estimate) ที่ได้ดีกว่าโมเดลการถดถอยแบบเชิงเส้น (Linear Regression Model) และประมาณค่าได้ดี และง่ายกว่าวิธีการประมาณค่าแบบเคอร์เนล (Kernel Estimate) ในรูปแบบของการถดถอยแบบนอนพาราเมตริก (Nonparametric Regression) การใช้โมเดล RKHS ในการประมาณค่าเป็นเรื่องของฟังก์ชัน 2 ตัวแปรเป็นค่าจริงที่สมมาตรของ K โดยค่าของ K ต้องไม่ติดลบ คือ $K \geq 0$ การวิเคราะห์ความแปรปรวนร่วมแบบเคอร์เนล (Covariance Kernels) การใช้วิธีการของ RKHS โดยการใช้ผลรวมกำลังสองทำให้ความสามารถในการทำนายมีค่าค่อนข้างสูงและถูกต้องแม่นยำกว่าวิธีอื่น ประยุกต์ไปสู่การวิเคราะห์และแก้ปัญหาในด้านของวิศวกรรม และการปรับประยุกต์โมเดลไปสู่การใช้ฟังก์ชันทางคณิตศาสตร์ ในด้านของฟังก์ชัน

เอกซ์โพเนนเชียล ควอไทล์ และฟังก์ชันลอการิทึม รวมถึงข้อมูลอนุกรมเวลาและยังจัดกระทำได้ทั้งข้อมูลจริง และการจำลองข้อมูล ซึ่งจะนำไปสู่การประยุกต์และพัฒนาการใช้วิธีการวัดที่มีความเหมาะสมถูกต้อง และแม่นยำ

จากการศึกษาวิธีการวิเคราะห์ข้อมูลของข้อมูลที่ค่าความคลาดเคลื่อนมีความสัมพันธ์กัน การใช้วิธีประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดไม่มีประสิทธิภาพและส่งผลให้ตัวประมาณค่าไม่มีคุณสมบัติเป็น BLUE โดยเฉพาะการพยากรณ์ข้อมูลทางเศรษฐศาสตร์ เพราะข้อมูลทางเศรษฐศาสตร์เป็นข้อมูลที่มีการจัดเก็บเป็นช่วงเวลาที่เรียกว่า ข้อมูลอนุกรมเวลา ทำให้ค่าสังเกตและค่าความคลาดเคลื่อนมีความสัมพันธ์กัน ดังนั้นในการพยากรณ์ข้อมูลทางเศรษฐศาสตร์ จึงจำเป็นต้องตระหนักถึงวิธีการประมาณค่าที่เหมาะสมกับข้อมูลเป็นอย่างมาก เพราะถ้าข้อมูลที่ต้องการใช้ไม่เป็นไปตามข้อตกลงเบื้องต้น อาจทำให้เกิดความคลาดเคลื่อนในการพยากรณ์ได้ มีผู้เคยทำและสร้างสมการพยากรณ์ด้วยวิธีต่าง ๆ เช่น วิธี Smoothing Method การวิเคราะห์การถดถอยที่ใช้วิธีกำลังสองน้อยที่สุด (Ordinary Least Square: OLS) ในการแก้ปัญหาการประมาณราคาของผลผลิตข้าว แต่วิธีการเหล่านี้มีข้อตกลงเบื้องต้นข้อหนึ่งคืออนุกรมเวลาที่ $\ldots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, \ldots$ ไม่มีสหสัมพันธ์กัน ซึ่งบ่อยครั้งที่ปรากฏว่าเงื่อนไขไม่เป็นไปตามข้อตกลงเบื้องต้นคือมีหลายกรณีที่อนุกรมเวลา $(\ldots, Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, \ldots)$ มีสหสัมพันธ์ ถ้าเป็นเช่นนี้การใช้วิธีที่มีข้อตกลงเบื้องต้นที่ตัวแปรเป็นอนุกรมเวลาไม่มีสหสัมพันธ์จะไม่เหมาะสม เพราะวิธีการต่าง ๆ นั้น ไม่ได้นำสหสัมพันธ์ที่ปรากฏไปใช้ประโยชน์ในการสร้างตัวแบบพยากรณ์ การมีวิธีการสร้างตัวแบบพยากรณ์ สำหรับอนุกรมเวลาที่นำเอาสหสัมพันธ์ที่ปรากฏไปวิเคราะห์ใช้ประโยชน์ และโดยทั่วไปวิธีเหล่านี้จะให้ผลการพยากรณ์ที่ดีกว่าในอีกหลายวิธี ในปัจจุบันวิธีที่รู้จักและใช้กันมาก คือ วิธีกำลังสองน้อยที่สุดทั่วไป การเลือกปรับข้อมูลในลักษณะที่ไม่เป็นเส้นตรง (Nonlinear) แบบนอนพาราเมตริกซ์ ทำให้เกิดความคลาดเคลื่อนจากการพยากรณ์ค่อนข้างสูง ทำให้การพยากรณ์มีความแม่นยำน้อย ประกอบกับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) เป็นวิธีการแบบนอนพาราเมตริก ที่มีกรอบแนวคิดสำหรับการประมาณค่าฟังก์ชันที่มีความยืดหยุ่น ในการสร้างตัวแบบทางสถิติที่มีข้อมูลทั้งโดยตรงและโดยอ้อม ที่สามารถจัดการกับปัญหาและการกระจายของข้อมูลที่หลากหลาย การประมาณค่าที่มีค่าความแกร่ง (Robust Estimation) สามารถจัดการกับค่าจากการสังเกตที่ซับซ้อนที่เกี่ยวกับอนุพันธ์ ปริพันธ์ ได้แต่วิธีการนี้ยังมีค่าคลาดเคลื่อนจากการประมาณที่ค่อนข้างสูง ประกอบกับข้อมูลที่นำมาวิเคราะห์อาจเป็นอนุกรมเวลาที่ความคลาดเคลื่อนมีอัตตสหสัมพันธ์กัน ทำให้วิธีการประมาณค่าพารามิเตอร์การถดถอยแบบ Ferraty (2007) ไม่เป็นที่นิยมนำมาใช้ในการพยากรณ์ เพราะมีค่าความคลาดเคลื่อนที่สูง

ผู้วิจัยจึงสนใจที่จะพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) เพื่อนำไปใช้ในการประมาณค่าพารามิเตอร์การถดถอย โดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ (Reproducing Kernel Hilbert Space) ให้สามารถควบคุมค่าความคลาดเคลื่อนที่เกิดขึ้นจากการพยากรณ์ให้มีค่าน้อยลงและค่าพยากรณ์มีความถูกต้องมากยิ่งขึ้น ในกรณีที่ข้อมูลเป็นอนุกรมเวลาที่ความคลาดเคลื่อนมีอัตตสหสัมพันธ์ และนำวิธีประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ ไปใช้ในการปรับข้อมูลการพยากรณ์ราคาข้าวเปลือกเจ้าล่วงหน้าของไทย ที่มีตัวแบบ คือ ARIMA (1, 0, 0) ในการพยากรณ์ เพื่อให้ผู้ที่เกี่ยวข้องนำผลการศึกษาที่ได้ไปใช้ในเรื่องของการวางแผนการผลิตการเพิ่มมูลค่าสินค้าทางการเกษตรในด้านของราคาของผลผลิต เพื่อการส่งออกและการวางแผนการผลิตที่มีประสิทธิภาพให้ดีขึ้นทั้งในด้านปริมาณและคุณภาพ อันจะนำไปสู่การพัฒนาการผลิตที่ยั่งยืนต่อไป ซึ่งสอดคล้องกับงานวิจัยของจิตราภรณ์ ฟั่นศิริ (2547) ศึกษาการพยากรณ์ราคาส่งออกข้าวด้วยวิธีบอกซ์-เจนกินส์ เพื่อศึกษารูปแบบและพยากรณ์ราคาส่งออกข้าวของไทย ในการวิเคราะห์ใช้ข้อมูลราคาส่งออกข้าวเป็นรายเดือนในช่วงเดือนมกราคม พ.ศ. 2531 ถึงเดือนธันวาคม พ.ศ. 2546 จำนวน 192 เดือน ข้อมูลจากกรมการค้าต่างประเทศ วิธีการศึกษาจะทดสอบความนิ่งของข้อมูลโดยใช้วิธีการทดสอบ Unit root และกำหนดตัวแบบด้วยวิธีของบอกซ์-เจนกินส์ ผลการทดสอบปรากฏว่าข้อมูลราคาส่งออกข้าวมีลักษณะไม่นิ่ง จึงทำการหาผลต่างอันดับ 1 และจากการพิจารณาค่าสหสัมพันธ์ในตัวเองพบว่า ตัวแบบที่เหมาะสมที่มีค่าสัมประสิทธิ์ของตัวแบบเท่ากับ 0.360 และ 0.228 ตามลำดับ และมีนัยสำคัญทางสถิติที่ระดับ .01 สำหรับการตรวจสอบความถูกต้องพบว่าค่าประมาณความคลาดเคลื่อนมีลักษณะเป็นเชิงสุ่ม ที่ระดับนัยสำคัญทางสถิติ .01 จากค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสองและค่าสัมประสิทธิ์ Thiel ที่มีค่าต่ำสุด จะได้ว่าตัวแบบที่มีความเหมาะสมที่สุดซึ่งมีสมการพยากรณ์คือ $y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$ เป็นสมการที่ดีที่สุดเมื่อเทียบกับตัวแบบอื่น ๆ ดังนั้นในการศึกษานี้จึงนำตัวแบบไปพยากรณ์ราคาข้าวเปลือกเจ้าของไทย ในช่วงปี พ.ศ. 2557

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์
2. เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ (2x6x9) แบนวิดจ์ 2 วิธี คือ วิธีที่ 1 Rules of Thumb และวิธีที่ 2 Silverman’s Rules of Thumb ขนาดตัวอย่าง 6 ขนาด (5, 10, 15, 30, 50 และ100) และค่าอัตตสหสัมพันธ์ 9 ระดับ (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) โดยวิธีการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล
3. เพื่อพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าโดยใช้ตัวแบบ ARIMA

กรอบแนวคิดในการวิจัย

การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยแบบปรับใหม่ โดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนมีอัตตสหสัมพันธ์ เริ่มต้นด้วยการศึกษาวิธีการประมาณค่าพารามิเตอร์การถดถอยแบบเชิงเส้น ในกรณีข้อมูลที่นำมาวิเคราะห์ไม่เป็นไปตามข้อตกลงเบื้องต้นบางข้อ ส่วนมากจะเกิดเฉพาะกรณีที่ความคลาดเคลื่อนสุ่มมีสหสัมพันธ์กัน กล่าวคือ $E(\varepsilon_t, \varepsilon_s) \neq 0$ สำหรับ $t \neq s$ ซึ่งการเกิดเหตุการณ์เช่นนี้เรียกว่า อัตตสหสัมพันธ์ของความคลาดเคลื่อน (Serial Correlation หรือ Auto Correlation) การเกิดอัตตสหสัมพันธ์จะเกิดกับกรณีค่าของตัวแปรตามและตัวแปรต้นมีการเก็บข้อมูลตามเวลาเรียกว่า อนุกรมเวลา (Box et al., 1994, pp. 739-750) ข้อมูลอนุกรมเวลา (Time Series Data) อัตตสหสัมพันธ์ของความคลาดเคลื่อนสุ่มที่พบโดยทั่วไป เป็นอัตตสหสัมพันธ์ที่หนึ่งช่วงเวลา (Autocorrelation ที่ lag 1) ผลจากการเกิดปัญหาอัตตสหสัมพันธ์ในความคลาดเคลื่อนสุ่ม จะทำให้ตัวประมาณค่าของพารามิเตอร์ β_1 เป็นตัวประมาณค่าที่ไม่เอนเอียงเชิงเส้นที่ดีที่สุด เนื่องจากมีค่าความแปรปรวนไม่ต่ำที่สุด แม้จะยังคงเป็นตัวประมาณค่าที่ไม่เอนเอียงก็ตาม (Raoc & Grilliches, 1969, pp. 14-23) มีผลให้การอนุมานเกิดความผิดพลาดร้ายแรง เช่น ถ้าเกิดปัญหาอัตตสหสัมพันธ์ขึ้น โดยเฉพาะอย่างยิ่งอัตตสหสัมพันธ์ทางบวก (Positive Autocorrelation) (Bower man & O’Connell, 1993, pp. 541-549; Box et al., 1994, pp. 739-750)

การแก้ปัญหาอัตตสหสัมพันธ์ ในกรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์อันดับที่ 1

แก้ได้หลายวิธี เช่น การเพิ่มตัวแปรพยากรณ์ในตัวแบบ การหาตัวแบบการถดถอยใหม่ การสร้างตัวแบบการถดถอยออโต้ การสร้างตัวแบบการถดถอยของเปอร์เซ็นต์ที่เปลี่ยนแปลง (ทรงศิริ แต้สมบัติ, 2548, หน้า 182) และในกรณีที่ความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์ เท่าที่ศึกษาปรากฏว่า ยังไม่มีงานวิจัยใดที่นำวิธีการประมาณค่าพารามิเตอร์การถดถอยด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซมาแก้ปัญหาข้อมูล ในกรณีที่ความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์กัน

การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยแบบปรับใหม่ กรณีความคลาดเคลื่อนมีอัตตสหสัมพันธ์เป็น ดังนี้

วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีฟังก์ชันเคอร์เนลแบบเกาส์เซียน คือ

1. ฟังก์ชันเคอร์เนล คือ $\hat{f}_h(\mathrm{x}) = \frac{1}{nh}\sum_{i=1}^{n} K(\mathrm{u})$

ฟังก์ชันในรูปแบบฟังก์ชัน (Gaussian Kernel Function) คือ

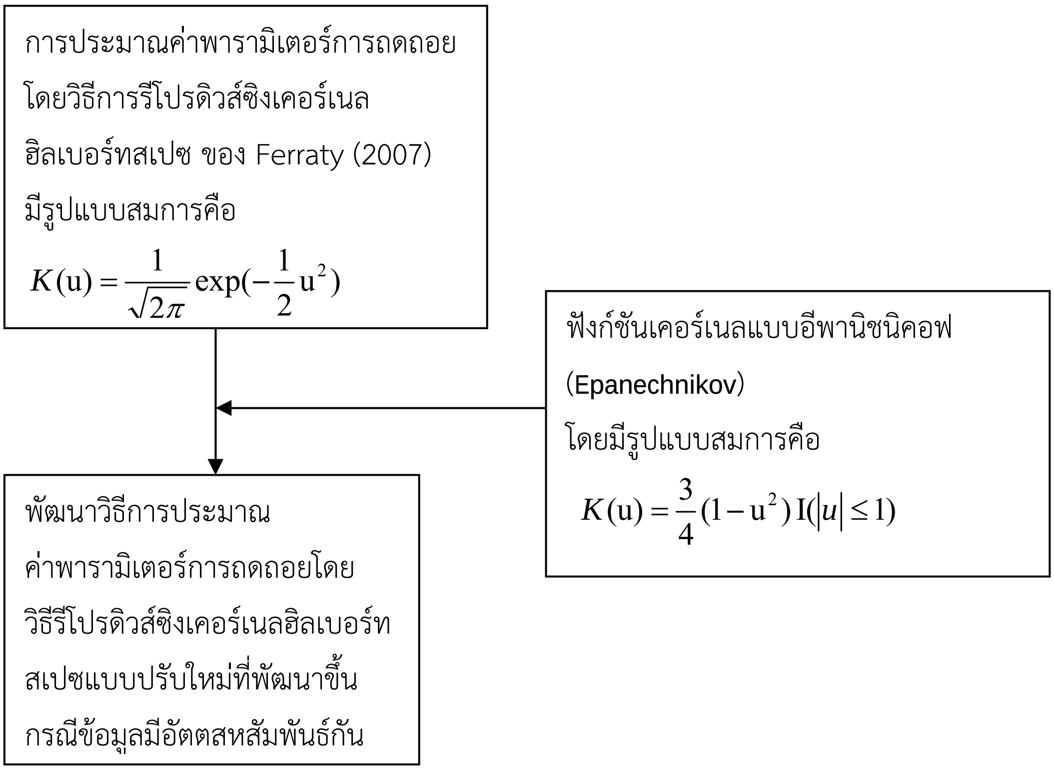
$$K(\mathrm{u}) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\mathrm{u}^2) \quad \text{เมื่อ} \quad u = \frac{x - X_i}{h}$$

2. พัฒนาฟังก์ชันในรูปแบบฟังก์ชันแบบเกาส์เซียน ให้เป็นฟังก์ชันอยู่ในรูปแบบของอีพานิชนิคอฟ (Epanechnikov Kernel Function) ที่มีรูปแบบของฟังก์ชัน คือ

$$K(\mathrm{u}) = \frac{3}{4}(1-\mathrm{u}^2)\,\mathrm{I}(|u| \le 1) \quad \text{เมื่อ} \quad u = \frac{x - X_i}{h}$$

และใช้สมการหาค่าความคลาดเคลื่อนค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error: *MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (Standard Error: *Std. error*) เพื่อเปรียบเทียบประสิทธิภาพวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ที่นำเสนอ

จากที่กล่าวมาข้างต้น กรอบแนวคิดในการวิจัยสำหรับการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ดังภาพที่ 1-1



ภาพที่ 1-1 กรอบแนวคิดในการวิจัย

สมมติฐานของการวิจัย

1. วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองต่ำกว่าวิธีการประมาณค่าพารามิเตอร์วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์

2. วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ มีค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีการประมาณค่าพารามิเตอร์การถดถอยที่ได้จากวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์

3. การพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าโดยใช้ตัวแบบ ARIMA มีค่าสัมประสิทธิ์การทำนาย (R^2) สูงกว่า 50%

ขอบเขตของการวิจัย

วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์กัน ได้ศึกษาภายใต้เงื่อนไข ดังนี้

สำหรับข้อมูลจำลอง (วัตถุประสงค์ข้อ 2)

1. ตัวแปรต้น คือ วิธีการประมาณค่าพารามิเตอร์การถดถอย จำนวน 2 วิธี ได้แก่
   1.1 วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่
   1.2 วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)
2. ตัวแปรตาม คือ ประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์การถดถอยของฟังก์ชันที่พิจารณาได้จาก
   2.1 ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง
   2.2 ค่าความคลาดเคลื่อนมาตรฐาน
3. รูปแบบของความสัมพันธ์

$$y = F(\mathrm{x}) + \varepsilon$$

4. กำหนดให้ตัวอย่างสุ่มเป็นตัวอย่างสุ่มที่เป็นอิสระต่อกัน และมีฟังก์ชันการแจกแจงความน่าจะเป็นเดียวกัน (independent and identically distributed: iid)
5. กำหนดฟังก์ชันเคอร์เนล คือ $\hat{f}_h(\mathrm{x}) = \dfrac{1}{nh}\sum_{i=1}^{n} K(\mathrm{u})$
   5.1 ฟังก์ชันในรูปแบบฟังก์ชันแบบเกาส์เซียน (Gaussian Kernel Function) โดยมีรูปแบบฟังก์ชันคือ $K(\mathrm{u}) = \dfrac{1}{\sqrt{2\pi}}\exp(-\dfrac{1}{2}\mathrm{u}^2)$ เมื่อ $u = \dfrac{x - X_i}{h}$
   5.2 ฟังก์ชันในรูปแบบฟังก์ชันอีพานิชนิคอฟ (Epanechnikov Kernel Function) โดยมีรูปแบบของฟังก์ชัน คือ $K(\mathrm{u}) = \dfrac{3}{4}(1-\mathrm{u}^2)\,\mathrm{I}(|u| \le 1)$ เมื่อ $u = \dfrac{x - X_i}{h}$
6. กำหนดแบนวิดจ์ที่ใช้ ดังนี้
   6.1 วิธี Rules of thumb (Deheuvels, 1977: Silverman, 1986)

$$h_{ROT} = 1.06\hat{\sigma}(\mathrm{n}^{\frac{1}{5}})$$

โดยที่ h_{ROT} คือ แบนวิดจ์ที่ได้จากวิธี Rules of thumb

n คือ ขนาดตัวอย่าง

$\hat{\sigma}$ คือ ค่าประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม

$X_i : i = 1,2,...,n$

โดยที่ $\hat{\sigma} = \sqrt{MSE}$

$$= \sqrt{\frac{\sum (\mathrm{x}_i - \bar{x})^2}{n}}$$

6.2 วิธี Silverman's Rules of Thumb (Silverman, 1986)

$$h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$$

โดยที่ h_{SROT} คือ แบนวิดจ์ที่ได้จากวิธี Silverman's Rules of Thumb

n คือ ขนาดตัวอย่าง

A คือ ค่าที่น้อยที่สุดระหว่างค่าประมาณส่วนเบี่ยงเบนมาตรฐานของ

ตัวแปรสุ่ม $X_i : i = 1,2,...,n$ กับ $\frac{IQR}{1.34}$ คือ $\min\left[\hat{\sigma}.\frac{IQR}{1.34}\right]$

IQR คือ พิสัยระหว่างควอไทล์ (Interquartile range)

$\hat{\sigma}$ คือ ค่าประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม $X_i : i = 1,2,...,n$

7. ตัวแปรสุ่มที่นำมาศึกษา

7.1 การแจกแจงแบบมาตรฐาน

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad \text{เมื่อ } -\alpha < \theta < \alpha$$

7.2 การแจกแจงแบบโคชีมาตรฐาน (Cauchy Distribution Standard)

โดยมีรูปแบบของฟังก์ชัน คือ

$$f(x) = \frac{1}{\pi\beta\left[1 + \left(\frac{x-\theta}{\beta}\right)^2\right]} \quad , -\alpha < \theta < \alpha$$

8. ฟังก์ชันของตัวแปรสุ่มที่ศึกษา คือ $g(\mathrm{X}_1, \mathrm{X}_2, ..., \mathrm{X}_m) = \sum_{i=1}^{n} X_i$ ที่ $m = 2$

9. กำหนดให้ฟังก์ชันเคอร์เนลเป็นฟังก์ชันอันดับที่ 2 กล่าวคือ $k = 2$

10. ศึกษาในกรณีที่ข้อมูลความคลาดเคลื่อนมีอัตตสหสัมพันธ์

11. ขนาดตัวอย่างที่ใช้ในการศึกษามี 6 ขนาด คือ 5, 10, 15, 30, 50 และ 100

12. ระดับของอัตตสหสัมพันธ์มี 9 ระดับ คือ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 แบ่งเป็น 3 ระดับ คือ ระดับต่ำ เท่ากับ 0.1 ถึง 0.3 ระดับกลาง เท่ากับ 0.4 ถึง 0.6

และระดับสูง เท่ากับ 0.7 ถึง 0.9

13. การวิจัยนี้ศึกษาอัตตสหสัมพันธ์เฉพาะอัตตสหสัมพันธ์ของค่าความคลาดเคลื่อนทางบวก (Positive Autocorrelation) โดยการจำลองข้อมูลให้ค่าความคลาดเคลื่อนมีอัตตสหสัมพันธ์อันดับที่ 1 (AR (1))

14. ค่าความคลาดเคลื่อน (ε_i) มีอัตตสหสัมพันธ์กัน โดยกำหนดรูปแบบความสัมพันธ์ AR (1) เป็น $\varepsilon_i = \rho\varepsilon_{i-1} + a_i$; $i = 1, 2, ..., n$

15. กำหนดจำนวนการกระทำซ้ำในแต่ละสถานการณ์เป็น 1,000 ครั้ง

16. การวิจัยนี้ใช้วิธีการจำลอง (Simulation) ให้มีสถานการณ์ตามที่กำหนดด้วยโปรแกรมอาร์

17. เปรียบเทียบวิธีการประมาณค่าพารามิเตอร์ว่า วิธีการใดเหมาะสมมากที่สุด พิจารณาจากเกณฑ์การเปรียบเทียบ คือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error: *MSE*) และค่าความคลาดเคลื่อนมาตรฐาน (Standard Error: *Std. error*) วิธีใดให้ค่า *MSE* และ *Std. error* น้อยกว่า จะเป็นวิธีการประมาณค่าพารามิเตอร์การถดถอย ในกรณีข้อมูลมีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์ที่มีประสิทธิภาพมากกว่า

สำหรับข้อมูลจริง (วัตถุประสงค์ข้อ 3)

1. ข้อมูลจริงเป็นข้อมูลทุติยภูมิราคาผลผลิตทางการเกษตร ได้แก่ ราคาข้าวเปลือกเจ้า 15% ในประเทศไทยของสำนักงานเศรษฐกิจทางการเกษตร กระทรวงเกษตรและสหกรณ์ ปี พ.ศ. 2552–2556

1.1 กลุ่มตัวอย่าง ราคาข้าวเปลือกเจ้าของประเทศไทยปีที่ $i-1$

1.2 วิธีการที่ศึกษา คือ วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่

2. ตัวแปรที่ศึกษา ได้แก่ สัมประสิทธิ์การกำหนด (R^2)

3. กำหนดตัวแบบของความสัมพันธ์ของตัวแปร ประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่

4. ประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่

5. พยากรณ์ราคาผลผลิตข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าด้วยตัวแบบ ARIMA

6. นำผลการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยที่ได้จากด้วยวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ไปเทียบกับราคาข้าวเปลือกเจ้าของไทยปี พ.ศ. 2557

ประโยชน์ที่คาดว่าจะได้รับจากการวิจัย

1. ได้วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนล ฮิลเบิร์ทสเปซ แบบปรับใหม่ที่พัฒนาขึ้น ในกรณีความคลาดเคลื่อนมีอัตตสหสัมพันธ์
2. ได้วิธีการประมาณค่าพารามิเตอร์การถดถอยที่เหมาะสมในการวิเคราะห์ข้อมูลการถดถอย ในกรณีข้อมูลเกิดปัญหาความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์
3. ผู้ใช้นำวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนล ฮิลเบิร์ทสเปซ แบบปรับใหม่ ไปแก้ปัญหาข้อมูลที่ความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์ได้
4. เป็นแนวทางในการนำไปประยุกต์กับข้อมูลจริงที่เกิดปัญหาความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์อันดับหนึ่งในการวิเคราะห์การถดถอย

นิยามศัพท์เฉพาะ

อัตตสหสัมพันธ์ (Autocorrelation) หมายถึง เหตุการณ์ที่ค่าความคลาดเคลื่อนมีความสัมพันธ์กัน

ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error: *MSE*) หมายถึง ค่าเฉลี่ยของความแตกต่างของระหว่างค่าจริงกับค่าประมาณกำลังสอง

ค่าความคลาดเคลื่อนมาตรฐาน (Standard Error; *Std. error*) หมายถึง ผลต่างระหว่างค่าที่วัดได้กับค่าที่แท้จริง

เทคนิคมอนติคาร์โล (Monte Carlo Technique) หมายถึง วิธีการทางคณิตศาสตร์ที่ใช้เทคโนโลยีทางคอมพิวเตอร์มาช่วยในการจำลองสถานการณ์ (Simulation) โดยการอาศัยตัวเลขสุ่ม (Random Number) ด้วยการสร้างตัวแปรให้สอดคล้องกับสถานการณ์ที่ต้องการและมีการทดลองซ้ำหลาย ๆ ครั้ง เพื่อให้ได้ค่าที่เกิดจากการจำลองสถานการณ์ลู่เข้าสู่ค่าจริง โดยอาศัยกฎจำนวนมาก (Law of Large Number) ในการนำไปใช้เป็นข้อสรุปของสถานการณ์ โดยในงานวิจัยนี้ใช้โปรแกรมอาร์ (Programm R) ในการจำลอง

ค่าประมาณความหนาแน่นแบบเคอร์เนล (Kenel Density Estimator) หมายถึง ค่าความหนาแน่นที่ประมาณขึ้น โดยวิธีการประมาณความหนาแน่นแบบเคอร์เนล

ตัวประมาณเคอร์เนล (Kernel Estimator) หมายถึง วิธีทางนอนพาราเมตริกในการประมาณฟังก์ชันความหนาแน่นของการแจกแจงความน่าจะเป็น โดยคำนวณจากตัวอย่างขนาด n โดยแทนที่แต่ละค่าของข้อมูลด้วย “เคอร์เนล” ของพื้นที่ $\frac{1}{n}$ ผลลัพธ์จะแสดงโค้งของฟังก์ชันความหนาแน่นที่ประมาณได้

วิธีการประมาณค่าพารามิเตอร์โดยวิธีรีโปรดิวส์ซิงเคอร์เนล หมายถึง วิธีทางนอนพาราเมตริกในการประมาณค่าฟังก์ชันความหนาแน่นของการแจกแจงความน่าจะเป็น โดยคำนวณซ้ำจากตัวประมาณค่าเคอร์เนล ตัวอย่างขนาด n โดยแทนที่แต่ละค่าของข้อมูลด้วย “ค่ารีโปรดิวส์ซิงเคอร์เนล”

วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ หมายถึง วิธีทางนอนพาราเมตริกในการประมาณค่าฟังก์ชันความหนาแน่นของการแจกแจงความน่าจะเป็น โดยคำนวณซ้ำจากตัวประมาณค่าเคอร์เนลที่ปรับใหม่ ตัวอย่างขนาด n โดยแทนที่แต่ละค่าของข้อมูลด้วย “ค่ารีโปรดิวส์ซิงเคอร์เนล” จากวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบิร์ทสเปซ แบบปรับใหม่

วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) หมายถึง วิธีทางนอนพาราเมตริกในการประมาณค่าฟังก์ชันความหนาแน่นของการแจกแจงความน่าจะเป็น โดยคำนวณซ้ำจากตัวประมาณค่าเคอร์เนล ตัวอย่างขนาด n โดยแทนที่แต่ละค่าของข้อมูลด้วย “ค่ารีโปรดิวส์ซิงเคอร์เนล” ตามวิธีการของ Ferraty

ตัวแบบ ARIMA (ARIMA Model) หมายถึง ตัวแบบที่แสดงความสัมพันธ์ของอนุกรมเวลาราคาข้าวเปลือกเจ้า 15% ของไทยกับเวลา

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ กรณีค่าความคลาดเคลื่อนมีอัตตสหสัมพันธ์ เปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ กับวิธีประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ (2x6x9) ได้แก่ แบนวิดจ์ 2 วิธี คือ วิธีที่ 1 Rules of Thumb และวิธีที่ 2 Silverman’s Rules of Thumb ขนาดตัวอย่าง 6 ขนาด (5, 10, 15, 30, 50 และ 100) และค่าอัตตสหสัมพันธ์ 9 ระดับ (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) โดยวิธีการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) และเพื่อพยากรณ์ราคาข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA ซึ่งผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง โดยนำเสนอผลการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้อง ดังนี้

ตอนที่ 1 การประมาณค่าของฟังก์ชันเคอร์เนล

ตอนที่ 2 การประมาณค่าของฟังก์ชัน (Reproducing Kernel Hilbert Space)

ตอนที่ 3 การพยากรณ์ด้วยอนุกรมเวลาและการสร้างตัวแบบ ARIMA

ตอนที่ 4 การตรวจสอบความเหมาะสมของอนุกรมเวลา

ตอนที่ 5 อัตตสหสัมพันธ์ในข้อมูลอนุกรมเวลา

ตอนที่ 6 การจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล

ตอนที่ 7 งานวิจัยที่เกี่ยวข้อง Reproducing Kernel Hilbert Space

ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับข้าวเปลือกเจ้า

ตอนที่ 1 การประมาณค่าของฟังก์ชันเคอร์เนล

การประมาณค่าความหนาแน่นแบบเคอร์เนล เป็นวิธีการนอนพาราเมตริกเพื่อการประมาณค่าฟังก์ชันความหนาแน่น ความน่าจะเป็นของตัวแปรสุ่ม การประมาณค่าความหนาแน่นแบบเคอร์เนล (Kernel) เป็นปัญหาพื้นฐานของการปรับข้อมูลให้เรียบ เมื่อมีการอ้างอิงสรุปถึง ประชากร ตามข้อมูลที่แน่นอนจากกลุ่มตัวอย่าง

1. ฮิสโตแกรม

ฮิสโตแกรม (Histogram) เป็นวิธีการที่มีประโยชน์แต่มีข้อจำกัด ในการประมาณค่าหรือช่วยให้เห็นภาพความหนาแน่นที่แฝงอยู่จริงในข้อมูลที่ได้จากการสังเกตที่ไม่ทราบการแจกแจง ฮิสโตแกรมเป็นฟังก์ชันที่จำเป็นแต่มีขั้นตอนไม่ต่อเนื่อง ดังนั้น หากเราเชื่อว่าข้อมูลที่สังเกตได้เกิดจากความหนาแน่นที่ต่อเนื่อง หรือความหนาแน่นที่จำแนกความแตกต่างได้ก็ตาม วิธีการประมาณค่าที่คล้ายฮิสโตแกรมวิธีหนึ่ง น่าจะเป็นวิธีที่น่าสนใจกว่าฮิสโตแกรมของฟังก์ชันเคอร์เนล คือ ลักษณะทั่วไปของฮิสโตแกรมปกติมีความเกี่ยวข้องกับฟังก์ชันแต่ละจุดข้อมูลที่เรียกว่าฟังก์ชันเคอร์เนลฮิสโตแกรมของฟังก์ชันเคอร์เนล คือ ผลรวมที่ได้รับการปรับให้เป็นปกติอย่างเหมาะสมแล้วของฟังก์ชัน โดยปกติฟังก์ชันเคอร์เนลจะขึ้นอยู่กับพารามิเตอร์อย่างหนึ่งที่ปกติจะเรียกกันว่าแบนวิดจ์ (Bandwith) ความกว้างของข้อมูล ที่มีผลอย่างชัดเจน แสดงอย่างมีนัยสำคัญต่อความขรุขระหรือความราบเรียบของฮิสโตแกรมของฟังก์ชันเคอร์เนลที่ถูกสร้างขึ้นซึ่งบางทีมีการสับสนว่าฟังก์ชันเคอร์เนลก็คือฟังก์ชันความหนาแน่นนั่นเอง

การเลือกการแจกแจงเป้าหมายที่สร้างข้อมูลแบบสุ่มจากการแจกแจงนั้น และเลือกประเภทของฟังก์ชันเคอร์เนล ตัวอย่างเช่น ฟังก์ชันเคอร์เนล แบบ Epanechnikov มีคุณสมบัติแบบ Asymptotic ที่ชัดเจนประการหนึ่งที่ทำให้เป็นฟังก์ชันเคอร์เนลที่พึงประสงค์ แม้ว่าจะสามารถได้รับฮิสโตแกรมของฟังก์ชันเคอร์เนลที่คล้ายคลึงอย่างมากกับฮิสโตแกรมปกติ โดยการเลือกฟังก์ชันเคอร์เนลรูปแบบเดียวกันแล้วเลือกผลที่เป็นจริงแบบสุ่ม (Random Realization) จากการแจกแจงที่เป็นเป้าหมายและดูว่าฮิสโตแกรมของฟังก์ชันเคอร์เนล สอดคล้องกับความหนาแน่นจริงที่แฝงอยู่

ความสนใจในเทคนิคการประมาณค่าของฟังก์ชันเคอร์เนล มีการใช้เทคนิคที่คล้าย ๆ กันไปประมาณค่าแบบนอนพาราเมตริกฟังก์ชัน σ ในสมการดิฟเฟอเรนเชียลสต็อกแคสติก

$$dX_t = \mu(X_t)d_t + \sigma(X_t)dW_t$$

เมื่อ W_t คือ การเคลื่อนที่แบบบราวเนียนมาตรฐาน แต่นอกจากนี้ เทคนิคการประมาณค่าของฟังก์ชันเคอร์เนล ยังถูกนำไปใช้ ตัวอย่างเช่น ในการประมาณค่าฟังก์ชันในสมการการถดถอยที่ไม่เป็นเชิงเส้น

$$Y_t = m(X_t) + \sigma(X_t)$$

เมื่อ t เป็นลำดับที่แจกแจงเหมือนกัน และอิสระในลักษณะที่ $t \sim N$ อีกด้วย

การทำแบนวิดจ์ให้เล็กจะให้ผลเกี่ยวกับข้อมูลเชิงสุ่มที่ถูกสร้างขึ้นตามกฎของการแจกแจงเป้าหมายที่เลือกมา แต่การทำแบนวิดจ์ให้เล็กก็ยังทำให้เกิดผลเป็นฮิสโตแกรมของฟังก์ชันเคอร์เนลได้อย่างดี การทำแบนวิดจ์ให้ใหญ่มาก ๆ จะทำให้รอยขรุขระบนฮิสโตแกรมของฟังก์ชันเคอร์เนลราบเรียบลง แต่อาจมีผลให้ฮิสโตแกรมของฟังก์ชันเคอร์เนลไม่สามารถคงลักษณะที่น่าสนใจที่ไม่ธรรมดาของข้อมูลไว้ได้ ในขณะที่ฮิสโตแกรมของฟังก์ชันเคอร์เนลกำลังปรับตัวเพื่อที่จะให้คู่ขนานและสอดคล้องความหนาแน่นแฝงที่มีจริงนั้นอัตราของการปรับตัวเพื่อที่จะให้คู่ขนานและสอดคล้องก็ดูเหมือนจะไม่เร็วนัก แต่ก็เห็นชัดเจนแล้วว่าหากการแจกแจงจริงที่แฝงอยู่ของข้อมูลราบเรียบพอแล้ว อัตราของการปรับตัวเพื่อที่จะให้คู่ขนานและสอดคล้องในความหมายแบบ L^2 ก็จะเป็น $O(n^{-4/5})$ หรืออาจจะกล่าวได้ว่าฮิสโตแกรมของฟังก์ชันเคอร์เนลกำลังปรับตัวเพื่อที่จะให้คู่ขนานและสอดคล้องในอัตราที่เร็วกว่าอัตราที่คล้ายคลึงกัน ในทฤษฎีขีดจำกัดกลาง โปรดดูเพิ่มเติมที่ทฤษฎีเสริมของ Kolmogorov ต่อทฤษฎีของ Gilvenko-Cantelli ฮิสโตแกรมของฟังก์ชันเคอร์เนลที่สร้างขึ้นครั้งนี้ ยังมิได้รับการปรับให้เข้ากับข้อตกลงเบื้องต้นที่แฝงอยู่ใด ๆ ที่เกี่ยวข้องกับการสนับสนุนการแจกแจงที่เป็นเป้าหมาย การใช้ฮิสโตแกรมของฟังก์ชันเคอร์เนลกับการประมาณค่าความหนาแน่นที่ยกกำลังดูตัวแปรสุ่มยกกำลังมีค่าเป็นลบและมีค่าความน่าจะเป็นเท่ากับศูนย์ แต่โดยเสมือนจริงนั้นฮิสโตแกรมของฟังก์ชันเคอร์เนลทั้งหมดที่ใช้เพื่อประมาณค่าความหนาแน่นที่ยกกำลังจะต้องเป็นค่าบวกไปทางซ้ายของค่าศูนย์อย่างเคร่งครัด อย่างไรก็ตาม ยังมีเทคนิคอีกมากมายที่จะใช้จัดการคุณลักษณะที่ไม่พึงประสงค์เช่นนี้ นอกจากนี้ก็ยังมีอีกหลายเทคนิคที่จะใช้เลือกแบนวิดจ์ “ให้ดีที่สุด” แม้จะไม่ทราบการแจกแจงที่แฝงอยู่ก็ตาม

2. การประมาณค่าความหนาแน่น (Density Estimation)

ฟังก์ชันการแจกแจงความหนาแน่นของความน่าจะเป็น (Probability Density Function) เป็นฟังก์ชันที่ใช้อธิบายถึงลักษณะของตัวแปรสุ่ม เมื่อพิจารณาตัวอย่างสุ่ม X ซึ่งมีฟังก์ชันความหนาแน่นและความน่าจะเป็น

$$P(\mathrm{a} < X < \mathrm{b}) = \int_a^b f(\mathrm{x})\,\mathrm{d}(\mathrm{x}) \quad \text{เมื่อ } a < b$$

ถ้า $X_1, X_2, ..., X_n$ เป็นตัวอย่างสุ่มขนาด n จากประชากร โดยไม่ทราบฟังก์ชันความหนาแน่นน่าจะเป็น ต้องทำการประมาณฟังก์ชันความหนาแน่น โดยวิธีการประมาณนี้ เรียกว่า การประมาณความหนาแน่น (Density Estimation) เป็นวิธีที่สร้างฟังก์ชันความหนาแน่นจากตัวอย่างสุ่มสามารถ

แบ่งเป็น 2 วิธี คือ วิธีการทางด้านพาราเมตริก และวิธีการทางด้านนอนพาราเมตริก (Silverman, 1986, p. 6)

การประมาณความหนาแน่นแบบพาราเมตริก จะต้องคำนึงถึงค่าพารามิเตอร์ของการแจกแจงที่คาดว่าจะเป็น เช่น การแจกแจงแบบปกติ ซึ่งมีพารามิเตอร์ค่าเฉลี่ย μ และพารามิเตอร์ความแปรปรวน σ^2 ที่ไม่ทราบค่า ในการประมาณความหนาแน่นจึงต้องเริ่มจาก ประมาณค่าเฉลี่ย μ และความแปรปรวน σ^2 จากชุดข้อมูล จากนั้นนำไปแทนที่พารามิเตอร์ในฟังก์ชันแจกแจงแบบปกติการประมาณความหนาแน่นแบบนอนพาราเมตริก ไม่ต้องคำนึงถึงค่าพารามิเตอร์ใด ๆ กล่าวคือ การประมาณความหนาแน่นแบบนอนพาราเมตริกเป็นการประมาณฟังก์ชันความหนาแน่นจากชุดข้อมูลโดยตรง

วิธีการประมาณความหนาแน่นแบบนอนพาราเมตริกที่ใช้กันในปัจจุบันมีหลายวิธี ซึ่งส่วนใหญ่มีพื้นฐานจากการประมาณความหนาแน่นแบบฮิสโตแกรม การประมาณค่าความหนาแน่นแบบฮิสโตแกรม (Histogram) เป็นวิธีการประมาณความหนาแน่นแบบนอนพาราเมตริกที่เก่าแก่ที่สุด (Wand & Jones, 1995, pp. 97-116) ซึ่งถ้ากำหนดให้ $X_1, X_2, ..., X_n$ เป็นตัวแปรสุ่มที่เป็นอิสระต่อกันและมีฟังก์ชันการแจกแจกความน่าจะเป็นเดียวกัน (Independent and Identically Distributed; iid)

กำหนดให้ X_0 เป็นจุดเริ่มต้น และ h เป็นความกว้างของแท่ง (Binwidth) โดยที่ h มีค่ามากกว่า 0 และกำหนดให้ความกว้างของแท่งมีค่าแบบช่วงระหว่าง $(\mathrm{X}_0 + \mathrm{mh}, \mathrm{X}_0(\mathrm{m}+1)\mathrm{h})$ โดยที่ m เป็นจำนวนเต็มบวกหรือจำนวนเต็มลบ

การคำนวณเริ่มต้นจากกำหนดให้ $a_0 < X_{(1)}$ และ $a_n < X_{(\mathrm{n})}$ โดยที่ $a_0 < a_1 < ... < a_n$ โดย $X_{(1)} = \min(\mathrm{X}_1, ..., \mathrm{X}_n)$ และ $X_{(1)} = \max(\mathrm{X}_1, ..., \mathrm{X}_n)$ จากนั้นแบ่งช่วงของ $(a_0, a_n]$ ออกเป็น k ช่วงโดยแต่ละช่วงมีความกว้างเท่า ๆ กัน โดยแต่ละช่วงมีความกว้างเท่ากับ $h = \frac{a_n - a_0}{k}$ แต่ก็สามารถกำหนดให้ค่าความกว้างแต่ละช่วงมีค่าไม่เท่ากันก็ได้ แต่โดยส่วนมากจะนิยมใช้ในกรณีที่กำหนดให้มีค่าความกว้างของช่วงเท่ากัน

กล่าวได้ว่าอัตราส่วนของจำนวนของตัวแปรสุ่มที่มีค่าอยู่ในแต่ละช่วงต่อจำนวนตัวแปรสุ่มทั้งหมด นั่นคือ ความถี่ของช่วงที่ j จะมีค่าเท่ากับ $\frac{n_j}{n}$ โดยที่ n_j คือ จำนวนของตัวแปรสุ่มที่อยู่ในช่วงที่ j ดังนั้นตัวประมาณความหนาแน่นแบบฮิสโตแกรม คือ

$$\hat{f}(\mathrm{x}) = \sum_{j=1}^{k} \frac{n_j}{nh_j} I_{(\mathrm{a}_0, \mathrm{a}_n)}(\mathrm{x})$$

$$= \frac{1}{n}\sum_{j=1}^{k}\frac{n_j}{h_j} I_{(a_0,a_n)}(x)\sum_{i=1}^{n} I_{(a_0,a_n)}(x_i)$$

โดยที่ h แทน ความกว้างของแท่ง

n แทน จำนวนข้อมูลทั้งหมด

$I_{(a_0,a_n)}$ แทน จำนวนข้อมูลที่อยู่ในช่วงข้อมูล

3. การประมาณค่าความหนาแน่นแบบเคอร์เนล (Kernel Density Estimation)

การประมาณค่าความหนาแน่นแบบเคอร์เนล คือ วิธีการนอนพาราเมตริกเพื่อการประมาณค่าฟังก์ชันความหนาแน่น ความน่าจะเป็นของ ตัวแปรสุ่ม การประมาณค่าความหนาแน่นแบบ Kernel เป็นพื้นฐานของการปรับข้อมูลให้เรียบ เมื่อมีการอ้างอิงสรุปถึง ประชากรตามข้อมูลที่แน่นอนจากกลุ่มตัวอย่าง และวิธีการนี้ยังเป็นที่รู้จักกันทั่วไปในชื่อว่า วิธีวินโดว์ของ Parzen-Rosenblatt ซึ่งตั้งชื่อตาม Parzen Emanuel และ Rosenblatt Murray ที่เป็นผู้สร้างวิธีการรูปแบบดังกล่าว และเป็นที่รู้จักกันทั่วไปในหมู่ผู้เชี่ยวชาญในวงการวิชาการบางวิชา เช่น วิชาการประมวลผลสัญญาณ (Signal Processing) วิชาเศรษฐมิติ (Econometrics) ส่วนในแวดวงวิชาการสถิติและทางคณิตศาสตร์ การประมาณค่าความหนาแน่นแบบเคอร์เนล ค่อนข้างจะเป็นที่แพร่หลายมากกว่าการกำหนดให้ $X_1, X_2, ..., X_n$ เป็นตัวอย่างที่อิสระและมีการกระจายคล้าย ๆ กันที่มาจากฟังก์ชันความหนาแน่น f ดังนั้น ค่าประมาณการความหนาแน่นแบบเคอร์เนล f ก็คือ

$$\hat{f}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)$$

เมื่อ K คือ ฟังก์ชันเคอร์เนลซึ่งโดยปกติจะมีลักษณะสมมาตรและไม่เป็นค่าลบ

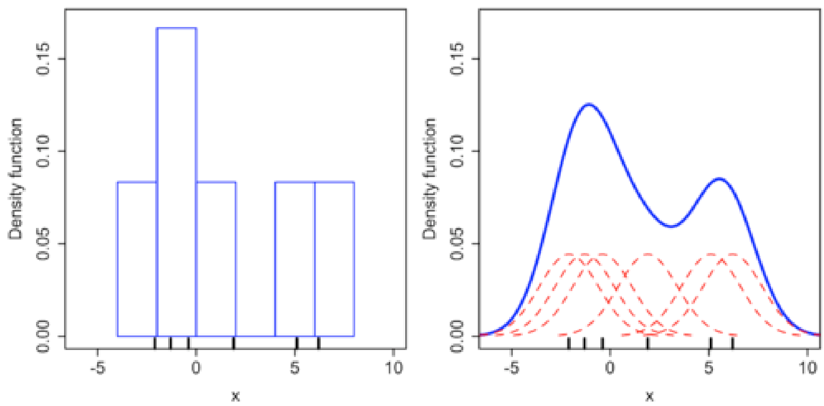
และ h คือ ตัวพารามิเตอร์เพื่อปรับข้อมูลให้เรียบที่มีค่าบวก ที่เรียกกันว่า แบนวิดจ์

ฟังก์ชันเคอร์เนลปรับขนาดก็คือ $K_h(x) = \frac{1}{h}K(\frac{x}{h})$ ฟังก์ชันเคอร์เนลที่ใช้กันทั่วไปมีหลายแบบ เช่น แบบกล่อง แบบสามเหลี่ยม แบบ Triweight แบบ Biweight แบบ Epanechnikov และแบบปกติ ฟังก์ชันเคอร์เนล แบบ Epanechnikov เป็นแบบที่ดีที่สุด

ในแง่ที่มีความแปรปรวนต่ำสุด แม้การสูญเสียประสิทธิภาพจะมีน้อยสำหรับฟังก์ชันเคอร์เนลแบบต่าง ๆ ดังที่กล่าวมาแล้ว แต่ด้วยคุณสมบัติทางคณิตศาสตร์ที่มีความสะดวก ฟังก์ชันเคอร์เนลแบบปกติ จึงมักจะถูกนำมาใช้บ่อยครั้งคือ

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2) \quad \text{เมื่อ} \quad u = \frac{x - X_i}{h}$$

ฟังก์ชันความหนาแน่นของความน่าจะเป็น มาตรฐานปกติ นั่นคือ เป็น ความหนาแน่นปกติที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น 1 ดังนั้น $K_h(\mathrm{x}-\mathrm{X}_i)=\frac{1}{h\sqrt{2\pi}}\exp\left[\frac{-(x-X_i)^2}{2h^2}\right]$ จึงเป็นฟังก์ชันความหนาแน่นปกติ ที่มีค่าเฉลี่ย X_i และความแปรปรวน h^2 ค่าประมาณของความหนาแน่นแบบเคอร์เนล ที่มีความสัมพันธ์อย่างใกล้ชิดกับฮิสโตแกรม แม้ว่า อย่างแรกจะมีข้อดีหลายประการ ลองเปรียบเทียบ โครงสร้างของฮิสโตแกรม กับ ค่าประมาณของความหนาแน่นแบบเคอร์เนล โดยใช้จุดข้อมูลทั้ง 6 จุด ดังนี้ $X_1=-2.1, \mathrm{X}_2=-1.3, \mathrm{X}_3=-0.4, X_4=1.9, X_5=5.1, X_6=6.2$ สำหรับ ฮิสโตแกรมนั้นใน ตอนแรกจะแบ่งแกนนอนออกเป็นช่วงย่อย หรือ ที่ว่างซึ่งครอบคลุมช่วงของข้อมูล ในกรณีนี้มี 6 ช่องในแต่ละความกว้าง 2 เมื่อใดก็ตามที่ จุดข้อมูลตกอยู่ภายในช่วงย่อยนี้จะวางกล่องที่มีความสูง $\frac{1}{2}$ ลงไป หากมีจุดข้อมูลมากกว่าหนึ่ง อยู่ภายในที่ว่างเดียวกันก็จะวางกล่องซ้อนกันที่ด้านบนของกล่องแต่ละกล่อง สำหรับค่าประมาณของความหนาแน่นแบบเคอร์เนล จะวางเคอร์เนลปกติที่มีความแปรปรวน 2.25 (โดยแสดงเป็นเส้นประ) ในแต่ละจุดข้อมูล X_i ฟังก์ชันเคอร์เนลจะได้รับการรวมสรุปให้เป็นค่าประมาณของความหนาแน่นแบบเคอร์เนล (เส้นโค้งด้านบนของแข็ง) ความราบเรียบและเนียนของค่าประมาณของความหนาแน่นแบบเคอร์เนล จะปรากฏชัดเมื่อเทียบกับฮิสโตแกรมที่มีลักษณะแยกส่วน (Discreteness) ที่ไม่ต่อเนื่อง ซึ่งลักษณะที่ไม่ต่อเนื่องเช่นนี้ เป็นผลจากการไม่มีประสิทธิภาพเชิงสถิติโดยธรรมชาติของฮิสโตแกรม เองเมื่อเทียบกับตัวประมาณค่าแบบเคอร์เนล



ภาพที่ 2-1 การประมาณค่าแบบฮิสโตแกรม กับค่าประมาณการความหนาแน่นแบบเคอร์เนล

จากแผนภาพการเปรียบเทียบ ฮิสโตแกรม (ซ้าย) กับกราฟผลค่าประมาณการความหนาแน่นแบบเคอร์เนล (ขวา) ที่สร้างโดยใช้ข้อมูลชุดเดียวกัน โดย 6 ฟังก์ชันเคอร์เนลแต่ละฟังก์ชัน เป็นเส้นโค้งประสีแดง ส่วนผลค่าประมาณการความหนาแน่นแบบเคอร์เนลจะเป็นเส้นโค้งสีฟ้า

และจุดข้อมูลจะอยู่บนแกนนอนในการสร้างผลค่าประมาณการความหนาแน่นแบบเคอร์เนล พบว่ามีการตีความในสาขาวิชาอื่นนอกเหนือไปจากการประมาณค่าความหนาแน่นด้วย ตัวอย่างเช่นในวิชาอุณหพลศาสตร์ (Thermodynamics) โดยค่าประมาณนี้ จะเทียบเท่ากับ ปริมาณความร้อนที่ถูกสร้างขึ้นเมื่อฟังก์ชันเคอร์เนลความร้อน (วิธีแก้โจทย์พื้นฐานของ สมการความร้อน) อยู่ที่ตำแหน่ง X_i วิธีการในลักษณะเดียวกันนี้จะถูกนำไปใช้ในการสร้างตัวปฏิบัติการลาปลาซ ประเภทไม่ต่อเนื่องกับแบบจุด

การประมาณความหนาแน่นเคอร์เนล (Kernel Density Estimation) เป็นวิธีการประมาณความหนาแน่นที่นิยมใช้กันมากที่สุด (Silverman, 1986, pp. 35-36) เป็นวิธีการที่ค่าสังเกตแต่ละค่าจะมีน้ำหนักถ่วง เสนอเป็นครั้งแรก โดย Rosenblatt ในปี ค.ศ. 1956 ตัวประมาณค่าความหนาแน่นแบบเคอร์เนล ($\hat{f}(x)$) คือ

$$\hat{f}_h(x) = \frac{1}{nh}\sum_{i=1}^{k} K(\frac{x - X_i}{h})$$

โดยที่ h คือ ค่าพารามิเตอร์ปรับเรียบหรือแบนวิดจ์ (Smoother Parameter or Bandwidth)

x คือ ค่าเฉลี่ยของตัวแปรสุ่ม

X_i คือ ค่าของตัวแปรสุ่ม X โดยที่ $i = 1, 2, ..., n$

K คือ ฟังก์ชันเคอร์เนล (Kernel Function)

ทั้งนี้ Wolfgang Hardle (1990) อ้างถึงใน (ปิยะฉัตร ลีลาศิลปะศาสน์, 2549, หน้า 2) ได้สรุปฟังก์ชันเคอร์เนล ดังตารางที่ 2-1 ตารางรูปแบบของฟังก์ชันเคอร์เนล

ตารางที่ 2-1 ตารางรูปแบบของฟังก์ชันเคอร์เนล

ฟังก์ชันเคอร์เนล	$K(\mathrm{u})$
1. ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov)	$\frac{3}{4}\left(1-u^2\right)I\left(\lvert u\rvert \le 1\right)$
2. ฟังก์ชันเคอร์เนลแบบควอติค (Quartic)	$\frac{15}{16}\left(1-u^2\right)^2 I\left(\lvert u\rvert \le 1\right)$
3. ฟังก์ชันเคอร์เนลไตรเวท (Triweight)	$\frac{35}{32}\left(1-u^2\right)^3 I\left(\lvert u\rvert \le 1\right)$
4. ฟังก์ชันเคอร์เนลแบบเกาส์เซียน (Gaussian)	$\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right)$
5. ฟังก์ชันเคอร์เนลแบบสามเหลี่ยม (Triangular)	$\left(1-\lvert u\rvert\right)I(\lvert u\rvert \le 1)$
6. ฟังก์ชันเคอร์เนลแบบยูนิฟอร์ม (Uniform)	$\frac{1}{2}I\left(\lvert u\rvert \le 1\right)$
7. ฟังก์ชันเคอร์เนลแบบคอซายอัส (Cosinus)	$\frac{\pi}{4}\cos\left(\frac{\pi}{2}u\right)I\left(\lvert u\rvert \le 1\right)$

โดยที่ $u = \frac{x - X_i}{h}$

จากตารางที่ 2-1 ที่แสดงถึงฟังก์ชันเคอร์เนลชนิดต่าง ๆ สามารถหาสมการฟังก์ชันเคอร์เนลได้จาก $K(\mathrm{u,p}) = \frac{(1-\mathrm{u}^2)^p}{2^{2p+1}B(\mathrm{p}+1)(\mathrm{p}+1)}$ โดยที่ $\{\lvert u \le 1\rvert\}$............................(2-1)

โดยที่ u คือ ตัวแปร u โดยที่ $u = \frac{x - X_i}{h}$

p คือ ค่าคงที่ p โดยที่ p มีค่าตั้งแต่ 0 เป็นต้นไป

และ $B(\mathrm{a,b}) = \frac{\Gamma(\mathrm{a})\Gamma(\mathrm{b})}{\Gamma(\mathrm{a+b})}$ โดยที่ $\Gamma(\mathrm{c}) = (\mathrm{c}-1)!$ เมื่อ c เป็นค่าคงที่ ดังนั้นฟังก์ชันเคอร์เนลมีรูปแบบแตกต่างกันไปตามลักษณะแนวโน้มของการแจกแจงของข้อมูล ยกตัวอย่างเช่น

ในกรณีที่ p = 0 จากสมการที่ 2-1 จะได้ว่า

$$K(\mathrm{u,p})=\frac{(1-\mathrm{u}^2)^p}{2^{2p+1}B(\mathrm{p}+1)(\mathrm{p}+1)}$$ โดยที่ $\{|u\le 1|\}$

$$K(\mathrm{u},0)=\frac{(1-\mathrm{u}^2)^0}{2^{2(0)+1}B(0+1)(0+1)}$$ โดยที่ $\{|u\le 1|\}$

$$=\frac{1}{2B(1,1)}$$

จาก $B(\mathrm{a,b})=\frac{\Gamma(\mathrm{a})\Gamma(\mathrm{b})}{\Gamma(\mathrm{a+b})}$ จะกล่าวได้ว่า $B(1,1)=\frac{\Gamma(1)\Gamma(1)}{\Gamma(1+1)}$

$$K(u,0)=\frac{1}{2}$$ โดยที่ $\{|u\le 1|\}$

ฟังก์ชันเคอร์เนลที่ได้ก็คือ ฟังก์ชันเคอร์เนลแบบยูนิฟอร์ม ในกรณีที่ p=1 จากสมการที่ 2-1 จะได้ว่า

$$K(\mathrm{u,p})=\frac{(1-\mathrm{u}^2)^p}{2^{2p+1}B(\mathrm{p}+1)(\mathrm{p}+1)}$$ โดยที่ $\{|u\le 1|\}$

$$K(\mathrm{u},1)=\frac{(1-\mathrm{u}^2)^1}{2^{2p+1}B(1+1)(1+1)}$$

$$K(\mathrm{u},1)=\frac{(1-\mathrm{u}^2)^1}{2^3B(2)(2)}$$

จาก $B(\mathrm{a,b})=\frac{\Gamma(\mathrm{a})\Gamma(\mathrm{b})}{\Gamma(\mathrm{a+b})}$ จะกล่าวได้ว่า $B(1,1)=\frac{\Gamma(2)\Gamma(2)}{\Gamma(2+2)}$

$$\mathrm{K}(\mathrm{u},1)=\frac{3}{4}(1-\mathrm{u}^2)$$ โดยที่ $\{|u\le 1|\}$

ฟังก์ชันเคอร์เนลที่ได้ก็คือ ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

คุณสมบัติของฟังก์ชันเคอร์เนล

1. ฟังก์ชันเคอร์เนลเป็นฟังก์ชันแบบสมมาตร โดยมีค่าระหว่าง 0 ถึง 1 ซึ่งกล่าวได้ว่า $K(\mathrm{u})=\mathrm{K}(-\mathrm{u})$

2. ถ้าฟังก์ชันเคอร์เนลเป็นฟังก์ชันความหนาแน่น ตัวประมาณค่าความหนาแน่นแบบเคอร์เนลจะเป็นฟังก์ชันความหนาแน่นด้วยเช่นกันกล่าวคือ ถ้า $\int_{-\infty}^{\infty}\mathrm{K}(\mathrm{u})d\mathrm{u}=1$ แล้ว $\int_{-\infty}^{\infty}\hat{\mathrm{f}}_{\mathrm{h}}(\mathrm{u})d\mathrm{u}=1$

3. เมื่อกำหนดค่าแบนวิดจ์ h และฟังก์ชันความหนาแน่นเคอร์เนล K ในข้อมูล 1 ชุด จะมีฟังก์ชันเคอร์เนลได้เพียง 1 ฟังก์ชัน

4. ถ้าฟังก์ชันเคอร์เนล K สามารถหาอนุพันธ์ได้อย่างต่อเนื่อง ตัวประมาณค่าความหนาแน่นแบบเคอร์เนลก็จะสามารถหาอนุพันธ์อย่างต่อเนื่องเช่นกัน

5. $\int_{-\infty}^{\infty} \mathrm{u}^{\mathrm{j}}\mathrm{K}(\mathrm{u})\mathrm{du} = 0$ สำหรับ $j = 1, 2, ..., \mathrm{k}-1$

6. $\int_{-\infty}^{\infty} \mathrm{u}^{\mathrm{k}}\mathrm{K}(\mathrm{u})\mathrm{du} \neq 0$

โดยทั่วไปแล้วจะเรียกฟังก์ชันเคอร์เนล K ว่าฟังก์ชันเคอร์เนลลำดับที่ k ซึ่งส่วนใหญ่จะศึกษาในกรณีที่ $k = 2$ เนื่องจากคุณสมบัติของฟังก์ชันเคอร์เนล นั่นคือ $\int_{-\infty}^{\infty} \mathrm{u}^{\mathrm{k}}\mathrm{K}(\mathrm{u})\mathrm{du} \neq 0$ อาจจะกล่าวได้ว่าเมื่อ $k = 2$ ค่าของ K ที่ได้จะมีค่าเป็นบวก กล่าวคือ ฟังก์ชันเคอร์เนล K จะกลายเป็นฟังก์ชันความหนาแน่นของความน่าจะเป็น และตัวประมาณความหนาแน่นแบบเคอร์เนลจะเข้าไปแทนที่ค่าสังเกต X_i แต่ละค่า แต่สำหรับกรณีที่ $k > 2$ ค่าของ K ที่ได้อาจจะมีค่าเป็นค่าลบและส่งผลให้ตัวประมาณค่าความหนาแน่นแบบเคอร์เนลไม่มีการเคลื่อนที่ และค่าประมาณความหนาแน่นแบบเคอร์เนลก็จะเป็นค่าลบด้วยเช่นกัน

ในการวิจัยนี้จะศึกษาการประมาณความหนาแน่นแบบเคอร์เนล แบบตัวแปรสุ่มเดิมในกรณีที่ $k = 1$ และ $k = 2$ โดยที่เมื่อ $k = 1$ จะได้ $\int_{-\infty}^{\infty} \mathrm{K}(\mathrm{x})\mathrm{dx} \neq 0$ และเมื่อ $k = 2$ จะได้ว่า $\int_{-\infty}^{\infty} \mathrm{x}^2\mathrm{K}(\mathrm{x})\mathrm{dx} \neq 0$

ฟังก์ชันเคอร์เนลที่ใช้ในการวิจัย

ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov Kernel Function) มีรูปแบบฟังก์ชัน คือ

$$K(\mathrm{u}) = \frac{3}{4}(1 - \mathrm{u}^2)\,\mathrm{I}(|u| \leq 1)$$

โดยที่ $u = \dfrac{x - X_i}{h}$

h คือ ค่าแบนวิดจ์ มีค่าเป็นจำนวนจริงบวก

X_i คือ ค่าของตัวแปรสุ่ม X โดยที่ $i = 1, 2, ..., n$

ประสิทธิภาพของฟังก์ชันเคอร์เนล

Silverman (1986, pp. 46-48) ได้กล่าวถึงฟังก์ชันเคอร์เนลที่เหมาะสมว่าควรจะเป็นฟังก์ชันเคอร์เนลที่ทำให้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเฉลี่ยรวม (*MISE*) มีค่าต่ำ และได้ให้สูตรการคำนวณหาประสิทธิภาพสัมพัทธ์ (Relative Efficiency) ของฟังก์ชันเคอร์เนล ดังนี้

$$eff(\mathrm{K}) = \frac{5}{4}C(\mathrm{K})\left\{\int f''(x)d(x)n^{\frac{-4}{5}}\right\}$$

โดยที่ $C(\mathrm{K}) = \mathrm{k}_2^{\frac{2}{5}}\left\{\int K(\mathrm{t})^2 dx\right\}^{\frac{4}{5}}$

สำหรับประสิทธิภาพของฟังก์ชันเคอร์เนลที่ศึกษาแสดงในตารางที่ 2-2 แสดงค่าประสิทธิภาพสัมพัทธ์ของฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

ตารางที่ 2-2 ค่าประสิทธิภาพสัมพัทธ์ของฟังก์ชันเคอร์เนล

ฟังก์ชันเคอร์เนล	ค่าประสิทธิภาพสัมพัทธ์
ฟังก์ชันอีพานิชนิคอฟ (Epanechnikov)	1.0000
ฟังก์ชันควอติค (Quartic)	0.9939
ฟังก์ชันสามเหลี่ยม (Triangular)	0.9859
ฟังก์ชันเกาส์เซียน (Gaussian)	0.9512

2. วิธีการเลือกแบนวิดจ์ (Bandwidth Selection)

2.1 วิธี Rules of thumb (Deheuvels, 1977., Silverman, 1986)

วิธีการเลือกแบนวิดจ์ด้วยวิธี Rules of thumb เป็นวิธีการเลือกแบนวิดจ์ที่ง่ายที่สุด ทำได้โดยการแทนค่าส่วนที่ไม่ทราบค่า นั่นคือ $\int f''(x)^2 dx$ ลงในค่าแบนวิดจ์ที่เหมาะสม (Optimal bandwidth)

$$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}}) \quad \text{..........}(2\text{-}1)$$

เมื่อ h_{ROT} คือ แบนวิดจ์ที่ได้จากวิธี Rules of Thumb

n คือ ขนาดตัวอย่าง

$\hat{\sigma}$ คือ ค่าประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม $X_i; i = 1,2,...n$

2.2 วิธี Silverman's rules of thumb (Silverman, 1986)

วิธีการเลือกแบนวิดจ์ด้วยวิธี Silverman's Rules of Thumb เป็นวิธีการเลือกแบนวิดจ์ที่ถูกปรับปรุงมาจากการเลือกแบนวิดจ์ด้วยวิธี Rules of Thumb

$$h_{SROT} = 0.9A(n^{\frac{1}{5}})$$

เมื่อ h_{SROT} คือ แบนวิดจ์ที่ได้จากวิธี Silverman’s Rules of Thumb

n คือ ขนาดตัวอย่าง

A คือ ค่าที่น้อยที่สุดระหว่างค่าประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม $X_i; i = 1,2,...n$ กับ $\frac{IQR}{1.34} = \min[\hat{\sigma}.\frac{IQR}{1.34}]$

IQR คือ พิสัยระหว่างควอไทล์ (Interquartile range)

$\hat{\sigma}$ คือ ค่าประมาณส่วนเบี่ยงเบนมาตรฐานของตัวแปรสุ่ม $X_i; i = 1,2,...n$

3. เกณฑ์การพิจารณา

การพิจารณาถึงความเหมาะสมของตัวประมาณค่าโดยทั่วไปจะพิจารณาจากความแตกต่างระหว่างฟังก์ชันความหนาแน่นที่แท้จริง $f(x)$ และฟังก์ชันความหนาแน่นที่ประมาณขึ้น $f_h(x)$ ดังนี้

3.1 ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Squared Error: *MSE*)

เป็นวิธีการวัดความแตกต่างของตัวประมาณความหนาแน่น $\hat{f}_h(x)$ จากความหนาแน่นที่แท้จริง ซึ่งมีหลายวิธีที่นำมาศึกษา แต่เมื่อพิจารณาถึงการประมาณแบบจุด มักจะนิยมใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง โดย

$$MSE(\hat{f}(x)) = E\left[\hat{f}_h(x) - f(x)\right)^2$$

$$MSE(\hat{f}(x)) = E\left[\hat{f}_h(x) - f(x)\right)^2 + \operatorname{var}\hat{f}(x) \quad \text{.......................................(2-2)}$$

จากสมการที่ 2-3 สามารถเขียนรูปแบบอย่างง่าย คือ

$$MSE(\hat{f}(x)) = \left[bias\hat{f}(x)\right]^2 + \operatorname{var}\hat{f}(x) \quad \text{..(2-3)}$$

เนื่องจาก

$$bias\hat{f}(x) = \frac{1}{2}h^2 f''(x)k_2 + o(h^2)$$

$$\operatorname{var}(\hat{f}(x)) = n^{-1}h^{-1}\int_{-\infty}^{\infty} K(t)^2\,dt + o(nh)^{-1}$$

ดังนั้น

$$MSE(\hat{f}(x)) = \left[\frac{1}{2}h^2 f''(x)k_2 + o(h^2)\right]^2 + n^{-1}h^{-1}\int_{-\infty}^{\infty} K(t)^2\,dt + o(\mathrm{nh})^{-1} \quad \text{...........(2-4)}$$

4. การแจกแจงต่าง ๆ ที่ใช้ในงานวิจัย

4.1 การแจกแจงแบบโคชี (Cauchy Distribution)

โอกุส ติน ลุยส์ โคชี (Augustin Louis Cauchy, 1857) นักคณิตศาสตร์ชาวฝรั่งเศส เป็นผู้ค้นพบการแจกแจงแบบโคชี (Cauchy Distribution) จะพบการแจกแจงแบบโคชี ในงานทางด้านวิทยาศาสตร์และวิศวกรรมศาสตร์ เช่น ในกลศาสตร์และทฤษฎีไฟฟ้า (มานพ ศรีวรกุล, 2547) ตัวแปรสุ่มต่อเนื่อง X มีการแจกแจงแบบโคชีด้วยพารามิเตอร์ θ โดยที่ $-\infty<\theta<\infty$ และ $\beta>0$ ฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงแบบโคชี คือ

$$f(\mathrm{x})=\frac{1}{\pi\beta\left[1+\left(\frac{x-\theta}{\beta}\right)^2\right]},-\infty<\theta<\infty$$

เขียนแทนด้วย $x\sim C(\theta,\beta)$ หาก $\theta=0$ และ $\beta=1$ จะเขียนแทนด้วย $x\sim C(0,1)$ และจะเรียกว่า x มีการแจกแจงแบบโคชีมาตรฐาน มานพ ศรีวรกุล (2547)

สถิติวิเคราะห์เชิงพรรณนาสำหรับการแจกแจงโคชี

ฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงโคชี

นิยาม ถ้าตัวแปรสุ่ม X มีการแจกแจงแบบโคชีแล้วฟังก์ชันความหนาแน่นน่าจะเป็น $f(x)$ ของตัวแปรสุ่ม X คือ

$$f(\mathrm{x})=\frac{1}{\pi\beta\left[1+\left(\frac{x-\theta}{\beta}\right)^2\right]},-\infty<x<\infty,-\infty<n<\infty\,\text{และ}\,b>0$$

จากนิยามของฟังก์ชันความหนาแน่นน่าจะเป็นของตัวแปรสุ่มต่อเนื่อง ซึ่งกล่าวว่า ถ้า X เป็นตัวแปรสุ่มต่อเนื่องแล้ว $f(x)$ จะเป็นฟังก์ชันความหนาแน่นน่าจะเป็นของตัวแปรสุ่ม X ก็ต่อเมื่อ $f(x)$ สอดคล้องกับคุณสมบัติต่อไปนี้

1) $f(x)\geq 0$

2) $\int_{-\infty}^{\infty} f(x)dx=1$

เพื่อแสดงว่าฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงแบบโคชีมีคุณสมบัติดังกล่าวสามารถพิจารณาได้ ดังนี้

จากนิยามฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงแบบโคชี จะพบว่าตัวแปรสุ่ม $-\infty \le X \le \infty$ และพารามิเตอร์ $b > 0$ ดังนั้นจะทำให้ $f(x)$ มีค่ามากกว่าศูนย์หรือเท่ากับศูนย์ด้วยนั่น คือ $f(x) \ge 0$

สำหรับการแสดงว่า $\int_{-\infty}^{\infty} f(x)dx = 1$ พิสูจน์ได้ดังนี้

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]} dx$$

$$= \frac{1}{\pi b}\int_{-\infty}^{\infty} \frac{1}{1+\left(\frac{x-a}{b}\right)^2} dx$$

$$= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{b}{b^2+(x-a)^2} dx$$

$$= \frac{b}{\pi}\int_{-\infty}^{\infty} \frac{1}{b^2+(x-a)^2} dx$$

เนื่องจาก $\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{a} f(x)dx + \int_{a}^{\infty} f(x)dx$

$$\int_{-\infty}^{\infty} f(x)dx = \frac{b}{\pi}\int_{-\infty}^{a} \frac{1}{b^2+(x-a)^2} d(x-a) + \frac{b}{\pi}\int_{a}^{\infty} \frac{1}{b^2+(x-a)^2} d(x-a)$$

$$= \frac{b}{\pi}\left[\frac{1}{b}\tan^{-1}\frac{(x-a)}{b}\right]_{-\infty}^{a} + \frac{b}{\pi}\left[\frac{1}{b}\tan^{-1}\frac{(x-a)}{b}\right]_{a}^{\infty}$$

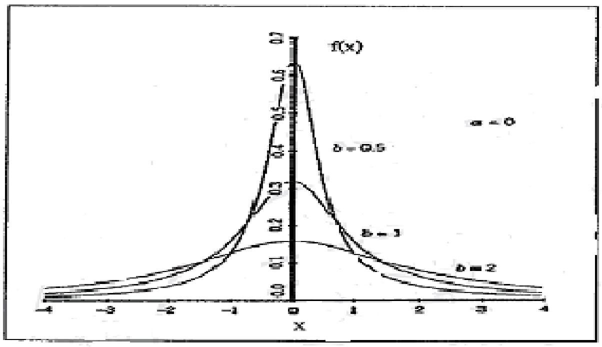
$$= \frac{1}{\pi}\left[0-\left(-\frac{\pi}{2}\right)\right] + \frac{1}{\pi}\left[\frac{\pi}{2}-0\right]$$

$$= \frac{1}{\pi}\left(\frac{\pi}{2}+\frac{\pi}{2}\right)$$

$$= 1$$

ดังนั้นสรุปได้ว่า ฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงโคชีมีความสอดคล้องกับคุณสมบัติของฟังก์ชันความน่าจะเป็นหนาแน่นน่าจะเป็นของตัวแปรสุ่มต่อเนื่องที่กล่าวในขั้นต้น

จากลักษณะฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงโคชี เมื่อนำมาเขียนกราฟจะพบว่าลักษณะของกราฟมีความสมมาตรรอบค่าพารามิเตอร์ a และเมื่อค่าพารามิเตอร์ a มีค่าคงที่ กราฟจะมีลักษณะโด่งมากขึ้นเมื่อพารามิเตอร์ b มีค่าลดลง ดังแสดงในรูป



ภาพที่ 2-2 ฟังก์ชันความหนาแน่นน่าจะเป็นของการแจกแจงโคชีเมื่อพารามิเตอร์ a=0 และ b=0.5, 1, 2

ฟังก์ชันการแจกแจงสะสมของการแจกแจงโคชี

ทฤษฎีบท 3.5 ถ้าตัวแปรสุ่ม X มีการแจกแจงโคชี ด้วยฟังก์ชันความหนาแน่นน่าจะเป็น $f(x)$ แล้วฟังก์ชันการแจกแจงสะสม (Cumulative Distribution Function) ของการแจกแจงโคชี คือ

$$F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x-a}{b}\right)$$

พิสูจน์จาก $F(x) = P(X \le x)$

$$= \int_{-\infty}^{x} \frac{1}{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]} dx$$

$$= \frac{1}{\pi b}\int_{-\infty}^{x} \frac{1}{1+\left(\frac{x-a}{b}\right)^2} dx$$

$$F(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{b}{b^2+(x-a)^2} dx$$

$$= \frac{b}{\pi}\int_{-\infty}^{x} \frac{1}{b^2+(x-a)^2} d(x-a)$$

$$= \frac{b}{\pi}\left[\frac{1}{b}\tan^{-1}\frac{(x-a)}{b}\right]_{-\infty}^{x}$$
$$= \frac{1}{\pi}\left[\tan^{-1}\left(\frac{x-a}{b}\right)-\left(-\frac{\pi}{2}\right)\right]$$
$$= \frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(\frac{x-a}{b}\right)$$

สถิติบรรยายสำหรับการแจกแจงโคชี

สำหรับการแจกแจงโคชีนั้นมีลักษณะพิเศษกว่าการแจกแจงลอกนอร์มอล เนื่องจากสถิติพรรณนาของการแจกแจงโคชีนั้นมีเฉพาะค่ามัธยฐานและค่าฐานนิยมเท่านั้น ในส่วนของค่าเฉลี่ย ความแปรปรวน สัมประสิทธิ์ความเบ้ และสัมประสิทธิ์ความโด่งนั้นไม่สามารถหาค่าของการอินทิเกรดได้ ซึ่งค่ามัธยฐานและค่าฐานนิยมนั้นสามารถหาได้ ดังนี้

ค่ามัธยฐานและค่าฐานนิยมของการแจกแจงโคชี

ทฤษฎี 3.6 ถ้าตัวแปรสุ่ม X มีการแจกแจงโคชี ด้วยพารามิเตอร์ a,b โดยที่ $b>0$ แล้ว ค่ามัธยฐาน (median) และค่าฐานนิยม (mode) ของการแจกแจงโคชี กำหนดโดย

ค่ามัธยฐาน $= a$

ค่าฐานนิยม $= a$

การหาค่ามัธยฐานและค่าฐานนิยม แสดงได้ดังนี้

ค่ามัธยฐาน คือ ค่าของ X ที่ทำให้พื้นที่ใต้โค้งฟังก์ชันความหนาแน่นน่าจะเป็น: $f(x)$

ด้านซ้ายมือของ x มีค่าเท่ากับด้านขวามือเท่ากับ $\frac{1}{2}$

จาก $\int_{-\infty}^{median} f(x)dx = \frac{1}{2}$

$$\frac{1}{2} = \int_{-\infty}^{median} \frac{1}{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]}dx$$

$$= \frac{1}{\pi b}\int_{-\infty}^{median} \frac{1}{1+\left(\frac{x-a}{b}\right)^2}dx$$

$$= \frac{1}{\pi}\int_{-\infty}^{median}\frac{b}{b^2+(x-a)^2}dx$$

$$= \frac{b}{\pi}\int_{-\infty}^{median}\frac{1}{b^2+(x-a)^2}d(x-a)$$

$$= \frac{b}{\pi}\left[\frac{1}{b}\tan^{-1}\frac{(x-a)}{b}\right]_{-\infty}^{median}$$

$$= \frac{1}{\pi}\left[\tan^{-1}\left(\frac{median-a}{b}\right)-\left(-\frac{\pi}{2}\right)\right]$$

$$\frac{1}{2} = \frac{1}{2}+\frac{1}{\pi}\tan^{-1}\left(\frac{median-a}{b}\right)$$

$$0 = \frac{1}{\pi}\tan^{-1}\left(\frac{median-a}{b}\right)$$

$$0 = \tan^{-1}\left(\frac{median-a}{b}\right)$$

ดังนั้น $\frac{median-a}{b} = 0$

ดังนั้น $median = a$

นั่นคือค่ามัธยฐานของการแจกแจงโคชี คือ a

ค่าฐานนิยม คือค่าของ x ที่ให้ค่า $f(\mathrm{x})$ สูงที่สุด ซึ่งสามารถหาได้ ดังนี้

จาก $f(x) = \dfrac{1}{\pi b\left[1+\left(\frac{x-a}{b}\right)^2\right]}$

ดังนั้น $\dfrac{df(x)}{dx} = \dfrac{2\pi\left(\frac{x-a}{b}\right)}{(\pi b)^2\left[1+\left(\frac{x-a}{b}\right)^2\right]}$

$$0 = \frac{2\pi\left(\frac{x-a}{b}\right)}{(\pi b)^2\left[1+\left(\frac{x-a}{b}\right)^2\right]}$$

$$0 = 2\pi\left(\frac{x-a}{b}\right)$$

$$0 = x-a$$

$$x = a$$

ดังนั้นค่าฐานนิยมของการแจกแจงโคชี คือ a

3. การแจกแจงแบบปกติ (Normal Distribution)

การแจกแจงแบบปกติเป็นการแจกแจงแบบต่อเนื่องที่พบมากที่สุดและมีความสำคัญมาก ทั้งทางด้านสถิติประยุกต์และสามารถนำไปใช้ประโยชน์ในการประมาณค่าของประชากร และทดสอบสมมติฐานทางสถิติ

อับราฮัม เดอร์มัวร์ (Abraham De Moire, pp. 1667-1754) นักคณิตศาสตร์ชาวฝรั่งเศส เป็นผู้ค้นพบการแจกแจงแบบปกติเมื่อปี พ.ศ. 1733 ต่อมา ปิแอร์ ลาปลาส (Pierre Laplace, pp. 1749-1827)

นักคณิตศาสตร์ชาวฝรั่งเศส กับคาร์ล เกาส์ (Carl Gauss, pp. 1777-1855) เป็นนักคณิตศาสตร์และวิทยาศาสตร์ชาวเยอรมัน ได้ค้นพบการแจกแจงแบบปกติโดยไม่ทราบผลงานของอับราฮัม เดอร์มัวร์ มาก่อนเลย ซึ่งพบว่าการแจกแจงของความคลาดเคลื่อนในการวัดทางวิทยาศาสตร์กายภาพสามารถประมาณได้อย่างใกล้เคียงโดยใช้โค้งปกติ ซึ่งเขาเรียกว่า “โค้งปกติของความคลาดเคลื่อน” (The Laws of Chance) ผลงานของลาปลาส และเกาส์ เป็นที่รู้จักกันอย่างแพร่หลายและนำไปใช้ประโยชน์กันอย่างกว้างขวาง (วินัย วีระวัฒนานนท์, 2537) ในบางครั้งมักจะเรียก การแจกแจงแบบปกติว่า “การแจกแจงลาปลาส” (Laplacian Distribution) หรือ “การแจกแจงเกาส์เซียน” (Gaussian Distribution) การแจกแจงแบบปกติมีฟังก์ชันความหนาแน่นน่าจะเป็น (pdf) คือ

ให้ X เป็นตัวแปรสุ่ม กล่าวว่า X มีการแจกแจงแบบปกติ ถ้า X มีฟังก์ชันกำหนดโดย

$$f(\mathrm{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\mathrm{x}-\mu)^2}{2\sigma^2} \right]$$

เมื่อ σ^2 คือ ความแปรปรวนของประชากร

σ คือ ส่วนเบี่ยงเบนมาตรฐานของประชากร

μ คือ ค่าเฉลี่ยของประชากร

π = 3.14159

คุณสมบัติของการแจกแจงแบบปกติ

1. ลักษณะโค้งเป็นรูประฆังคว่ำ (bell shaped) สมมาตรกับแกนตั้งที่ลากผ่านเส้นค่าเฉลี่ย μ ฟังก์ชันความหนาแน่นน่าจะเป็น มีลักษณะเป็นรูประฆังคว่ำที่สมมาตร ดังนั้น ค่าเฉลี่ยของตัวแปรสุ่ม X คือ

$$\mu_x = E(\mathrm{X}) = \mu$$

ความแปรปรวนของตัวแปรสุ่ม X คือ

$$\sigma_x^2 = E[(x-\mu)^2] = \sigma^2$$

2. กราฟลดลงอย่างต่อเนื่องทั้งสองข้าง มีจุดเปลี่ยนเว้าที่ $x = \mu \pm \sigma'$ โดยปลายโค้งจะเข้าใกล้แกน X เมื่อ X มีค่าห่างจาก μ ออกไปแต่จะไม่ตัดแกน X

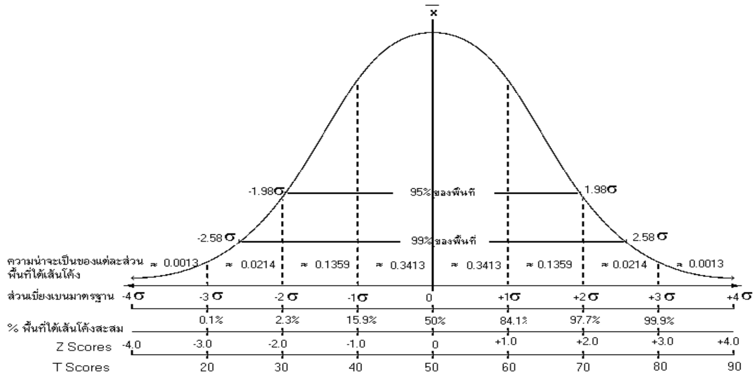
3. เป็นโค้งที่มีจุดสูงสุดเพียงจุดเดียว (Unimodal) อยู่ที่ $x = \mu$ ซึ่งเป็นค่าฐานนิยม

4. ค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยม มีค่าเท่ากันอยู่ที่จุด $x = \mu$

5. ค่าความโด่ง (Kurtosis) เท่ากับ 3 และค่าความเบ้ (Skewness) เท่ากับ 0

6. μ และ σ^2 เป็นค่าพารามิเตอร์ของการแจกแจงแบบปกติโดยที่ μ และ σ^2 จะเป็นตัวกำหนดตำแหน่งจุดของเส้นโค้งและลักษณะของเส้นโค้งว่าจะเป็นแบบแบน หรือโด่ง

7. พื้นที่ใต้โค้งปกติที่อยู่ระหว่าง $x = \mu \pm \sigma'$, $\mu \pm 2\sigma$ และ $\mu \pm 3\sigma$ จะมีค่าเป็น 68.27%, 95.45% และ 99.73% ตามลำดับ



ในกรณีที่การแจกแจงแบบปกติ ที่มีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ 1 จะเรียกการแจกแจงแบบปกตินี้ว่า การแจกแจงแบบปกติมาตรฐาน (Standard Normal Distribution) สามารถเขียนได้ว่า $Z \sim N(0,1)$ เมื่อ Z คือตัวแปรที่มีการแจกแจงแบบปกติมาตรฐาน

4. การประมาณค่าแบบนอนพาราเมตริก

Nonparametric Regression เป็นการวิเคราะห์สมการถดถอยที่เป็นทางเลือกกรณีเมื่อข้อตกลงเกี่ยวกับส่วนเกี่ยวกับส่วนเหลือ (Residual) หรือความคลาดเคลื่อน (Error) มิได้เป็นไปตามข้อตกลงในความจริงแล้วจะพบว่า u มิได้เป็นจริงตามข้อตกลง เราก็สามารถใช้วิธี OLS ได้เพียงแต่ต้องปรับข้อมูลตามวิธีที่เหมาะสม หรือเปลี่ยนวิธีการประมาณค่าที่เหมาะสม แต่ในบางครั้งการเลือก

วิธี Nonparametric Regression อาจยังมีเหตุผลอื่น ๆ อีก คือ การไม่ประสงค์จะกำหนดหรือทราบสมการความสัมพันธ์ระหว่าง Y กับ X หรือ Y กับ $\mathrm{X's}$ คือไม่ต้องยุ่งเกี่ยวหรือมีภาระต้องทราบต้องพัฒนาสมการ $Y = f(\mathrm{X's})$ เช่นที่ปฏิบัติใน Parametric Regression แต่มุ่งไปที่การคาดคะเนข้อมูลที่เป็น Nonparametric Regression ใช้มักจะเป็นข้อมูลอันดับของคะแนน (Rank) มิใช่คะแนน

เหตุนี้วิธีการประมาณค่าจึงเกี่ยวข้องกับมัธยฐาน (Median) เช่นวิธี Weighted Median วิธี $Minimize\sum_{i}^{n}\left\{rang(e_i)-\frac{n+1}{2}\right\}e_i$ หรืออาจใช้ตัวคะแนนโดยรอบของ Y_i มาเฉลี่ยกันด้วยน้ำหนักที่เหมาะสมเช่นวิธี Locally Weighted Average Predictor (LWAP) เป็นต้น (มนตรี พิริยะกุล, 2544, หน้า 515)

การประมาณค่าแบบนอนพาราเมตริก คือ วิธีการทางสถิติที่ช่วยให้ได้ข้อมูลในรูปแบบของฟังก์ชันที่เหมาะสม ในกรณีที่ไม่มีเงื่อนไขข้อจำกัดต่าง ๆ ทางทฤษฎีเลย ดังนั้น ขั้นตอนการประมาณค่าแบบนอนพาราเมตริกจึงไม่มีค่าพารามิเตอร์ที่สำคัญ ๆ เกี่ยวข้อง มีเทคนิคแบบนอนพาราเมตริกที่สำคัญ ๆ 2 ประเภท คือ เครือข่ายคล้ายเส้นใยประสาท (Artificial Neural Networks) และการประมาณค่าแบบฟังก์ชันเคอร์เนล (Kernel Estimation) เครือข่ายคล้ายเส้นใยประสาทจะสร้างตัวแบบฟังก์ชันที่ไม่ทราบค่าออกมาในรูปของผลรวมถ่วงน้ำหนักของค่าผลรวมหลายตัว ที่ปกติจะถูกเลือกให้เป็นเส้นโค้งโลจิท ที่แต่ละโค้งคือฟังก์ชันของตัวแปรเชิงอธิบายที่เกี่ยวข้องทั้งหมด ซึ่งเป็นรูปแบบของฟังก์ชันที่ยืดหยุ่นมาก ๆ ที่ในการประมาณค่าเป็นต้องใช้ชุดคำสั่งค้นหาแบบซ้ำ ๆ แบบค่ากำลังสองน้อยที่สุดไม่เป็นเชิงเส้นที่เป็นไปตามระดับความลาดชัน การประมาณค่าแบบฟังก์ชันเคอร์เนล จะกำหนดฟังก์ชัน

$$y = m(x) + e_i$$

เมื่อ $m(x)$ คือ ค่าคาดหวังตามเงื่อนไขของ y

โดยที่ไม่มีลักษณะของพารามิเตอร์เลย และความหนาแน่นของความคลาดเคลื่อน e_i ไม่ได้ถูกกำหนดโดยสิ้นเชิง ส่วนค่าที่ได้จากการสังเกตของ y_i และ x_i จำนวน N ค่าก็จะถูกนำมาประมาณค่าฟังก์ชันความหนาแน่นร่วมสำหรับ y และ x ความหนาแน่น ณ จุด $(\mathrm{y}_0, \mathrm{x}_0)$ ก็จะถูกประมาณค่าโดยดูว่าสัดส่วนใดของค่าทีได้จากการสังเกตจำนวน N ค่า “เข้าใกล้” จุด $(\mathrm{y}_0, \mathrm{x}_0)$

ตัวแบบการถดถอยเชิงเส้น $Y_i = m(\mathrm{X}_i) + \mathrm{b} + \sigma(\mathrm{X}_i)$ เมื่อ $\sigma(\mathrm{X}_i) \in t$ และขึ้นอยู่กับข้อตกลงเบื้องต้นสำคัญ ๆ จำนวนหนึ่ง ที่สำคัญมากที่สุดก็ คือ ความสัมพันธ์เชิงเส้นระหว่างตัวแปรร่วม X_t กับตัวแปรอิสระ Y_t วิธีการหนึ่งในการสร้างตัวแบบการถดถอยเชิงเส้นมาตรฐาน ก็คือ

ต้องขยายออกไปสู่รูปแบบที่มิใช่เชิงเส้น (nonlinear form) ให้เป็น $Y_i = m(\mathrm{X}_i) + \sigma(\mathrm{X}_i)$ เมื่อ $\sigma(\mathrm{X}_i) \in t$ โดย m ที่เป็นลบและ σ เป็นฟังก์ชันของค่าจริงและเป็นลำดับที่แจกแจงเหมือนกันและเป็นอิสระของ ตัวแปรสุ่มปกติมาตรฐาน จากการศึกษาข้อมูล 5 ชุด (3 ชุด คือ ข้อมูลจริง อีก 2 ชุดคือ ข้อมูลจำลอง) เพื่อแสดงให้เห็นว่าการถดถอยกำลังสองน้อยที่สุดของฟังก์ชันเคอร์เนลถ่วงน้ำหนัก สามารถนำไปใช้ประมาณค่าฟังก์ชันกลางที่มีอยู่ คือ $m(x) = E\left[Y / \mathrm{X}_t = x\right]$ มีผลงานมากมายเกี่ยวกับการถดถอยไม่เชิงเส้นและเทคนิคการประมาณค่าแบบนอนพาราเมตริก อ้างอิงไปที่หนังสือที่มีชื่อเสียงของ Fan and Yao และผลงานมากมายของ Bjerve and Doksum (1993, p. 890) กล่าวอย่างไม่ละเอียดนัก วิธีที่จะประมาณค่า m สามารถทำได้ดังต่อไปนี้ ขั้นแรก เลือกเอาชุดของจุดกำหนดที่มีระยะห่างเท่ากัน ไปไว้ที่ช่วงห่างภายในของการสนับสนุนตัวแปรร่วม X_t ต่อจากนั้น ณ จุดกำหนดแต่ละจุด แก้โจทย์กำลังสองน้อยที่สุดของฟังก์ชันเคอร์เนลให้เข้าได้พอดีเฉพาะที่กับลำดับแบบพหุนาม

สำหรับคำว่า “ฟังก์ชันเคอร์เนลถ่วงน้ำหนัก” นั้น หมายความว่า ข้อมูลได้รับการถ่วงน้ำหนักตามฟังก์ชันเคอร์เนลแบบ Epanechnikov $K_h = \frac{3}{4}(\frac{1-u^2}{h})$ โดยที่ $u = \frac{x - x_0}{h}$, x_0 คือจุดกำหนดและ h คือแบนวิดจ์ของฟังก์ชันเคอร์เนล ตัวแปร h สามารถควบคุมปริมาณของข้อมูลใกล้เคียงที่ได้รับอนุญาตให้มีอิทธิพลต่อค่าโดยประมาณของ m รวมทั้งอนุพันธ์ ณ จุดนั้น อย่างได้ผล ยังคงมีเทคนิคและแนวทางค้นหาอีกหลายวิธีที่จะเลือกตัวแปร h โดยอาจปรับขนาดของแบนวิดจ์ให้หลากหลายแบนวิดจ์เล็กลงจะทำให้ได้ลักษณะเฉพาะที่ของข้อมูลอีกมากมาย และในกรณีที่แบนวิดจ์ใหญ่ขึ้นก็จะทำให้ข้อมูลถูกปรับเรียบเกินไป ด้วยการแก้โจทย์การถดถอยกำลังสองน้อยที่สุดของฟังก์ชันเคอร์เนล เราก็จะได้ตัวประมาณค่าของ m และอนุพันธ์ 2 ตัวแรก ณ จุดกำหนดแต่ละจุด แล้วก็จะมีข้อมูลทั้งหมดที่จะปรับโดยใช้ (Spline) ได้พอดี

ตอนที่ 2 การประมาณค่าของฟังก์ชัน Reproducing Kernel Hilbert Space

1. การประมาณค่าของฟังก์ชัน Reproducing Kernel Hilbert Space

การวิเคราะห์ Reproducing Kernel Hilbert Space ช่วยให้สามารถสร้างแบบจำลองทางสถิติที่น่าสนใจ และช่วยในการวิเคราะห์ผลจากค่าที่วัดได้ของข้อมูลที่หลากหลายและการกระจัดกระจายและข้อมูลอื่น ๆ ไปพร้อมกันได้ โดยแสดงแนวคิด ความคิดที่เป็นนามธรรมต่าง ๆ พร้อมกับการวิเคราะห์ข้อมูลที่มีความเจาะจงหลายแบบ Reproducing Kernel Hilbert Space มีกรอบแนวคิดสำหรับการประมาณค่าฟังก์ชันที่มีความยืดหยุ่นและสร้างตัวแบบทางสถิติที่มีข้อมูลทั้งโดยตรงและ

โดยอ้อมที่กระจัดกระจาย และสับสนในขอบข่ายที่มีลักษณะทั่วไปค่อนข้างมากในแบบจำลองต่าง ๆ ตามแนววิเคราะห์ฟังก์ชันแบบ RKHS เป็นพื้นฐานสำหรับการประเมินโอกาสที่การประมาณค่าความคลาดเคลื่อนและวิธีการต่าง ๆ ในการปรับเข้าสู่ระบบและสามารถจัดการกับปัญหาและการกระจายต่าง ๆ ของข้อมูลที่หลากหลายเช่น ทฤษฎีของเกาส์ใช้สกุลเลขยกกำลังทั่วไป (General Exponential Families) การประมาณค่าที่มีประสิทธิภาพ (Robust Estimation) การสังเกตแบบช่วง (Interval Observations) ในแบบจำลองสามารถมีข้อจำกัดต่าง ๆ ได้ เช่น ความเป็นบวก (Positivity), การนูนออก (Convexity) และข้อจำกัดความไม่เสมอภาคเชิงเส้นอื่น ๆ (other inequality constraint) Reproducing Kernel Hilbert Space สามารถจัดการกับผลจากการสังเกตที่สับสนที่เกี่ยวกับอนุพันธ์ (Derivatives) ปริพันธ์ (Integrals) และฟังก์ชันเชิงเส้นอื่น ๆ มีกรอบแนวคิดสำหรับการรวมข้อมูลประเภทต่าง ๆ เช่น ค่าจากการสังเกตเฉลี่ย ณ บริเวณช่วงเวลาที่ผิดปกติและไม่สอดคล้องกันสามารถประมาณค่าปริพันธ์และอนุพันธ์ และค่าฟังก์ชันได้ และยังสามารถประมาณค่าการคาดหมายหรือส่วนประกอบ ของแบบจำลองได้อีกด้วย และมีวิธีการปรับตัวแบบจำลองเพื่อให้ได้ประโยชน์สูงสุดจากการอคติความแปรปรวนมีการตีความทั้งสองแบบ (Dual Interpretation) คือ ในรูปของ ค่าประมาณการตามทฤษฎีของ Bayes ก่อนอคติ-ความแปรปรวน (Bias-variance) หรือการปรับการสรุปโดยทั่วไปความคลาดเคลื่อน (Generalization-error) มีช่วงความเชื่อมั่นตามทฤษฎีของ Bayes ที่มีคุณสมบัติความถี่ (Frequents Properties) และสามารถรวมระบบพลวัต (Dynamical System) สมการต่าง ๆ และแบบจำลองทางกายภาพอื่น ๆ ให้อยู่ในแบบจำลองเชิงประจักษ์ (Empirical Model) ได้

2. ฟังก์ชันรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (Reproducing Kernel Hilbert Space)

ตามแนวคิดวิธีการของ Wabba (1990) นั้น RKHS ของ H (ที่มีค่าจริง) คือ Hilbert Space ของฟังก์ชันที่มีค่าจริง ที่กำหนดไว้บน อันตรภาค [0, 1] ที่มีตัวประเมินแบบจุด $L_t : H \to R_t$ $L_t(\mathrm{f}) = \mathrm{f(t)}$ L_t(f) = f(t) ที่มีลักษณะต่อเนื่อง ตามทฤษฎีการแทนค่าของ Riesz การกำหนดนิยามเช่นนี้หมายถึงการมีอยู่ของฟังก์ชัน 2 ตัวแปร $K(\mathrm{s},\mathrm{t})$ ดังเช่น

$$K(\mathrm{s},\bullet) \in H, \text{ for all } s \in [0,1]$$

(Reproducing Property) for every $f \in H$ and $t \in [0,1], hk(t,\bullet), f(H) = f(t)$

การกำหนดนิยาม RKHS อาจเริ่มจากฟังก์ชัน 2 ตัวแปรที่มีค่าบวกชัดเจน $K(\mathrm{s},\mathrm{t})$ และการสร้าง RKHS ให้สมบูรณ์เป็นขอบเขตเชิงเส้นของ $\{K(\mathrm{s},.), \mathrm{s} \in [0,1]\}$ ที่มีผลผลิตภายในกำหนดโดย $\{K(\mathrm{s},.), \mathrm{K}(\mathrm{t},.)\}_H = K(\mathrm{s},\mathrm{t})$ ในตัวแบบการถดถอยที่มีการตอบสนองฟังก์ชันนั้น ค่าของ F ที่มีค่าต่าง ๆ ใน Hilbert Space ของ H ดังนั้น RKHS ที่สร้างจะเป็น subset ของ $\{\mathrm{F} : \mathrm{H} \to \mathrm{H}\}$ การกำหนดนิยาม RKHS ในกรณีนี้จึงทำตามวิธีการของ Wabba (1990)

นิยามที่ 1 ฟังก์ชัน RKHS เมื่อ H เป็น Subset ของ $\{F \to F(x)\}$: โดยเป็น Hilbert Space ซึ่งมี Inner Product $\langle .,. \rangle H$, และมีตัวประเมินแบบจุดหรือคะแนนที่เป็นตัวกระทำเชิงเส้นแบบปิดนั่น คือ $\mathrm{L_x : F \to F(x)}$ เป็นตัวกระทำแบบปิด ตั้งแต่ H ถึง H สำหรับ $x \in H$ ใด ๆ

นิยามข้างต้นนี้ไม่มีประโยชน์กับการสร้าง RKHS นัก เนื่องจากจะนิยามความหมายของ Kernel ที่เกี่ยวข้องกับ Space ส่วนความต่อเนื่องของ $F \to F(x)$ สำหรับ $x \in H$ ใด ๆ จะเทียบเท่ากับความต่อเนื่องของ $F \to F(x)$

นิยามข้างต้นนี้ไม่มีประโยชน์สำหรับการสร้าง RKHS จุดมุ่งหมายที่ต้องการที่จะนิยามเคอร์เนลที่เกี่ยวข้องกับสเปซให้เด่นชัด ความต่อเนื่องของ $F \to F(x)$ สำหรับ $x \in H$ ใด ๆ จะเทียบเท่ากับความต่อเนื่องของการทำ Mapping $F \to F(x)$ สำหรับ $x \in H$ ใด ๆ และ $g \in H$ ด้วยทฤษฎีการแทนที่ที่ประยุกต์ใช้กับ Hilbert space H มี element $k_{g/x}$ อยู่ใน H ดังเช่นในสมการ

$$\left\langle K_x^g, F \right\rangle_H = \left\langle F(x), g \right\rangle_H$$

จากสมการข้างต้นนี้การทำ Mapping $g \to \dfrac{g}{x}$ เป็นเส้นตรงด้วยการจำกัดขอบเขต (Boundedness) ของตัวประเมินแบบจุด (Point Evaluation Operator) เราก็จะได้

$$\left\langle K_x^g, F \right\rangle_H = \left\langle F(x), g \right\rangle_H \le C \|F\|_H \|g\|_H$$

หมายความว่า การทำ Mapping $g \to k_x^g$ ก็ถูกจำกัดขอบเขตเช่นกัน และดังนั้น $g \to K_x^g(\mathrm{y})$ เป็นตัวประเมินเชิงเส้นที่ถูกจำกัดขอบเขตสำหรับ $x \in H$ ใด ๆ ซึ่งเราอาจแสดงให้เห็นโดย $K(\mathrm{x,y})$ นั่นคือ $K(\mathrm{x,y})\mathrm{g} := \mathrm{K}_x^g(\mathrm{y})$ และเรียก $K(.,.)$ ว่าเป็น Reproducing Kernel ที่เกี่ยวข้องกับ H ให้สังเกตว่าในกรณีนี้ คุณสมบัติ reproducing ได้รับการนิยามโดย $\left\langle K(x,.)\mathrm{g,F} \right\rangle_H = \left\langle F(\mathrm{x},), \mathrm{g} \right\rangle_H$ สำหรับ $x \in H, g \in H$ ใด ๆ ซึ่งก็แค่เพียงเขียนสมการ $\left\langle K_x^g, F \right\rangle_H = \left\langle F(x), g \right\rangle_H$ ที่เกิดขึ้นใหม่

คุณสมบัติ 2 ประการของ reproducing kernel ค่อนข้างจะใกล้เคียงกัน $\mathrm{(a)\,K(x,y)} = K\left(y,x\right)^T$ โดยยกกำลัง T หมายถึงตัวประเมินผูกพัน (Adjoint Operator) นี่ก็คือผลสืบเนื่องของเอกลักษณ์ในลำดับต่อไปนี้ (following sequence of identities) ที่ใช้ประโยชน์สมการ $\left\langle K_x^g, F \right\rangle_H = \left\langle F(x), g \right\rangle_H$

$$\left\langle K(\mathrm{x,y})\mathrm{g,f} \right\rangle_H = \left\langle K_x^g, K_y^f \right\rangle_H = \left\langle K_x^g(x), g \right\rangle_H = \left\langle g.K(\mathrm{y,x,f}) \right\rangle_H .\mathrm{(b)} \sum_{i=1}^{n} \sum_{i=j}^{n} \left\langle K(\mathrm{x}_i, \mathrm{x}_j)\mathrm{f}_i \mathrm{f}_j \right\rangle_H \ge 0$$

ซึ่งตามแบบ $\sum_{i,j} \left\langle K(\mathrm{x}_i, \mathrm{x}_j)\mathrm{f}_i, \mathrm{f}_j \right\rangle_H = \sum_{i,j} \left\langle K_x^f, K_{x_j}^{f_j} \right\rangle_H = \left\| \sum_i K_{x_i}^{f_i} \right\|_H$

จากการกล่าวถึงแล้วข้างต้น ได้นิยามของเคอร์เนลที่แน่นอนเป็นค่าบวก (Positive Definte Kernel) ดังต่อไปนี้

นิยามที่ 2 ฟังก์ชันเคอร์เนลที่แน่นอนค่าไม่เป็นลบ เป็นการทำ Mapping แบบ 2 ตัวแปร กับ $H \times H$ ที่ใช้ค่าใน $L(\mathrm{H})$ สเปซของตัวประมาณเชิงเส้นที่ถูกจำกัดขอบเขตจาก H ถึงตัวเอง ดังเช่นสมการ

$$K(\mathrm{x},\mathrm{y}) = \mathrm{K}(\mathrm{y},\mathrm{x})^T$$

และ

$$\sum_{i=1}^{n}\sum \left\langle K(\mathrm{x}_i,\mathrm{x}_j)\mathrm{f}_i,\mathrm{f}_j \right\rangle_H \geq 0$$

โดย $\{X_i\}$ และ $\{f_i\}$ คือ 2 Sequence ใด ๆ ใน H ถ้าผลรวม 2 ผลรวมข้างต้นมีค่าเป็นบวกโดยเคร่งครัดแล้ว โดยที่เมื่อใดก็ตามที่ $\{X_i\}$ ชัดเจนและ $\{f_i\}$ มิได้มีค่าเป็นศูนย์ทั้งหมด โดย K เคอร์เนลที่เป็นค่าบวกที่แน่นอน

3. ตัวแบบของข้อมูลฟังก์ชันเคอร์เนลฮิลเบอร์ทสเปซ

พิจารณาปัญหาการอนุมานและพยากรณ์สำหรับตัวแบบในสมการ $y_i = F(x_i) + \varepsilon_i$ หากพิจารณาฟังก์ชันตัวแปรร่วมและฟังก์ชันการตอบสนอง $\{(x_1, y_1),...,(x_n, y_n)\}$ ที่สังเกตได้แล้ว แนวทางทั่วไปที่จะประมาณค่า F ก็คือ จะต้องแก้ปัญหา Minimization ต่อไปนี้ :

$$\min_{F\varepsilon H}\sum_{i=1}^{n}\left\|y_i - F(x_i)\right\|_2^2 + \lambda\|F\|_H$$

โดยที่เงื่อนไข (Penalty Term) ในเกณฑ์ปกติ H ที่มีค่าแลมด้า มากกว่าศูนย์ ถูกเพิ่มเข้าในความเสี่ยงเชิงประจักษ์ (Empirical Risk) ในฐานที่เสร็จแล้วในฟังก์ชันการถดถอยเสมือนพหุนามเพื่อการปรับเรียบ (Smoothing Spline Regression) เราใช้ฟังก์ชันความสูญเสีย (Loss Function) แบบง่ายที่สุด $\|x\|_2^2 = \int_0^1 x^2(\mathrm{t})\,\mathrm{dt}$ ในที่นี้ แม้ว่าความสูญเสียแบบอื่น ก็สามารถนำมาพิจารณาได้ค่อนข้างแน่นอน โจทย์การ Optimization ข้างต้นได้ Optimize ทำให้เหมาะสมที่สุดกับสเปซที่มีมิติเป็นอนันต์ (Infinite-Dimensional Space) H ซึ่งโดยทั่วไปแล้วจะกระทำไม่ได้ แต่ในทางทฤษฎีแล้วตัวแทน (Representer) ข้างล่างได้ลดโจทย์นี้ให้อยู่ในมิติที่ค่าแน่นอน (หากพิจารณาให้ H เป็น Vector Space ที่มี Elements ใน H ที่ทำหน้าที่เป็น “ตัวคงที่”)

ทฤษฎีบทที่ 1 หากจะพิจารณาค่าที่ได้จากการสังเกต $\{(x_i, y_i)\}_{i-1}^n$ แล้ว วิธีแก้สมการ $\min_{F\in H}\sum_{i=1}^{n}\|y_i - F(\mathrm{x}_i)\|_2^2 + \lambda\|F\|_H$ จะมีการแทนค่าดังต่อไปนี้

$$\hat{F} = \sum_{i=1}^{n} K(\mathrm{x}_i,.)\alpha_i$$

ที่มีสัมประสิทธิ์ฟังก์ชันเป็น $\alpha_i \in H$

เมื่อถึงขั้นตอนนี้จะประสบกับปัญหายุ่งยาก 2 ประการ ประการแรก การสร้าง K โดยทั่วไปค่อนข้างยาก และการค้นคว้าจากงานวิจัยหรือบทความที่เกี่ยวข้องก็ไม่สามารถช่วยอะไรได้เกี่ยวกับวิธีที่จะสร้างเคอร์เนลที่แน่นอนค่าบวกโดยทั่วไป ประการที่สองแม้ว่าจะสามารถสร้างเคอร์เนลได้ แต่วิธีที่จะแก้สมการ $\min_{F \in H} \sum_{i=1}^{n} \|y_i - F(\mathrm{x}_i)\|_2^2 + \lambda \|F\|_H$ ก็ไม่ชัดเจนหลังจากที่แทนค่าในสมการ $\hat{F} = \sum_{i=1}^{n} K(\mathrm{x}_i, .)\alpha_i$ แล้วก็ตาม จากปัญหายุ่งยาก 2 ประการดังกล่าวจึงต้องถูกบังคับให้เลือกใช้ฟังก์ชันเคอร์เนลแบบง่ายที่สุด คือ $K(\mathrm{x}, \mathrm{y}) = \mathrm{a}(\mathrm{d}(\mathrm{x}, \mathrm{y}))\mathrm{I}$ เมื่อ $a(.)$ เป็นฟังก์ชันที่เป็นค่าบวกแน่นอนจำนวนจริง $I \in L(H)$ คือ ตัวกระทำเอกลักษณ์ และ d คือค่าเมตริกซ์ที่ H เลือกที่จะใช้ค่าเมตริกซ์ที่ง่ายที่สุด $d(x, y) = \|x - y\|_2$ ซึ่งพบว่า มีใช้เช่นกันในผลงานของ (Preda, 2007, pp. 829-840) แต่ไม่สามารถที่จะยกตัวอย่างของเคอร์เนลที่ซับซ้อนกว่านี้อีกได้ แต่ค่าประมาณค่านี้ก็ทำงานได้ค่อนข้างดีในการทดลองทั้งหมดและเป็นที่ชัดเจนว่า K ที่ได้รับการนิยามแบบนี้ คือ ค่าที่แน่นอนและไม่เป็น ค่าลบ แต่การพิสูจน์ว่าค่าที่แน่นอนและเป็นค่าบวกจำเป็นต้องมีการทำงานศึกษาเพิ่มเติม

ทฤษฎีบทที่ 2 ฟังก์ชันเคอร์เนล $K(\mathrm{x}, \mathrm{y}) = \mathrm{a}(\mathrm{d}(\mathrm{x}, \mathrm{y}))\mathrm{I}$ คือ ค่าที่แน่นอนและเป็นค่าบวก ถ้าหาก $a\left(\|x_i - x_j\|_2\right)$ คือ ฟังก์ชันที่แน่นอนเป็นจำนวนจริงบวก หลังจากการปรับให้ง่ายขึ้นได้สามารถแก้โจทย์สมการ $\min_{F \in H} \sum_{i=1}^{n} \|y_i - F(\mathrm{x}_i)\|_2^2 + \lambda \|F\|_H$ โดยให้ $a_{ij} = a\left(\|x_i - x_j\|_2\right)$ และใช้ประโยชน์จากทฤษฎีบท จะได้โจทย์การ Optimization ดังต่อไปนี้

$$\min_{\alpha} \sum_{i=1}^{n} \left\| y_i - \sum_{j} a_{ij} \right\|_2^2 + \lambda \sum_{i,j} a_{i,j}(\alpha_i, \alpha_j)$$

มีวิธีที่จะดำเนินการต่อมากกว่า 1 วิธี ตัวอย่างเช่น เราสามารถที่จะแทนค่า α_i แต่ละค่าด้วยการขยายด้วยหลักการพื้นฐานที่ได้เลือกแทนที่จะคำนวณสมการ $\min_{\alpha} \sum_{i=1}^{n} \left\| y_i - \sum_{j} a_{ij} \right\|_2^2 + \lambda \sum_{i,j} a_{i,j}(\alpha_i, \alpha_j)$ นี้อีก ภายใต้กรอบแนวคิดแบบ RKHS

โดยสมมติว่า H คือ RKHS เป็นจำนวนจริงในตัวมันเองที่มี Reproducing Kernel k. ในตอนแรกเงื่อนไขของพจน์ที่หาย ได้ถูกทำให้มีค่าวิกฤตไม่ต่อเนื่อง (Discretized) โดยถือว่าค่าที่สังเกตได้จากกริดสม่ำเสมอปกติ (Regular Grid) $\{t_1, ..., t_T\}$ บนข้อมูล $[0,1]$ แล้วการประยุกต์ใช้ทฤษฎีบทตัวแทน อีกแบบหนึ่งในกรณีของค่าจำนวนจริง (Real-Value Case)

ดังที่ระบุในทฤษฎีบทที่ 3 ข้างล่าง ก็จะแสดงให้เห็นวิธีการที่เป็น

□□□□□ $$\alpha_i = \sum_{l=1}^{T} b_j^i k(t_l)$$

และการแค่คำนวณสัมประสิทธิ์ $\{b_l^i\}$ แล้วจะได้ทฤษฎีบทที่ 3 อย่างเป็นทางการ

ทฤษฎีบทที่ 3 จงพิจารณาสมการที่ 3 อย่างที่เป็นข้อมูลดิบ

$$\min_{\alpha_i} \sum_{i=1}^{n} \sum_{t=1}^{T} \left[y_i(\mathrm{t}_i) - \sum_j a_{ij}(\mathrm{t}_i) \right]^2 + \lambda \sum_{i,j} a_{i,j} \left\langle \alpha_i, \alpha_j \right\rangle H$$

วิธีการแก้สมการที่ตามทฤษฎีบทที่ 3 จะอยู่ในรูปของสมการ $\alpha_i = \sum_{i=1}^{T} b_j^i K(\mathrm{t}_i)$

สำหรับสูตรอย่างละเอียดเกี่ยวกับการคำนวณวิธีแก้สมการตามทฤษฎีบทที่ 3 ตัวแบบสมการตามทฤษฎีบทที่ 3 ซึ่งจะคล้ายคลึงกับการประมาณค่าเสมือนพหุนามเพื่อการปรับเรียบแบบดั้งเดิม (Classical Smoothing Spline Estimation) ที่มีความแตกต่างอยู่ 2 ประการ ประการแรกในกรณีแรกของสมการตามทฤษฎีบทที่ 3 แทนที่จะพยายามปรับค่า y ให้เรียบด้วยฟังก์ชันเพียงฟังก์ชันเดียว แต่กลับพยายามประมาณการค่า y ที่ได้จากการสังเกตแต่ละค่า ในฐานที่เป็นผลรวมของเซตของฟังก์ชันร่วม $\{a_j\}_{j=1}^{n}$ สัมประสิทธิ์ของ a_{ij} สะท้อนให้เห็นถึงความคล้ายคลึงของตัวแปรร่วม x_i และ x_j ประการที่สอง ปรากฏว่ามีเงื่อนไขความคลาดเคลื่อน $\sum_{i,j} a_{ij} \left\langle \alpha_i, \alpha_j \right\rangle_H$ ซึ่งเกี่ยวข้องกับเงื่อนไขทั่วไปทั้งหมดของ $\left\langle \alpha_i, \alpha_j \right\rangle_H$ โดยมี $i \neq j$ ส่วนเงื่อนไขของพจน์ที่หายในสมการตามทฤษฎีบทที่ 3 นั้น ค่อนข้างจะเป็นธรรมชาติอยู่แล้ว พอจะมีความเข้าใจได้โดยไม่ต้องไปแก้ปัญหาทั้งหมดโดยใช้ฟังก์ชัน RKHS และอนุพันธ์ข้างต้น ถ้ามีปัญหาก็อาจใช้เงื่อนไขความคลาดเคลื่อน เช่น $\sum_{i,j} a_{ij} \left\langle \alpha_i, \alpha_j \right\rangle_H$ เพื่อปรับให้เป็นปกติได้ จะเรียกวิธีการแก้โจทย์สมการตามทฤษฎีบทที่ 3 โดยค่าประมาณ RKHS ที่ปรับแล้วที่มีเงื่อนไขของความคลาดเคลื่อนที่ง่ายขึ้นนี้ว่าเป็นวิธีที่ตรงกันข้ามกับวิธีแก้โจทย์สมการตามทฤษฎีบทที่ 3 ดั้งเดิมด้วยค่าประมาณ RKHS

การดำเนินการนั้น จะใช้เคอร์เนลแบบเกาส์กับ RKHS ทั้ง 2 แบบ ซึ่งก็คือ $K(\mathrm{x},\mathrm{y}) = \exp\left\{ -\frac{\|x-y\|_2^2}{2\sigma^2} \right\} I$ และ $\mathrm{k}(\mathrm{s},\mathrm{t}) = \exp\left\{ -\frac{\|x-y\|_2^2}{2\sigma^2} \right\}$ นั่นเอง ดังนั้นจึงมีค่าเป็นค่าเฉลี่ยของ σ, σ', λ อย่างน้อย 3 ค่าที่จะต้องระบุรายละเอียดเฉพาะ สำหรับค่าความกว้างของพารามิเตอร์ในเคอร์เนล เลือก σ ให้เป็นค่าเฉลี่ยของ $\|x_i - x_j\|_2, i, j = 1, 2, \ldots, n$ ทั้งหมด และเลือก σ' เป็นค่าเฉลี่ยของ $\|t_i - t_j\|_2, i, j = 1, 2, \ldots, \mathrm{T}$ การเลือกเช่นนี้มิใช่การเลือกที่เหมาะสมที่สุด แต่ที่พบจะให้ผลออกมาดีพอสมควร ดังนั้นจึงไม่สนใจนักที่จะค้นหาสเปซของค่าพารามิเตอร์เหล่านี้ ส่วนการเลือกค่าพารามิเตอร์ปรับเรียบนั้นยิ่งมีความสำคัญขึ้นไปอีก ซึ่งอาจใช้ GCV หรือ วิธีการตรวจสอบความถูกต้องแบบครอบคลุมทั่วไป (Generalized Cross-Validation) ที่ใช้มากในผลงานของ

Wabba (1990) สำหรับการเลือกค่า λ เมื่อมีการกำหนดกริดของค่าต่าง ๆ (Grid of values) สำหรับ λ แล้ว GCV ก็จะประมาณค่าความคลาดเคลื่อนได้โดย

$$V(\lambda) = \frac{1}{n} \frac{\| (I - A(\lambda))\, \mathrm{y} \|^2}{\left[\frac{1}{n} T_r (\mathrm{I} - \mathrm{A}(\lambda)) \right]^2}$$

โดย Y คือ $n \times t$ matrix ของค่าดิบจากค่าการสังเกต $\{y_i(t_l)\}$ และ $A(\lambda)$ คือ "influence matrix" และค่า λ ค่าสุดท้ายก็ถูกเลือกให้เป็นค่าที่ลดทอน $V(\lambda)$ ให้เหลือน้อยที่สุด ในผลงานของ Ferraty and Vieu (2003), Ferraty and Vieu (2004), and Ferraty et al. (2007) ได้มีการศึกษาค่าประมาณเคอร์เนล แบบนอนพาราเมตริก ซึ่งนิยามง่าย ๆ เป็น

$$F(\mathrm{x}) = \frac{\sum_i k(\| x_i - x \|) y_i}{\sum_i k(\| x_i - x \|)}$$

ผลงานเหล่านั้น ได้มีการศึกษาแต่เฉพาะค่าประมาณแบบเคอร์เนลสำหรับตัวแบบที่มีการตอบสนองแบบสเกลาร์เท่านั้น แต่ก็สามารถที่จะใช้ค่าประมาณค่านี้ได้อย่างชัดเจนเมื่อตัวแปรตามเป็นเส้นโค้ง นอกจากนี้ยังค่อนข้างจะเป็นธรรมชาติที่น่าจะปรับการตอบสนอง y_i ให้เรียบก่อนที่จะใส่เข้าไปในสมการ $F(\mathrm{x}) = \frac{\sum_i k(\| x_i - x \|) y_i}{\sum_i k(\| x_i - x \|)}$ หากค่าการสังเกตมีการรบกวนค่อนข้างมาก

ตอนที่ 3 การพยากรณ์ด้วยอนุกรมเวลาและการสร้างตัวแบบ ARIMA

การพยากรณ์ด้วยอนุกรมเวลาอยู่ภายใต้ข้อสมมติว่า รูปแบบของข้อมูลในอดีตยังคงเกิดขึ้นต่อไปในอนาคตหรืออาจกล่าวได้ว่า ลักษณะของการเปลี่ยนแปลงรูปแบบของข้อมูลหรือตัวแปรที่สนใจในอนาคตลักษณะก็ควรจะเป็นเช่นนั้นต่อไป ข้อมูลอนุกรมเวลาประกอบด้วยองค์ประกอบพื้นฐานที่สำคัญ 4 ปัจจัย ได้แก่ อิทธิพลของแนวโน้ม อิทธิพลของฤดูกาล อิทธิพลของวัฏจักร และอิทธิพลของเหตุการณ์ที่ผิดปกติ การพยากรณ์ด้วยอนุกรมเวลามีหลายวิธี วิธีที่นิยมใช้ได้แก่ 1) การพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 2) การพยากรณ์ด้วยตัวแบบ ARIMA 3) การพยากรณ์ด้วยตัวแบบ ARIMAX สามารถสรุปได้ดังนี้

1. การพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล

การพยากรณ์โดยวิธีปรับให้เรียบแบบเอกซ์โพเนนเชียล (Exponential Smoothing) เป็นเทคนิคการพยากรณ์ด้วยอนุกรมเวลา ซึ่งใช้พยากรณ์ข้อมูลระยะสั้นถึงระยะปานกลาง เทคนิคการพยากรณ์ด้วยวิธีนี้ เป็นการกำจัดอิทธิพลของความไม่แน่นอนออกไป ทำให้ค่าพยากรณ์ในอนาคตที่ได้มีความแม่นยำมากยิ่งขึ้น วิธีการพยากรณ์โดยวิธีปรับให้เรียบ แบบเอกซ์โพเนนเชียลเป็นเทคนิค

การพยากรณ์ด้วยอนุกรมเวลาที่พัฒนาขึ้น เพื่อแก้จุดด้อยของเทคนิคการพยากรณ์ด้วยวิธีค่าเฉลี่ยเคลื่อนที่เนื่องจากเทคนิคการพยากรณ์ด้วยวิธีค่าเฉลี่ยเคลื่อนที่ให้ความสำคัญของข้อมูลทุกค่าเท่ากันหมด แต่การพยากรณ์โดยวิธีปรับให้เรียบแบบเอกซ์โพเนนเชียล ให้ความสำคัญของข้อมูลหรือน้ำหนักไม่เท่ากัน (กัลยา วานิชย์บัญชา, 2554, หน้า 368-393) การพยากรณ์โดยวิธีปรับให้เรียบแบบเอกซ์โพเนนเชียล มีหลายวิธีในที่นี้จะกล่าวถึงวิธี ดังต่อไปนี้

1.1 การปรับให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่าย

การปรับให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่าย (Single Exponential Smoothing) เป็นเทคนิคการพยากรณ์ด้วยอนุกรมเวลาที่ให้ความสำคัญของข้อมูลไม่เท่ากัน วิธีนี้เหมาะสำหรับข้อมูลที่มีการเปลี่ยนแปลง หรือเคลื่อนไหวแบบไม่มีแนวโน้ม ไม่มีอิทธิพลของฤดูกาล มีเฉพาะความไม่แน่นอนเพียงอย่างเดียว วิธีนี้ให้ความสำคัญกับข้อมูลปัจจุบันมากที่สุด และความสำคัญของข้อมูลที่อยู่ในอดีตจะลดความสำคัญลงเรื่อย ๆ ขั้นตอนในการพยากรณ์ด้วยวิธีปรับให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่าย มีดังนี้

กำหนดน้ำหนัก (α) โดยที่ $0 \le \alpha \le 1$ ให้แก่ข้อมูลล่าสุด (Y_t) รองลงมาให้น้ำหนัก $\alpha(1-\alpha)$ ให้กับข้อมูล Y_{t-1} และให้น้ำหนัก $\alpha(1-\alpha)^2$ ให้กับข้อมูล Y_{t-2} ไปเรื่อย ๆ โดยที่ค่าพยากรณ์ของ Y ณ เวลา $t+1$ คือ F_{t+1} สามารถเขียนเป็นสมการได้ดังนี้

$$F_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \dots \quad (1)$$

หรือ

$$F_t = \alpha Y_{t-1} + \alpha(1-\alpha)Y_{t-2} + \alpha(1-\alpha)^2 Y_{t-3} + \dots \quad (2)$$

จากสมการที่ (1) จะได้

$$F_{t+1} = \alpha Y_t + (1-\alpha)[\alpha Y_{t-1} + \alpha(1-\alpha)Y_{t-2} + \alpha(1-\alpha)^2 Y_{t-3} + \dots] \quad (3)$$

แทนค่า

$F_t = \alpha Y_{t-1} + \alpha(1-\alpha)Y_{t-2} + \alpha(1-\alpha)^2 Y_{t-3} + \dots$ ในสมการที่ (3) จะได้

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

โดยกำหนดให้ $F_1 = Y_1$

ในการเลือกค่า α ให้เลือกค่า α ที่ใกล้ 1 มากที่สุดเพื่อให้ความสำคัญกับข้อมูลในปัจจุบันมากที่สุด สำหรับเกณฑ์ในการเลือกค่าน้ำหนัก α ที่เหมาะสมจะเลือกค่าที่ให้ SSE หรือ MSE ที่ต่ำที่สุด

1.2 การปรับให้เรียบแบบเอกซ์โพเนนเชียล 2 ครั้ง

วิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 2 ครั้ง (Double Exponential Smoothing) มีชื่อเรียกอีกอย่างหนึ่ง คือ Holt Linear Method ผู้ที่พัฒนาวิธีนี้ คือ Holt (1957)

วิธีนี้เหมาะกับใช้ในการพยากรณ์ข้อมูลที่มีแนวโน้มในรูปแบบเชิงเส้น (Linear) รวมอยู่ด้วย วิธีนี้เหมาะในการพยากรณ์ระยะสั้นถึงระยะปานกลาง วิธีของ Holt ยังใช้หลักการของการปรับให้เรียบแบบเอกซ์โพเนนเชียลอย่างง่าย คือ ให้ความสำคัญของข้อมูล แต่ละตัวไม่เท่ากัน ซึ่งวิธีนี้เหมาะกับข้อมูลที่มีเฉพาะความไม่แน่นอน แต่วิธีของ Holt จะใช้กับข้อมูลที่มีแนวโน้มและความไม่แน่นอนจึงมีค่าคงที่ในการปรับให้เรียบ 2 ค่า คือ α และ γ โดยที่

α คือ ค่าคงที่ในการปรับให้เรียบระหว่างข้อมูลกับค่าพยากรณ์โดยที่ $0 \le \alpha \le 1$ ถ้าให้ α มีค่าใกล้ 1 แสดงว่าให้ความสำคัญกับข้อมูลตัวนั้นมากกว่าค่าอื่น จะให้ความสำคัญกับข้อมูลปัจจุบันใกล้ 1 มากที่สุด

γ คือ ค่าคงที่ในการปรับให้เรียบระหว่างแนวโน้มจริงกับค่าประมาณของแนวโน้มโดยที่ $0 \le \gamma \le 1$ ถ้าให้ γ มีค่าใกล้ 1 แสดงว่าให้ความสำคัญกับข้อมูลตัวนั้นมากกว่าค่าอื่น จะให้ความ สำคัญกับข้อมูลปัจจุบันใกล้ 1 มากที่สุด สมการที่ใช้ในการพยากรณ์คือ

$$F_{t+m} = L_t + b_t m$$

b_t คือ ความชัน (Slope) ของข้อมูล ณ เวลา t

m คือ จำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า เช่น $m = 7$

หมายถึง การพยากรณ์ข้อมูลที่จะเกิดขึ้นอีก 7 เดือน ข้างหน้า โดยคำนวณ

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \gamma(L_t - L_{t-1}) + (1-\gamma)b_{t-1}$$

ขั้นตอนในการพยากรณ์ด้วยวิธีปรับให้เรียบแบบเอกซ์โพเนนเชียล 2 ครั้ง หรือวิธีของ Holt มีดังนี้

1. กำหนดค่าคงที่ในการปรับให้เรียบ 2 ค่า คือ α และ γ โดยที่ $0 \le \alpha \le 1$ และ $0 \le \gamma \le 1$ โดยการเลือก α และ γ ที่ทำให้ค่า SSE หรือ MSE ต่ำที่สุด

2. กำหนดค่าเริ่มต้น $L_1 = Y_1$, $b_1 = Y_2 - Y_1$

3. คำนวณค่า

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \gamma(L_t - L_{t-1}) + (1-\gamma)b_{t-1}$$

4. นำค่า L_t และ b_t มาหาค่าพยากรณ์โดยคำนวณจาก

$$F_{t+m} = L_t + b_t m$$

ในกรณีที่ $\alpha = \gamma$ จะเรียกวิธีนี้ว่า Brown’s Double Exponential Smoothing

1.3 การปรับให้เรียบแบบเอกซ์โพเนนเชียล 3 ครั้ง

วิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 3 ครั้ง (Triple Exponential Smoothing) ผู้ที่พัฒนาขึ้นมา คือ Winter (1960) หรือเรียกว่าวิธีของ Winter เหมาะสำหรับข้อมูลที่มีแนวโน้มและอิทธิพลของฤดูกาล ใช้พยากรณ์ระยะสั้นถึงระยะปานกลาง ข้อมูลไม่ควรเป็นรายปี เพราะจะทำให้ไม่สามารถแยกอิทธิพลของฤดูกาลได้ ข้อมูลควรอยู่ในรูปของ รายเดือน รายสัปดาห์ หรือรายไตรมาส อย่างน้อย 36 รายการขึ้นไป ถ้าเป็นรายเดือนอย่างน้อย 12 รายการขึ้นไป วิธีของ Winter ใช้หลักการของการปรับให้เรียบแบบเอกซ์โพเนนเชียล คือ ให้ความสำคัญแก่ข้อมูลไม่เท่ากัน และมีค่าปรับให้เรียบ 3 ค่า คือ

α คือ ค่าคงที่ในการปรับให้เรียบระหว่างข้อมูลกับค่าพยากรณ์โดยที่ $0 \le \alpha \le 1$ ถ้าให้ α มีค่าใกล้ 1 แสดงว่าให้ความสำคัญกับข้อมูลตัวนั้นมากกว่าค่าอื่น จะให้ความสำคัญกับข้อมูลปัจจุบันใกล้ 1 มากที่สุด

γ คือ ค่าคงที่ในการปรับให้เรียบระหว่างแนวโน้มจริงกับค่าประมาณของแนวโน้มโดยที่ $0 \le \gamma \le 1$ ถ้าให้ γ มีค่าใกล้ 1 แสดงว่าให้ความสำคัญกับข้อมูลตัวนั้นมากกว่าค่าอื่น จะให้ความสำคัญกับข้อมูลปัจจุบันใกล้ 1 มากที่สุด

δ คือ ค่าคงที่ในการปรับให้เรียบระหว่างฤดูกาลจริงกับค่าประมาณของฤดูกาล โดยที่ $0 \le \delta \le 1$ ถ้าให้ δ มีค่าใกล้ 1 แสดงว่าให้ความสำคัญกับข้อมูลตัวนั้นมากกว่าค่าอื่น จะให้ความสำคัญกับข้อมูลปัจจุบันใกล้ 1 มากที่สุด

สำหรับการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 3 ครั้ง (Triple Exponential Smoothing) หรือวิธีของ Winter สามารถพยากรณ์ได้ 2 รูปแบบ คือ

1. ตัวแบบเชิงคูณ (Multiplicative Seasonality Model)
2. ตัวแบบเชิงบวก (Additive Seasonality Model)

1. การพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 3 ครั้ง โดยใช้ตัวแบบเชิงคูณสมการที่ใช้พยากรณ์ คือ

$$F_{t+m} = (L_t + b_t m) S_{t-s+m}$$

เมื่อ s คือ จำนวนฤดูกาลใน 1 ปี ถ้าข้อมูลรายเดือน $s = 12$, ถ้าข้อมูลรายไตรมาส $s = 4$

m คือ ระยะเวลาที่ต้องการพยากรณ์ไปข้างหน้า

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)[L_{t-1} + b_{t-1}]$$
$$b_t = \gamma(L_t - L_{t-1}) + (1-\gamma)b_{t-1}$$
$$S_t = \delta \frac{Y_t}{L_t} + (1-\delta)S_{t-s}$$

การกำหนดค่าเริ่มต้นประกอบด้วยกำหนดค่า L ค่า b และ s โดยคำนวณจาก

$$L_s = \frac{1}{s}(Y_1 + Y_2 + ... + Y_s)$$
$$b_s = \frac{1}{s}(\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + ... + \frac{Y_{2s} - Y_s}{s})$$

และ

$$S_1 = \frac{Y_1}{L_s}, S_2 = \frac{Y_2}{L_s}, ..., S_s = \frac{Y_s}{L_s}$$

2. การพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 3 ครั้งโดยใช้ตัวแบบเชิงบวกสมการที่ใช้ในการพยากรณ์ คือ

$$F_{t+m} = L_t + b_t m + S_{t-s+m}$$

โดยที่

$$L_t = \alpha(Y_t + S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \gamma(L_t - L_{t-1}) + (1-\gamma)b_{t-1}$$
$$S_t = \delta(Y_t - L_t) + (1-\delta)S_{t-s}$$

การกำหนดค่าเริ่มต้น

$$S_1 = Y_1 - L_s, S_2 = Y_2 - L_s, ..., S_s = Y_s - L_s$$
$$L_s = \frac{1}{s}(Y_1 + Y_2 + ... + Y_s)$$
$$b_s = \frac{1}{s}(\frac{Y_{s+1} - Y_1}{s} + \frac{Y_{s+2} - Y_2}{s} + ... + \frac{Y_{2s} - Y_s}{s})$$

ขั้นตอนการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 3 ครั้ง หรือวิธีของ Winter มีดังนี้

1. นำข้อมูลมาพล็อตกราฟว่ามีส่วนประกอบใดบ้าง ในการพยากรณ์ด้วยวิธีการปรับให้เรียบแบบเอกซ์โพเนนเชียล 3 ครั้ง หรือวิธีของ Winter ต้องมีส่วนประกอบของแนวโน้มและฤดูกาล แล้วเลือกตัวแบบเชิงบวกหรือเชิงคูณ ขึ้นอยู่กับผู้วิจัย สำหรับตัวแบบที่นิยมใช้ คือ ตัวแบบเชิงคูณ

2. กำหนดค่าคงที่ในการปรับให้เรียบประกอบด้วยค่า α , γ และ δ ที่ทำให้ค่า SSE หรือ MSE ต่ำที่สุด

3. คำนวณค่า L_t, b_t และ S_t

4. คำนวณค่าพยากรณ์ตามตัวแบบที่เลือกถ้าเป็นตัวแบบเชิงบวกใช้สมการ $F_{t+m} = L_t + b_t m + S_{t-s+m}$

ถ้าเป็นตัวแบบเชิงคูณใช้สมการ $F_{t+m} = (L_t + b_t m)S_{t-s+m}$

2. การพยากรณ์ด้วยตัวแบบ ARIMA

ตัวแบบ ARIMA โดยวิธีการของบ็อกซ์–เจนกินส์ มีชื่อว่า Auto-Regressive Integrated Moving Average เขียนแทนด้วยสัญลักษณ์ ARIMA (p,d,q)(P,D,Q)$_s$ เป็นวิธีการพยากรณ์อนุกรมเวลาซึ่งเลือกตัวแบบที่ใช้ในการพยากรณ์ โดยพิจารณาจากลักษณะของสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และสหสัมพันธ์ในตัวเองส่วนย่อย (Partial Autocorrelation Function: PACF) ของข้อมูลอนุกรมเวลาที่พิจารณา ซึ่งมีคุณสมบัติ Stationary โดยตัวแบบที่เป็นไปได้อาจมีมากกว่าหนึ่งตัวแบบ ซึ่งต้องมีขั้นตอนการตรวจสอบเพื่อเลือกตัวแบบที่เหมาะสมที่สุดเพื่อใช้ในการพยากรณ์ต่อไป

ตัวแบบ ARIMA (p,d,q)(P,D,Q)$_s$ มีสมการทั่วไป ดังนี้

$$(1-\phi_1 B-\phi_2 B^2 \ldots -\phi_p B^p)(1-\Phi_1 B^S-\Phi_2 B^{2S}-\ldots-\Phi_P B^{PS})(1-B)^d(1-B^S)^D Y_t$$
$$= (1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)(1-\Theta_1 B^S-\Theta_2 B^{2S}-\ldots-\Theta_Q B^{QS})\varepsilon_t$$

โดยที่ Y_t คือ ค่าของข้อมูลอนุกรมเวลาที่ต้องการพยากรณ์ ณ เวลาที่ t และ Y_t ต้องมีคุณสมบัติ Stationary

p คือ อันดับที่ p ของขบวนการถดถอยในตัวเองแบบ Nonseasonal (Nonseasonal Auto-Regressive Process of Order p)

P คือ อันดับที่ P ของขบวนการถดถอยในตัวเองแบบ Seasonal (Seasonal Auto-Regressive Process of Order P)

q คือ อันดับที่ q ของขบวนการเฉลี่ยเคลื่อนที่แบบ Nonseasonal (Nonseasonal Moving Average Process of Order q)

Q คือ อันดับที่ Q ของขบวนการเฉลี่ยเคลื่อนที่แบบ Seasonal (seasonal Moving Average Process of Order Q)

d, D คือ จำนวนครั้งของการหาผลต่างแบบ Nonseasonal อันดับที่ d และผลต่างแบบ Seasonal อันดับที่ D ตามลำดับ เพื่อทำให้อนุกรมเวลามีคุณสมบัติ Stationary

S คือ จำนวนคาบเวลาของฤดูกาลใน 1 รอบ

t คือ เวลา

B คือ ตัวดำเนินการย้อนหลังเวลา (Backward Shift Operator)

$\phi_1, \phi_2, ..., \phi_p$ คือ พารามิเตอร์แสดงสัมประสิทธิ์การถดถอยในตัวเองแบบ Nonseasonal อันดับที่ 1, 2,...,*p* ตามลำดับ (Nonseasonal Auto-Regressive Process of Order 1, 2,...,*p*)

$\theta_1, \theta_2, ..., \theta_q$ คือ พารามิเตอร์แสดงสัมประสิทธิ์ของการเฉลี่ยเคลื่อนที่แบบ Nonseasonal อันดับที่ 1, 2,...,q ตามลำดับ (Nonseasonal Moving Average Process of Order 1, 2,...q)

$\Phi_1, \Phi_2, ..., \Phi_P$ คือ พารามิเตอร์แสดงสัมประสิทธิ์การถดถอยในตัวเองแบบ Seasonal อันดับที่ 1, 2,...,P ตามลำดับ (Seasonal Auto-Regressive Process of Order 1, 2,...P)

$\Theta_1, \Theta_2, ..., \Theta_Q$ คือ พารามิเตอร์แสดงสัมประสิทธิ์ของการเฉลี่ยเคลื่อนที่แบบ Seasonal อันดับที่ 1, 2,...Q ตามลำดับ (Seasonal Moving Average Process of Order 1, 2,...Q)

$(1-B)^d Y_t$ คือ ผลต่างอันดับที่ d ของอนุกรมเวลา Y_t

$(1-B^s)^D Y_t$ คือ ผลต่างอันดับที่ D ของอนุกรมเวลา Y_t

ε_t คือ ค่าความคลาดเคลื่อนจากการพยากรณ์ ณ เวลาที่ t

3. การพยากรณ์ด้วยตัวแบบ ARIMAX

ตัวแบบ ARIMAX มีชื่อว่า Auto-Regressive Integrated Moving Average with Exogenous Variables เป็นวิธีการพยากรณ์อนุกรมเวลาที่เป็นส่วนขยายของทฤษฎีของบ็อกซ์–เจนกินส์ ซึ่งจะนำอนุกรมเวลานำเข้า (Input Time Series: X_t) มาพิจารณาร่วมในตัวแบบเพื่อใช้ในการพยากรณ์อนุกรมเวลานำออก (Output Time Series: Y_t) โดยตัวแบบดังกล่าวจะเขียนแทนด้วยสัญลักษณ์ ARIMAX (p,d,q) (P, D, Q)$_S$

ตัวแบบ ARIMAX (p,d,q)(P,D,Q)$_S$ มีสมการทั่วไป ดังนี้

$$(1-B)^d(1-B^S)^D Y_t = \beta_0 + \beta_1[(1-B)^d(1-B^S)^D X_{1,t}] + \ldots + \beta_k[(1-B)^d(1-B^S)^D X_{k,t}]$$
$$+\frac{(1-\theta_1 B-\theta_2 B^2-\ldots-\theta_q B^q)(1-\Theta_1 B^S-\Theta_2 B^{2S}-\ldots-\Theta_Q B^{QS})}{(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)(1-\Phi_1 B^S-\Phi_2 B^{2S}-\ldots-\Phi_P B^{PS})}\varepsilon_t$$

โดยที่ Y_t คือ ค่าของข้อมูลอนุกรมเวลาที่ต้องการพยากรณ์ ณ เวลาที่ t และ Y_t ต้องมีคุณสมบัติ Stationary

p คือ อันดับที่ p ของขบวนการถดถอยในตัวเองแบบ Nonseasonal (Nonseasonal Auto-Regressive Process of Order p)

P คือ อันดับที่ P ของขบวนการถดถอยในตัวเองแบบ Seasonal (Seasonal Auto-Regressive Process of Order P)

q คือ อันดับที่ q ของขบวนการเฉลี่ยเคลื่อนที่แบบ Nonseasonal (Nonseasonal Moving Average Process of Order q)

Q คือ อันดับที่ Q ของขบวนการเฉลี่ยเคลื่อนที่แบบ Seasonal (seasonal Moving Average Process of Order Q)

$d.D$ คือ จำนวนครั้งของการหาผลต่างแบบ Nonseasonal อันดับที่ d และผลต่างแบบ Seasonal อันดับที่ D ตามลำดับ เพื่อทำให้อนุกรมเวลามีคุณสมบัติ Stationary

S คือ จำนวนคาบเวลาของฤดูกาลใน 1 รอบ

t คือ เวลา

B คือ ตัวดำเนินการย้อนหลังเวลา (Backward Shift Operator)

$X_{1,t},\ldots,X_{k,t}$ คือ ค่าของอนุกรมเวลาชุดที่ 1,..., ชุดที่ k ณ เวลาที่ k ที่มีอิทธิพลต่ออนุกรมเวลาที่ต้องการพยากรณ์ (Y_t) ($X_{1,t},\ldots,X_{k,t}$ ต้องมีคุณสมบัติ Stationary)

β_0 คือ พารามิเตอร์แสดงค่าคงที่ (Constant Term) ในตัวแบบ

$\beta_1,\ldots,\beta_k$ คือ พารามิเตอร์แสดงสัมประสิทธิ์การถดถอย (Regression Coefficients)

$\phi_1, \phi_2, ..., \phi_p$ คือ พารามิเตอร์แสดงสัมประสิทธิ์การถดถอยในตัวเองแบบ Nonseasonal อันดับที่ 1, 2,...*p* ตามลำดับ (Nonseasonal Auto Regressive Process of Order: 1, 2,...*p*)

$\theta_1, \theta_2, ..., \theta_q$ คือ พารามิเตอร์แสดงสัมประสิทธิ์ของการเฉลี่ยเคลื่อนที่แบบ Nonseasonal อันดับที่ 1,2,..q ตามลำดับ (Nonseasonal Moving Average Process of Order 1, 2,...q)

$\Phi_1, \Phi_2, ..., \Phi_P$ คือ พารามิเตอร์แสดงสัมประสิทธิ์การถดถอยในตัวเองแบบ Seasonal อันดับที่ 1, 2,...P ตามลำดับ (Seasonal Auto Regressive Process of Order 1, 2,...P)

$\Theta_1, \Theta_2, ..., \Theta_Q$ คือ พารามิเตอร์แสดงสัมประสิทธิ์ของการเฉลี่ยเคลื่อนที่แบบ Seasonal อันดับที่ 1, 2,...Q ตามลำดับ (Seasonal Moving Average Process of Order 1, 2,...Q)

$(1-B)^d Y_t$ คือ ผลต่างอันดับที่ d ของอนุกรมเวลา Y_t

$(1-B^s)^D Y_t$ คือ ผลต่างอันดับที่ D ของอนุกรมเวลา Y_t

ε_t คือ ค่าความคลาดเคลื่อนในการพยากรณ์ ณ เวลาที่ t

4. การสร้างตัวแบบ ARIMA ที่ใช้ในงานวิจัย

ในงานวิจัยนี้ใช้ตัวแบบ ARIMA (p, d, q) ในการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยโดยใช้วิธีประมาณค่าพารามิเตอร์ด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ใช้การแจกแจงแบบโคชี ในตัวแบบ ARIMA ประกอบด้วยตัวแบบ ARIMA (1, 0, 0) และตัวแปรราคาข้าว (a_t)

4.1 ปรับข้อมูลโดยวิธีประมาณค่าพารามิเตอร์ด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ใช้การแจกแจงแบบโคชี ตัวแบบที่ได้ คือ ARIMA (1, 0, 0) สามารถเขียนเป็นสมการ

$$a_t = y_t - \frac{3}{4nh^3} \sum (h^2 - (x - x_i)^2) I + \rho \varepsilon_{i-1}$$
$$\hat{a}_t = y_t - (0.0002151) + (-0.252)(1.89478x10^{-14})$$
$$\hat{a}_t = y_t - 0.0002151$$

4.2 สร้างตัวแบบ ARIMA (p, d, q) มีสมบัติทั่วไป คือ ARIMA (1, 0, 0)

$$(1 - \phi_1 B)(y_t - \theta_0) = \varepsilon_t$$
$$y_t - \theta_0 - \phi_1 B y_t + \theta_0 \phi_1 B = \varepsilon_t$$

$$y_t - \theta_0 - \phi_1 y_{t-1} + \theta_0 \phi_1 B = \varepsilon_t$$
$$y_t = \theta_0 - \theta_0 \phi_1 + \phi_1 y_{t-1} + \varepsilon_t$$

เมื่อ y_t คือ ค่าของข้อมูลอนุกรมเวลาที่ต้องการพยากรณ์ ณ เวลาที่ t

y_{t-1} คือ ค่าของข้อมูลอนุกรมเวลาที่ต้องการพยากรณ์ ณ เวลาที่ $t-1$

B คือ ตัวดำเนินการย้อนหลังเวลา (Backward shift Operator)

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลาที่ t

θ_0 คือ พารามิเตอร์แสดง Constant Term

ϕ_1 คือ พารามิเตอร์แสดงสัมประสิทธิ์การถดถอยในตัวเองแบบ Nonseasonal อันดับที่ 1 (Nonseasonal Auto-Regressive Process of Order 1)

4.3 ตรวจสอบความเหมาะสมของตัวแบบโดยใช้สถิติทดสอบ Q ของ Box-Ljung

ตอนที่ 4 การตรวจสอบความเหมาะสมของอนุกรมเวลา

การพยากรณ์จะต้องคำนึงจากองค์ประกอบที่สำคัญ ต้องเลือกวิธีการพยากรณ์ให้เหมาะสมกับสถานการณ์และลักษณะของปัญหาที่แตกต่างกันไป ดังนี้

1. รูปแบบของค่าพยากรณ์ เมื่อผู้วิจัยหรือผู้ที่ใช้เกี่ยวกับค่าพยากรณ์ สิ่งที่จำเป็นสำหรับผู้ที่พยากรณ์ คือ รูปแบบของการพยากรณ์เพื่อนำไปใช้ประโยชน์ในการบริหารจัดการ เช่น ต้องการรูปแบบการพยากรณ์แบบจุดหรือต้องการรูปแบบการพยากรณ์แบบช่วงเพราะบางวิธีสามารถให้ค่าพยากรณ์แบบจุดเท่านั้น

2. ความแม่นยำ ความแม่นยำหรือความถูกต้องของค่าพยากรณ์ เป็นส่วนประกอบที่สำคัญ ในการเลือกวิธีการพยากรณ์ เพราะถ้าวิธีการพยากรณ์ที่ให้ค่าความแม่นยำสูง สามารถใช้ในการบริหารจัดการได้ตามที่ต้องการ ช่วยลดต้นทุนต่าง ๆ และยังเป็นการเพิ่มความสามารถในการบริหารจัดการ สำหรับผู้ที่ใช้ค่าพยากรณ์เพราะวิธีการพยากรณ์บางวิธีให้ค่าความแม่นยำเฉพาะบางช่วงเวลา เช่นเหมาะสมในการพยากรณ์ระยะสั้น หรือบางวิธีเหมาะสมในการพยากรณ์ระยะยาว

3. กรอบเวลา การพยากรณ์เชิงปริมาณเป็นการทำนายเหตุการณ์ในอนาคต อาจจะเป็นระยะเวลารายวัน รายสัปดาห์ รายเดือน รายไตรมาส หรือรายปี เป็นต้น ช่วงเวลาเหล่านี้เรียกว่ากรอบเวลาโดยทั่วไปจำแนกกรอบเวลาตามความยาวของระยะเวลา ดังนี้ 1. การพยากรณ์ระยะใกล้ หมายถึง การพยากรณ์ในช่วงเวลาที่ห่างจากปัจจุบันน้อยกว่า 1 เดือน 2. การพยากรณ์ระยะสั้น หมายถึง การพยากรณ์ในช่วงเวลาที่ห่างจากปัจจุบันตั้งแต่ 1 เดือนแต่ไม่เกิน 3 เดือน 3. การพยากรณ์

ระยะกลาง หมายถึง การพยากรณ์ในช่วงเวลาที่ห่างจากปัจจุบัน มากกว่า 3 เดือน แต่น้อยกว่า 2 ปี และ 4. การพยากรณ์ระยะยาว หมายถึง การพยากรณ์ในช่วงเวลาที่ห่างจากปัจจุบัน ตั้งแต่ 2 ปี ขึ้นไป

4. ลักษณะของข้อมูล ผู้พยากรณ์ควรตรวจสอบข้อมูลที่มีอยู่เป็นข้อมูลประเภทใด เป็นค่าของตัวแปรที่จะพยากรณ์เท่านั้นหรือเป็นค่าของตัวแปรอื่นที่มาเกี่ยวข้องด้วย เพื่อจะได้เลือกวิธีในการพยากรณ์ได้ถูกต้องและให้ค่าที่แม่นยำสูง

5. ค่าใช้จ่าย การพยากรณ์แต่ละครั้งย่อมมีค่าใช้จ่ายหลายประการเกิดขึ้น เช่น ค่าใช้จ่ายในการเก็บรวบรวมข้อมูล ค่าใช้จ่ายเกี่ยวกับการวิเคราะห์ข้อมูล และในบางครั้งถ้าต้องการใช้ค่าพยากรณ์ที่มีความแม่นยำสูง ค่าใช้จ่ายที่ใช้ในการพยากรณ์ยิ่งสูงเพิ่มขึ้นด้วย นอกจากนี้ ถ้าพิจารณาการพยากรณ์ที่ซับซ้อนมากยิ่งขึ้น ต้องใช้เวลาในการคำนวณมากขึ้น เป็นสาเหตุที่ให้ค่าใช้จ่ายในการพยากรณ์สูงขึ้น กล่าวคือ ถ้าใช้วิธีการพยากรณ์ที่มีความซับซ้อนมาก และใช้จำนวนข้อมูลในการพยากรณ์มาก ค่าใช้จ่ายที่เกิดขึ้น จะสูงกว่าวิธีการพยากรณ์ที่ไม่ซับซ้อนและใช้ข้อมูลน้อย แต่วิธีการพยากรณ์ที่มีความซับซ้อนมาก และใช้จำนวนข้อมูลในการพยากรณ์มาก ๆ ก็ให้ค่าความแม่นยำสูงกว่าด้วยเช่นกัน

6. ข้อมูลที่มีอยู่ในการพยากรณ์ผู้พยากรณ์ต้องทราบว่า ชนิดของข้อมูลในอดีตเป็นข้อมูลที่ทันสมัย หรือเป็นอดีตมากเกินไป จนแตกต่างจากข้อมูลที่ทันสมัยมากกว่า เหมาะสมที่จะใช้ทั้งหมดหรือไม่ หรือจำเป็นต้องใช้แค่บางส่วน ข้อมูลที่ต้องการพยากรณ์และตัวแปรต่าง ๆ ที่เกี่ยวข้องมีอยู่สมบูรณ์หรือไม่ และจำนวนข้อมูลเท่าใดถึงจะเพียงพอกับเงื่อนไขในการพยากรณ์บางวิธี เพื่อให้ได้ค่าพยากรณ์ที่มีความแม่นยำสูง

7. ความซับซ้อนวิธีในการพยากรณ์ที่ใช้ จะทำให้ผู้ใช้ค่าพยากรณ์เกิดความเชื่อมั่นในการตัดสินใจในการตอบปัญหาต่าง ๆ ความซับซ้อนของวิธีการพยากรณ์ ควรอยู่ในระดับที่ผู้พยากรณ์สามารถเข้าใจ และอธิบายผลการนำไปใช้ได้ง่าย เป็นสิ่งที่จำเป็นอย่างยิ่งในการใช้วิธีในการพยากรณ์ กล่าวคือ วิธีที่ไม่ซับซ้อนที่ให้ความแม่นยำในการพยากรณ์น้อยกว่านั้น จะเป็นวิธีที่เหมาะสมกว่าวิธีในการพยากรณ์ที่มีความซับซ้อน แต่ให้ค่าพยากรณ์ที่มีความแม่นยำสูง ถ้าความแม่นยำแตกต่างกันไม่มาก และในบางปัญหาอาจมีวิธีการพยากรณ์มากกว่า 1 วิธี ผู้ใช้ค่าพยากรณ์ต้องปรับให้เหมาะกับปัญหาในการตอบโจทย์ในการใช้ค่าพยากรณ์

การตรวจสอบความเหมาะสมของอนุกรมเวลา พิจารณา ดังนี้ 1) กระบวนการสเตชันนารี (Stationary) 2) ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (Autocorrelation Function Coefficient: ACF) 3) ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองส่วนย่อย (Partial Autocorrelation

Function Coefficient: PACF) 4) ความคลาดเคลื่อนในการพยากรณ์ และ 5) การวิเคราะห์ส่วนตกค้าง (Residual) รายละเอียดมี ดังนี้

1. กระบวนการสเตชันนารี (Stationary)

ในการวิเคราะห์อนุกรมเวลาด้วยตัวแบบสโตแคสติค อนุกรมเวลาต้องมีสมบัติสเตชันนารี กล่าวคือ ค่าเฉลี่ยและความแปรปรวนของกระบวนการมีค่าคงที่ทุกหน่วยเวลา t ใด ๆ

ให้ $Y_{t_1}, Y_{t_2}, ..., Y_{t_s}$ แทน ตัวแปรเวลา ที่ $t_1, t_2, ..., t_s$ และ $Y_{t_{1+k}}, Y_{t_{2+k}}, ..., Y_{t_{s+k}}$ เมื่อ k แทนจำนวนจริงใด ๆ แทนตัวแปรเวลา ที่ $t_{1+k}, t_{2+k}, ..., t_{s+k}$ แล้ว กระบวนการสเตชันนารี คือกระบวนการที่มีการแจกแจงร่วมของตัวแปร $Y_{t_1}, Y_{t_2}, ..., Y_{t_s}$ เป็นการแจกแจงเดียวกับการแจกแจงร่วมของตัวแปร $Y_{t_{1+k}}, Y_{t_{2+k}}, ..., Y_{t_{s+k}}$ จะได้ว่า $E(\mathrm{Y}_t) = \mu$ และ $Var(\mathrm{Y}_t) = \sigma^2$ สำหรับทุกค่าหน่วยเวลา t ที่มีค่าคงที่ และค่าความแปรปรวนร่วม (Covariance) ระหว่าง Y_{t_1} และ Y_{t_2} มีค่าเท่ากับความแปรปรวนร่วมระหว่าง $Y_{t_{1+k}}$ และ $Y_{t_{2+k}}$ สามารถเขียนเป็นสัญลักษณ์ได้เป็น $\mathrm{Cov}(\mathrm{Y}_{t_1}, \mathrm{Y}_{t_2}) = \mathrm{Cov}(\mathrm{Y}_{t_{1+k}}, \mathrm{Y}_{t_{2+k}})$ เมื่อพิจารณาความแปรปรวนร่วมระหว่าง 2 คาบเวลาใด ๆ จะได้ $\mathrm{Cov}(\mathrm{Y}_{1_1}, \mathrm{Y}_{1+k}) = ... = \mathrm{Cov}(\mathrm{Y}_{t-k}, \mathrm{Y}_{s-k})$ ณ เวลาที่ t, s และ k ใด ๆ กล่าวคือกระบวนการมีสมบัติสเตชันนารี การแจกแจงร่วมระหว่าง Y_{t_s} และ $\mathrm{Y}_{t_{s+k}}$ ขึ้นอยู่กับคาบเวลา t และขึ้นอยู่กับระยะห่างของช่วงเวลา k หน่วย (Lag k) สรุปได้ว่า อนุกรมเวลา สเตชันนารี เป็นอนุกรมเวลาที่มีการเปลี่ยนแปลงรอบค่าเฉลี่ย โดยมีค่าความแปรปรวนคงที่ สำหรับอนุกรมเวลาที่มีการเปลี่ยนแปลงรอบค่าเฉลี่ยไม่คงที่ และความแปรปรวนของอนุกรมเวลาไม่คงที่ เรียกว่า อนุกรมเวลาไม่สเตชันนารี

2. ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (Autocorrelation Function Coefficient: ACF)

ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง (ρ) เป็นมาตรวัดความสัมพันธ์ระหว่าง Y_t และ Y_{t-k} ใช้สัญลักษณ์เป็น ρ_k สามารถคำนวณได้จาก

$$\rho_k = \frac{\mathrm{E}(\mathrm{Y}_t - \mu)(\mathrm{Y}_{t-k} - \mu)}{\sqrt{\mathrm{E}(\mathrm{Y}_t - \mu)^2}\sqrt{\mathrm{E}(\mathrm{Y}_t - \mu)^2}} = \frac{\gamma_k}{\sqrt{\mathrm{E}(\mathrm{Y}_t - \mu)^2}\sqrt{\mathrm{E}(\mathrm{Y}_t - \mu)^2}}$$

เนื่องจาก $\mathrm{Var}(\mathrm{Y}_t) = \mathrm{Var}(\mathrm{Y}_{t-k}) = \mathrm{E}(\mathrm{Y}_t - \mu)^2 = \gamma_0$

จะได้ $\rho_k = \dfrac{\gamma_k}{\gamma_0}$ และ $\rho_0 = 1$ สำหรับ γ_k ในกระบวนการสเตชันนารีมีสมบัติดังนี้

1. $\gamma_0 = \mathrm{Var}(\mathrm{Y}_t)$
2. $|\gamma_k| \le \gamma_0$

3. $\gamma_k = \gamma_{-k}$

4. $\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\gamma_{|t_i-t_j|} \geq 0$, ทุกเซตของ $t_1,...,t_n$ และจำนวนจริงใด ๆ $\alpha_1,...,\alpha_n$

สำหรับสมบัติของ ρ_k มีดังนี้

1. $\rho_0 = 1$

2. $|\rho_k| \leq 1$

3. $\rho_k = -\rho_k$

4. $\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\rho_{|t_i-t_j|} \geq 0$, ทุกเซตของ $t_1,...,t_n$ และจำนวนจริงใด ๆ $\alpha_1,...,\alpha_n$

เมื่อพิจารณาสหสัมพันธ์ในตัวเอง ในลักษณะที่เป็นฟังก์ชันของช่วงเวลาที่ห่างกัน k หน่วย เรียก ρ_k ว่าฟังก์ชันสหสัมพันธ์ในตัวเอง กราฟฟังก์ชันสหสัมพันธ์ในตัวเองของอนุกรมเวลาสเตชันนารีหรือ ACF เป็นกราฟที่มีลักษณะลดลงรวดเร็วเข้าสู่ศูนย์ หรือ กราฟซึ่งมีค่าเป็นศูนย์ทุกค่า เมื่อ $k > q$ สามารถเป็นจริงได้เมื่อ $q = 1$

ในการหาค่า ρ_k จากตัวอย่างใช้สัญลักษณ์เป็น r_k เรียกว่า สหสัมพันธ์ของตัวอย่าง เมื่อย้อนเวลาไป k หน่วย หรือ lag k คำนวณได้จาก

$$r_k = \frac{\sum_{t=k+1}^{n}(\mathrm{Y}_t - \bar{\mathrm{Y}})(\mathrm{Y}_{t-k} - \bar{\mathrm{Y}})}{\sum_{t=1}^{n}(\mathrm{Y}_t - \bar{\mathrm{Y}})^2} = \frac{\sum_{t=1}^{n-1}(\mathrm{Y}_t - \bar{\mathrm{Y}})(\mathrm{Y}_{t+k} - \bar{\mathrm{Y}})}{\sum_{t=1}^{n}(\mathrm{Y}_t - \bar{\mathrm{Y}})^2}$$

เมื่อ $\bar{Y} = \frac{\sum_{t=1}^{n} Y_t}{n}$

สำหรับกราฟ SACF ของอนุกรมเวลา สเตชันนารี มีลักษณะลดลงอย่างรวดเร็วเข้าสู่ศูนย์ เหมือนกราฟ ACF โดยมีค่าเป็นศูนย์ที่ระยะห่างของช่วงเวลาเท่ากับ q หรือกราฟ SACF ลดลงแบบเลขชี้กำลัง สำหรับกราฟ SACF มีลักษณะลดลงช้า ๆ แบบเชิงเส้น แสดงว่าอนุกรมเวลาไม่สเตชันนารี

3. ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองส่วนย่อย (Partial Autocorrelation Function Coefficient: PACF)

การพิจารณาสหสัมพันธ์ระหว่างตัวแปร Y_t และ Y_{t-k} สหสัมพันธ์ดังกล่าว อาจเป็นผลมาจากสหสัมพันธ์ระหว่างสองตัวแปรกับตัวแปร $Y_{t-1},...,Y_{t-k+1}$ เพื่อให้ได้ความสัมพันธ์ระหว่าง Y_t และ Y_{t-k} ในการขจัดความสัมพันธ์ของ $Y_{t-1},...,Y_{t-k+1}$ ต้องใช้การหาสหสัมพันธ์แบบมีเงื่อนไข

$Corr(\mathrm{Y}_t, \mathrm{Y}_{t-k} / \mathrm{Y}_{t-1},..., \mathrm{Y}_{t-k+1})$ เรียกว่า สหสัมพันธ์ในตัวเองบางส่วน (ϕ_{kk}) คำนวณได้จาก

$$\phi_{\mathrm{kk}} = \frac{\mathrm{Cov}((\mathrm{Y_t} - \hat{\mathrm{Y}}_\mathrm{t}),(\mathrm{Y_{t-k}} - \hat{\mathrm{Y}}_\mathrm{t-k})}{\sqrt{\mathrm{Var}((\mathrm{Y_t} - \mathrm{Y_t})}\sqrt{\mathrm{Var}(\mathrm{Y_{t-k}} - \mathrm{Y_{t-k}})}}$$

เมื่อ $\hat{\mathrm{Z}}_\mathrm{t} = \beta_1 \mathrm{Y_{t-1}} + \beta_2 \mathrm{Y_{t-2}} + ... + \beta_{\mathrm{k-1}} \mathrm{Y_{t-k+1}}$

สามารถเขียน ϕ_{kk} ในรูปของสมการการถดถอย คือ

$$\mathrm{Z_t} = \phi_{\mathrm{k1}} \mathrm{Y_{t-1}} + \phi_{\mathrm{k2}} \mathrm{Y_{t-2}} + ... + \phi_{\mathrm{kk}} \mathrm{Y_{t-p}} + \mathrm{a_t}$$

การคำนวณสหสัมพันธ์ในตัวเองบางส่วนของตัวอย่าง ($\hat{\phi}_{\mathrm{kk}}$) เมื่อช่วงเวลาห่างกัน k หน่วย สามารถคำนวณได้จาก

$$\hat{\phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j}, k = 2,3,... \ ,$$

โดยที่ $\hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk}\hat{\phi}_{k-1,k-j}$

สมบัติที่สำคัญสำหรับกระบวนการสเตชันนารี คือ กราฟของ ϕ_{kk} อาจมีค่าเป็นศูนย์เมื่อช่วงเวลาห่างกันมากกว่า q หรือลดลงเข้าสู่ศูนย์อย่างรวดเร็วเช่นเดียว ACF

4. ความคลาดเคลื่อนในการพยากรณ์

การพยากรณ์เป็นการคาดการณ์ข้อมูลในอนาคต การพยากรณ์ที่ดีนั้น ค่าพยากรณ์ที่ได้ควรมีค่าใกล้เคียงกับค่าจริงให้มากที่สุด ทำให้ความผิดพลาดในการพยากรณ์ต่ำที่สุด (มุกดา แม้นมินทร์, 2549, หน้า 11–13: ดวงกมล แก้วสุขมณี และสุชาดา กรเพชรปาณี, 2558) สำหรับสถิติที่ใช้วัดความผิดพลาดในการพยากรณ์ มีดังต่อไปนี้

4.1 ความคลาดเคลื่อนเฉลี่ย (Mean Error: ME)

$$Me = \frac{\sum_{t=1}^{n} e_t}{n}$$

เมื่อ e_t คือความคลาดเคลื่อนจากการพยากรณ์ = ค่าจริง–ค่าพยากรณ์

4.2 Mean Absolute Deviation (MAD)

$$MAD = \frac{\sum_{t=1}^{n} |e_t|}{n}$$

4.3 ค่าเฉลี่ยคลาดเคลื่อนกำลังสอง (Mean Squared Error: MSE)

$$MSE = \frac{\sum_{t=1}^{n} e_t^2}{n}$$

4.4 ค่าเบี่ยงเบนมาตรฐานของความคลาดเคลื่อน (Standard Deviation of Error: SDE)

$$SDE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n-1}}$$

4.5 ร้อยละของค่าคลาดเคลื่อน (Percentage Error: PE)

$$PE = (\frac{Y_t - F_t}{Y_t})x100 = \frac{e_t}{Y_t}x100$$

Y_t คือ ข้อมูล ณ เดือนที่ t

F_t คือ ค่าพยากรณ์ ณ เดือนที่ t

4.6 ร้อยละของค่าคลาดเคลื่อนเฉลี่ย (Mean Percentage Error: MPE)

$$MPE = \frac{\sum_{t=1}^{n} PE}{n}$$

4.7 ร้อยละของค่าคลาดเคลื่อนเฉลี่ยสัมบูรณ์ (Mean Absolute Percentage Error: MAPE)

$$MAPE = \frac{\sum_{t=1}^{n} |PE|}{n}$$

5. การวิเคราะห์ส่วนตกค้าง (Residual)

ในการเลือกตัวแบบ ARIMA หรือตัวแบบ ARIMAX ที่เหมาะสมแล้วและการประมาณค่าพารามิเตอร์มีค่าใกล้เคียงกับค่าจริง ส่วนตกค้าง (a_t) มีสมบัติดังต่อไปนี้

1. ส่วนตกค้างไม่มีสหสัมพันธ์ในตัวเอง
2. ส่วนตกค้างมีการแจกแจงแบบปรกติมีค่าเฉลี่ยเป็น 0 และความแปรปรวนคงที่

สำหรับการตรวจสอบส่วนตกค้างไม่มีสหสัมพันธ์ในตัวเองใช้สถิติ สถิติทดสอบ Q ของ Box-Ljung คือ สถิติที่ใช้ในการทดสอบสหสัมพันธ์ในตัวเองของ Residual ตั้งแต่ Lag ที่ 1, 2,...,k ว่ามีค่าแตกต่างจากศูนย์หรือไม่ มีขั้นตอน ดังนี้

สมมติฐานของการทดสอบ

$\mathrm{H}_0 : \rho_{\mathrm{e}}(1) = \rho_{\mathrm{e}}(2) = ... = \rho_{\mathrm{e}}(\mathrm{k}) = 0$ (ตัวแบบการพยากรณ์ที่พิจารณามีความเหมาะสม)

$\mathrm{H}_1 : \rho_{\mathrm{e}}(\mathrm{k})$ อย่างน้อย 1 ค่าที่แตกต่างจากศูนย์ (ตัวแบบการพยากรณ์ที่พิจารณาไม่มีความเหมาะสม)

1. สถิติทดสอบ คือ

$$Q = n(\mathrm{n}+2)\sum_{k=1}^{k}[\frac{r_e^2(\mathrm{k})}{n-k}]$$

โดยที่ $$r_{e(\mathrm{k})} = \frac{\sum_{t=k+1}^{n}(\mathrm{e}_t-\overline{\mathrm{e}})(\mathrm{e}_{t-k}-\overline{\mathrm{e}})}{\sum_{t=1}^{n}(\mathrm{e}_t-\overline{\mathrm{e}})^2}$$

$r_{e(k)}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเองของส่วนเหลือ ณ Lag k

k คือ ช่วงเวลาที่จะพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง

$$k = 1, 2, ..., \frac{n}{4}$$

n คือ จำนวนของส่วนตกค้าง

3) เกณฑ์ที่ใช้ในการตัดสินใจ คือ จะปฏิเสธ H_0 ถ้า $Q > \chi^2_{\alpha, df=k-p}$

โดยที่ p คือ จำนวนพารามิเตอร์ในตัวแบบที่พิจารณา

ตอนที่ 5 อัตตสหสัมพันธ์ในข้อมูลอนุกรมเวลา

การวิเคราะห์ด้วยการถดถอย มีข้อตกลงเบื้องหนึ่งที่เราจะต้องนำมาพิจารณาด้วยเสมอ คือ การที่ค่าความคลาดเคลื่อน (ε_i) ซึ่งเป็นกลุ่มตัวแปรสุ่ม แต่ละค่าของ ε_i จะต้องมีการแจกแจงปกติและเป็นอิสระต่อกัน (Independent Normal Random Variable) ในบางกรณีเราพบว่า

ความคลาดเคลื่อน อาจมีความสัมพันธ์กันเอง กล่าวคือ $E(\varepsilon_i\varepsilon_j) \neq 0$ เมื่อ $i \neq j$ ซึ่งเป็นการขัดแย้งกับข้อตกลงสำหรับการวิเคราะห์ด้วยการถดถอยที่กำหนดไว้ว่า $E(\varepsilon_i\varepsilon_j) \neq 0$ เมื่อ $i \neq j$ เหตุการณ์เช่นนี้มักพบได้เสมอ โดยเฉพาะอย่างยิ่งการนำเทคนิคการวิเคราะห์ถดถอยไปใช้ในการวิเคราะห์ในทางธุรกิจ และเศรษฐศาสตร์ ซึ่งเป็นข้อมูลที่เก็บรวบรวมตามเวลาที่เกิดขึ้นตามลำดับก่อนหลัง การที่ข้อมูลมีปัญหาอัตตสัมพันธ์ที่เกิดขึ้น ผู้วิจัยจะต้องทำการปรับแก้ไขข้อมูลนั้นเสียก่อนที่จะนำสมการถดถอยไปใช้ในการพยากรณ์ มิฉะนั้นแล้วค่าพยากรณ์ที่ได้อาจเกิดความผิดพลาดได้

1. สาเหตุการเกิดปัญหาการมีอัตตสหสัมพันธ์

การเกิดอัตตสหสัมพันธ์ในตัวเอง (Autocorrelation หรือ Serial Correlation) อาจเกิดขึ้นจากสาเหตุต่าง ๆ ดังนี้

1.1 การไม่ได้พิจารณาตัวแปรอิสระที่สำคัญ

โดยปกตินักวิจัยมีความจำเป็นต้องกำหนดตัวแปรอิสระเพื่อใช้ในการอธิบายตัวแปรตามที่เราสนใจให้ครบถ้วนที่สุดเท่าที่จะกระทำได้ เพราะถ้าสามารถกำหนดตัวแปรอิสระได้ครบถ้วนมากเพียงใด เราย่อมทราบธรรมชาติตลอดจนพฤติกรรมในอดีต–ปัจจุบัน–อนาคตของ ตัวแปรตามได้มากขึ้นเท่านั้น แต่อย่างไรก็ตามอาจจะมีความผิดพลาดเกิดขึ้นได้ เพราะการที่กำหนดตัวแปรอิสระให้ได้ครบถ้วนนั้นเป็นเรื่องยาก อาจเป็นไปได้ที่เราไม่ได้พิจารณาตัวแปรอิสระบางตัวหรือหลายตัวไปโดยไม่ตั้งใจ นอกจากนี้เราอาจตัดตัวแปรอิสระบางส่วนทิ้งด้วยเหตุจำเป็นเพราะไม่อาจวัดค่าสังเกตได้ ตัวแปรที่ไม่ได้พิจารณานี้มีการควบคุมการเปลี่ยนแปลงของ Y อยู่นอกสมการ $Y = f(X's)$ โดยส่งอิทธิพลผ่านตัวแปรสุ่ม ε ดังนั้นตัวแปรสุ่ม ε จึงขาดความเป็นอิสระ ทำให้เกิดความสัมพันธ์ระหว่างตัวแปรสุ่ม ε ขึ้นได้ สหสัมพันธ์ในตัวเองลักษณะนี้ จึงไม่ใช่ธรรมชาติแท้จริงของของ ε หรือไม่ใช่สหสัมพันธ์แท้ ๆ ของ ε แต่แสดงให้เห็นเสมือนว่า ε มีสหสัมพันธ์ เรียกว่า Quasi–Autocorrelation

1.2 การกำหนดความสัมพันธ์ทางคณิตศาสตร์ผิดพลาด

การระบุความสัมพันธ์ทางคณิตศาสตร์ระหว่างตัวแปรอิสระกับตัวแปรตามผิดจากสภาพความสัมพันธ์ที่แท้จริงจะมีผลให้ตัวแปรสุ่ม ε เกิดความสัมพันธ์กันเองขึ้นได้ เช่น ความสัมพันธ์ที่แท้จริงของสมการ $Y = f(X's)$ คือ ความสัมพันธ์ชนิดพหุนามกำลังสอง แต่นักวิจัยกำหนดให้เป็นสมการเชิงเส้น ความคลาดเคลื่อนจึงเกิดขึ้นและความคลาดเคลื่อนเหล่านี้จะถูกนำไปรวมกันไว้ในตัวแปรสุ่ม ε ขาดความเป็นอิสระตามธรรมชาติเดิม และเกิดปัญหาสหสัมพันธ์ในตัวเองขึ้น

1.3 การระบุข้อตกลงว่า $E(\varepsilon_i\varepsilon_j) = 0$ ทั้ง ๆ ที่ $E(\varepsilon_i\varepsilon_j) \neq 0$ เมื่อ $i \neq j$

โดยปกติแล้วตัวแปรเป็นจำนวนมากมักมีธรรมชาติของความเกี่ยวเนื่องกันภายในตัวเอง ในระหว่างช่วงเวลาอยู่แล้ว เช่น ราคาสินค้าปัจจุบันสัมพันธ์ในทางบวกกับราคาในอดีตกล่าวคือ การกำหนดราคาสินค้า ผู้กำหนดราคานิยมเอาราคาในอดีตเป็นฐาน ซึ่งการกระทำเช่นนี้มีผลให้ตัวแปร ราคาเองมีสหสัมพันธ์ในตัวเอง ขณะเดียวกันตัวแปรอิสระอาจมีความสัมพันธ์กันเองได้ด้วย สาเหตุ อื่น ๆ หลายประการ เช่น การนัดหยุดงานของคนงานมีผลให้ปริมาณการผลิต ยอดขาย สภาพคล่อง และอื่น ๆ กระทบกระเทือนเป็นลูกโซ่หลายช่วงเวลา (Period) ด้วยเหตุที่ตัวแปรอิสระมักมี ความสัมพันธ์ในตัวเองอยู่แล้วตามธรรมชาติ (True Autocorrelation) การกำหนดให้ $E(\varepsilon_i \varepsilon_j) \neq 0; i \neq j$ โดยไม่ได้ตรวจสอบ จะส่งผลต่อประสิทธิภาพและความแม่นยำในการวิเคราะห์ ข้อมูล

2. ผลกระทบของการเกิดอัตตสหสัมพันธ์ต่อวิธีกำลังสองน้อยที่สุด

เมื่อเกิดปัญหาสหสัมพันธ์ในตัวเองขึ้น คือ $Var(\varepsilon \text{a} \neq \sigma_\varepsilon^2 \text{I}_n)$ แต่ยังคงใช้วิธีกำลัง สองน้อยที่สุดในการประมาณค่า β อยู่เช่นเดิม จะส่งผลให้ตัวประมาณที่ได้ขาดความแม่นยำ โดยเฉพาะในแง่ของคุณภาพของ b ดังนี้

1) b ยังเป็นตัวประมาณค่าที่ไม่เอนเอียงของ β แต่โดยทั่วไปจะมีประสิทธิภาพ ต่ำกว่าวิธีอื่น ๆ นั่นคือ

$$\begin{aligned} E(\text{b}) &= \text{E}[(\text{X}'\text{X})^{-1} X'Y] \\ &= E[(\text{X}'\text{X})^{-1} X'(\text{X}\beta + \varepsilon)] \\ &= \beta + (\text{X}'\text{X})^{-1} X'E(\varepsilon = \beta) \end{aligned}$$

ดังนั้น $E(\text{b}) = \beta$

2) ถ้าเกิดปัญหาสหสัมพันธ์ในตัวเองขึ้นโดยเฉพาะอย่างยิ่ง คือ Positive Autocorrelation ค่าประมาณของ σ^2 ซึ่งประมาณขึ้นโดยอาศัยค่าความคลาดเคลื่อน คือ $\hat{\sigma}^2 = \dfrac{\sum e_i^2}{n-k-1}; i = 1, 2, 3, ..., n$ จะมีค่าต่ำกว่าความเป็นจริง ทั้งนี้เพราะในกรณีเช่นนี้ค่าจริง ของ ε จะจับกลุ่มกันห่างไกลจากสมการเส้นตรงจริง (True Line) ขณะที่ค่าประมาณ ε จะจับกลุ่ม กันใกล้สมการเส้นตรงที่ประมาณ (Estimated Line)

3) เมื่อเกิดปัญหาสหสัมพันธ์ในตัวเองขึ้น ค่าประมาณของ $Var(\beta)$ คือ $Var(\text{b})$ จะต่ำกว่าความเป็นจริง โดยเฉพาะอย่างยิ่งเมื่อเราประมาณค่า β โดยอาศัยวิธีกำลังสองน้อยที่สุด กล่าวคือ $Var(\text{b}) = \hat{\sigma}^2 (\text{X}'\text{X})^{-1}$ เหตุที่เป็นเช่นนี้เพราะ $\hat{\sigma}^2$ มีค่าต่ำกว่าความเป็นจริงประการหนึ่ง

และในกรณีของสหสัมพันธ์ในตัวเองนั้น $E(\varepsilon_i\varepsilon_j) \neq 0; \mathrm{i} \neq \mathrm{j}$ แสดงว่า $Cov(\varepsilon_i\varepsilon_j)$ จะต้องมีบทบาทในการกำหนดค่า $Var(\mathrm{b})$ ด้วย ถ้าตัด $Cov(\varepsilon_i\varepsilon_j)$ ทิ้งไป โดยถือเสมือน $Cov(\varepsilon_i\varepsilon_j) = 0$ ผลที่ตามมา คือ $Var(\mathrm{b})$ มีค่าต่ำกว่าความเป็นจริง ค่า $Var(\mathrm{b})$ ที่ต่ำ ๆ ในกรณีนี้หากผู้วิจัยไม่ระมัดระวังและไม่มีความเข้าใจในเรื่องเหล่านี้ อาจเข้าใจผิดคิดว่างานของตนมีคุณภาพสูงทั้ง ๆ ที่ความจริงมิได้เป็นเช่นนั้น

4) เมื่อเกิดปัญหาอัตตสหสัมพันธ์ในตัวเองขึ้น ตัวประมาณโดยวิธีกำลังสองน้อยที่สุดจะมีประสิทธิภาพต่ำ เพราะ $Var(\mathrm{b})$ โดยวิธีกำลังสองน้อยที่สุด จะมีค่าสูงกว่า $Var(\mathrm{b})$ ที่ประมาณได้โดยอาศัยวิธีที่มีความเหมาะสมกว่า

5) ค่า MSE ที่คำนวณได้อาจจะมีค่าต่ำกว่าความเป็นจริง

6) การประมาณช่วงความเชื่อมั่นและการทดสอบสมมติฐาน โดยใช้การแจกแจง t และ F อาจเกิดความผิดพลาดได้

3. การตรวจสอบการเกิดอัตตสหสัมพันธ์

การตรวจสอบปัญหาการเกิดอัตตสหสัมพันธ์ โดยพิจารณาจากแผนภาพการกระจายของค่าความคลาดเคลื่อนกับเวลา ซึ่งเป็นวิธีที่นิยมใช้เมื่อกรณีมีตัวอย่างเป็นจำนวนมากพอแต่ถ้าตัวอย่างมีไม่มากพอ การเขียนแผนภาพกระจาย อาจจะทำให้เกิดข้อสรุปเกี่ยวกับอัตตสหสัมพันธ์ผิดพลาดได้ ซึ่งในกรณีเช่นนี้จะตรวจสอบโดยใช้วิธีการทดสอบสมมติฐาน และใช้สถิติทดสอบ Dubin-Watson

ก่อนที่จะดำเนินการทดสอบสมมติฐาน พิจารณาตัวแบบการถดถอยเชิงเส้นอย่างง่ายที่ประกอบด้วยตัวแปรอิสระเพียงตัวเดียว โดยมีความคลาดเคลื่อนสุ่ม (Random Error) อยู่ในเทอมของ First-order Autoregressive Process ดังนี้

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

โดยที่

$$\varepsilon_t = \rho\varepsilon_{t-1} + U_t$$

เมื่อ ρ เป็นค่าสัมประสิทธิ์สหสัมพันธ์ในตัวเอง และ $|\rho| < 1$

U_t เป็นตัวแปรสุ่มอิสระ และมีการแจกแจงปกติ (U_t อาจจะเรียกว่า disturbance)

จากสมการถดถอยเชิงเส้นอย่างง่ายจะเห็นว่าในกรณีนี้ค่าความคลาดเคลื่อน ε_t ประกอบด้วยค่าความคลาดเคลื่อนที่เกิดขึ้นในช่วงเวลาที่เกิดขึ้นก่อนหน้าช่วงเวลาปัจจุบัน 1 ช่วงเวลา

(เมื่อ *p > 0*) และเทอม disturbance ได้ว่าในกรณีที่ *p=0* ซึ่งหมายถึงการไม่มีอัตตสหสัมพันธ์ในตัวเอง ค่า $\varepsilon_t = U_t$ มีการแจกแจงปกติที่มีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ σ^2

คุณสมบัติของค่าความคลาดเคลื่อน

ในกรณีที่เกิดปัญหาอัตตสหสัมพันธ์ ตัวแบบถดถอยที่พิจารณาจะเรียกว่า Generalized Regression Models ทั้งนี้เนื่องจากตัวแบบการถดถอยดังกล่าวมีค่าคลาดเคลื่อนซึ่งสัมพันธ์กัน แต่อย่างไรก็ตามค่าความคลาดเคลื่อนยังคงมีคุณสมบัติที่สำคัญ คือ มีค่าเฉลี่ยเท่ากับ 0 และมีค่าความแปรปรวนคงที่ โดยที่

$$E(\varepsilon_t) = 0$$

$$\sigma^2(\varepsilon_t) = \frac{\sigma^2}{1-\rho^2}$$

และ $$Cov(\varepsilon_t \varepsilon_{t-1}) = \rho\left(\frac{\sigma^2}{1-\rho^2}\right)$$

นอกจากนี้เราอาจพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความคลาดเคลื่อนในแต่ละเทอมได้ในรูปของ

$$\rho(\varepsilon_t \varepsilon_{t-1}) = \frac{\rho(\frac{\sigma^2}{1-\sigma^2})}{\left(\sqrt{\frac{\sigma^2}{1-\sigma^2}}\right)\left(\sqrt{\frac{\sigma^2}{1-\sigma^2}}\right)} = \rho$$

นั่นคือค่าสัมประสิทธิ์สัมพันธ์ในตัวเองมีค่าเท่ากับ ρ โดยที่ $|\rho| < 1$

การทดสอบอัตตสหสัมพันธ์โดยใช้สถิติทดสอบ Durbin-Watson

การทดสอบของ Durbin–Watson มีข้อกำหนดที่สำคัญ คือ ตัวแบบการถดถอยจะต้องมีค่าความคลาดเคลื่อนอยู่ในเทอมของ First-order Autoregressive Error Models การทดสอบนี้จะหาข้อสรุปว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความคลาดเคลื่อน (ρ) มีค่าเท่ากับศูนย์หรือไม่ ถ้า $|\rho| < 0$ แสดงว่า $\varepsilon_t = U_t$ นั่นคือ ε_t จะเป็นอิสระต่อกันซึ่งชี้ให้เห็นว่า ค่าความคลาดเคลื่อนไม่เกิดปัญหาอัตตสหสัมพันธ์

โดยทั่วไปปัญหาการเกิดอัตตสหสัมพันธ์ มักจะพบได้บ่อยในข้อมูลเชิงธุรกิจและเศรษฐศาสตร์ ซึ่งมีแนวโน้มเป็นความสัมพันธ์ในลักษณะเชิงบวก (Positive Serial Correlation) โดยมีสมมติฐาน ดังนี้

$H_0 : \rho = 0$

$H_1 : \rho > 0$

สถิติทดสอบ $D = \sum_{t=2}^{n} \frac{(\mathrm{e}_t - e_{t-1})}{e_t^2}$

โดยที่ $e_t = Y_t - \hat{Y}_t$ ซึ่งคำนวณค่า $\hat{Y}_t$ จากสมการถดถอยที่ประมาณได้จากวิธีกำลังสองน้อยที่สุด

การตัดสินใจยอมรับหรือปฏิเสธสมมติฐาน จะพิจารณาจากค่า d_L และ d_U ซึ่งเป็นค่าขอบเขตล่างและขอบเขตบนที่เปิดได้จากตารางเขตวิกฤตของ Durbin-Watson ตามจำนวนตัวแปรอิสระในสมการถดถอย

ถ้า $D > d_U$ ตัดสินใจยอมรับ H_0

ถ้า $D < d_U$ ตัดสินใจยอมรับ H_1 (ปฏิเสธ H_0)

ถ้า $d_L < D < d_U$ ไม่สามารถสรุปผลการทดสอบได้

จะเห็นได้ว่าการที่ค่า D ที่คำนวณได้มีค่าน้อย จะนำไปสู่การปฏิเสธ H_0 แสดงว่าเรายอมรับว่า $\rho > 0$ นั่นคือค่า ε_t และ ε_{t-1} มีความคล้ายคลึงกันจะส่งผลให้ผลต่างของ e_t กับ e_{t-1} มีค่าน้อย เมื่อ $\rho > 0$

4. การแก้ไขปัญหากรณีเกิดอัตตสหสัมพันธ์ (Remedial Measures for Autocorrelation)

โดยทั่วไปเมื่อเกิดปัญหาการมีอัตตสหสัมพันธ์ในค่าความคลาดเคลื่อนต้องมีการปรับแก้เสียก่อนที่จะนำสมการถดถอยที่ได้ไปใช้ในการพยากรณ์ การแก้ไขปัญหาอาจจะกระทำได้ 2 วิธี คือ การเพิ่มตัวแปรอิสระเข้าไปในตัวแบบการถดถอย และการแปลงตัวแปร (Transformed Variables) ซึ่งมี 3 วิธี ที่เป็นที่นิยมใช้กันอย่างแพร่หลาย คือ Cochran-Orcutt Procedure, Hildreth-Lu Procedure และ First Difference Procedure ซึ่งในที่นี้จะกล่าวเฉพาะรายละเอียดของวิธี Cochran-Orcutt เท่านั้น ส่วนรายละเอียดของวิธีอื่น ๆ ศึกษาได้จาก Applied Linear Statistical Model ของ Neter et al.

4.1 การเพิ่มตัวแปรอิสระเข้าไปในตัวแบบการถดถอย (Addition of Predictor Variables to Regression Model)

ตามที่ได้กล่าวมาแล้วในตอนต้นว่าการเกิดอัตตสหสัมพันธ์ของค่าความคลาดเคลื่อนมีสาเหตุสำคัญอันหนึ่ง คือ การไม่ได้นำตัวแปรอิสระที่สำคัญเข้าไปในแบบถดถอย ดังนั้นเมื่อเกิดปัญหาอัตตสหสัมพันธ์ขึ้น สิ่งที่ควรพิจารณาเป็นอันดับแรกคือ การพิจารณาว่าตัวแบบการถดถอย

ขาดตัวแปรอิสระที่สำคัญตัวใดบ้าง ซึ่งในบางกรณีเราอาจพิจารณาตัวแปรหุ่น (Dummy Variable) สำหรับอิทธิพลของฤดูกาลที่อาจจะมีต่อตัวแปรตามที่เราสนใจ เข้าไปในตัวแบบถดถอย เช่น ในการพยากรณ์ยอดขายสินค้าตามระยะเวลาที่เกิดขึ้น กรณีเช่นนี้ฤดูกาลอาจจะมีผลต่อยอดขายด้วยนอกเหนือจากตัวแปรอิสระที่เราได้พิจารณานำเข้ามาในตัวแบบการถดถอยแล้ว

4.2 การแปลงตัวแปร (Use of Transformed Variable)

กรณีการแก้ปัญหาอัตตสหสัมพันธ์โดยการเพิ่มตัวแปรอิสระเข้าไปในตัวแบบแล้วพบว่ายังคงเกิดปัญหาอัตตสหสัมพันธ์ ในกรณีเช่นนี้เราอาจจะใช้วิธีการแปลงตัวแปร ซึ่งมี 3 วิธี ที่สำคัญดังที่ได้กล่าวไว้ในตอนต้นของหัวข้อ

การแก้ปัญหาโดยใช้ 3 วิธีดังกล่าวข้างต้นอยู่บนพื้นฐานของตัวแบบการถดถอยที่อยู่ในรูปของ First-order Autoregressive Error Terms ซึ่งในที่นี้จะพิจารณาจากตัวแบบการถดถอยเชิงเส้นอย่างง่าย

จากตัวแบบการถดถอย $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$

$$\varepsilon_t = \rho\varepsilon_{t-1} + U_t$$

พิจารณาการแปลงตัวแปรโดยกำหนดให้

$$Y_t' = Y_t - \rho Y_{t-1}$$

เนื่องจาก $Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$ และ

$$Y_{t-1} = \beta_0 + \beta_1 X_t + \varepsilon_{t-1}$$

ดังนั้น $Y_t' = (\beta_0 + \beta_1 \mathrm{X}_t + \varepsilon_t) - \rho(\beta_0 + \beta_1 \mathrm{X}_t + \varepsilon_{t-1})$

$$= \beta(1-\rho) + \beta_1(\mathrm{X}_t - \rho X_{t-1}) + (\varepsilon_t + \rho\varepsilon_{t-1})$$

แต่ $U_t = \varepsilon_t - \rho\varepsilon_{t-1}$

ดังนั้น $Y_t' = \beta(1-\rho) + \beta_1(\mathrm{X}_t - \rho X_{t-1}) + U_t$

จะเห็นได้ว่าเมื่อเราใช้วิธีการแปลงตัวแปร Y_t' แล้วตัวแบบการถดถอยที่ได้ภายหลังการแปลงตัวแปรมีค่าความคลาดเคลื่อนที่เป็นอิสระต่อกัน กล่าวคือ $U_t \sim N(0,\sigma^2)$ และสมการถดถอยที่ได้ดังกล่าวยังคงเป็นสมการถดถอยเชิงเส้นอย่างง่าย โดยที่สามารถเขียนในรูปของ

$$Y_t' = \beta_0' + \beta_1' X_t' + \varepsilon_t$$

โดยที่ $Y_t' = Y_t - \rho Y_{t-1}$

$$X_t' = X_t - \rho X_{t-1}$$

$$\beta_0' = \beta_0(1-\rho)$$

และ $\beta_1' = \beta_1$

จากสมการที่ได้ภายหลังจากการแปลงตัวแปร เราต้องประมาณค่า ρ ซึ่งเป็นพารามิเตอร์ที่ไม่ทราบค่า ด้วยค่า r ซึ่งเป็นค่าประมาณของ ρ อยู่ในรูปของ

$$r = \frac{\sum_{t=2}^{n} e_{t-1}e_t}{\sum_{t=2}^{n} e_{t-1}^2} \qquad \text{โดยที่} \quad |r| < 1$$

จะได้ว่า $Y_t' = Y_t - rY_{t-1}$

$$X_t' = X_t - rX_{t-1}$$

ดังนั้นสมการถดถอยของตัวอย่าง จึงอยู่ในรูปของ

$$\hat{Y}_t = b_0' + b_1' X_t$$

เมื่อขจัดปัญหาการเกิดอัตตสหสัมพันธ์ในการวิเคราะห์แล้ว จะทำการแปลงสมการดังกล่าวกลับไปเป็นตัวแบบถดถอยเริ่มต้น โดยมีสมการถดถอยเป็น ดังนี้

$$\hat{Y}_t = b_0 + b_1 X_t$$

โดยที่ $b_0' = b_0(1-\mathrm{r})$

ดังนั้น $b_0 = \dfrac{b_0'}{1-r}$

และ $b_t = b_1'$

นอกจากนี้ยังอาจประมาณค่าเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์การถดถอยของตัวแบบการถดถอยเริ่มต้นได้ ดังนี้

$$S(\mathrm{b}_0) = \frac{S(\mathrm{b}_0')}{1-r}$$

$$S(\mathrm{b}_1) = S(\mathrm{b}_1')$$

ที่ได้กล่าวมาทั้งหมดนี้เป็นทฤษฎีพื้นฐานของการใช้วิธีการแปลงตัวแปร ซึ่งหลักการที่สำคัญของการแปลงตัวแปร คือ การประมาณค่าสัมประสิทธิ์อัตตสหสัมพันธ์ (ρ) ด้วยค่า r ต่อจากนั้นจึงดำเนินการด้วยวิธีการต่อไปนี้ คือ วิธี Cochran-Orcutt Procedure

การขจัดปัญหาการเกิดอัตตสหสัมพันธ์ในค่าความคลาดเคลื่อน โดยวิธีของ Cochran-Orcutt เป็นวิธีหนึ่งที่อาศัยหลักการแปลงตัวแปร และเป็นที่นิยมใช้กันอย่างแพร่หลาย โดยมีหลักการที่สำคัญดังนี้ คือ

1. ประมาณค่า ρ ด้วยค่า r ตามสูตรที่ได้กล่าวมาแล้วข้างต้น
2. สร้างสมการถดถอยด้วยวิธีกำลังสองน้อยที่สุด โดยใช้ตัวแปรอิสระและตัวแปรตามที่ได้จากการแปลงตัวแปร กล่าวคือ

$$Y_t' = Y_t - rY_{t-1}$$

และ $X_t' = X_t - rX_{t-1}$

3. ทดสอบค่าความคลาดเคลื่อนที่ได้จาก 2) โดยใช้สถิติทดสอบของ Durbin-Watson ว่าเกิดปัญหาอัตตสหสัมพันธ์หรือไม่ ถ้าไม่เกิดปัญหาอัตตสหสัมพันธ์ถือว่าปัญหาดังกล่าวได้ถูกขจัดไปแล้ว

4. แปลงสมการที่ได้กลับไปเป็นสมการถดถอยเริ่มต้น เพื่อนำไปใช้ในการพยากรณ์ต่อไป

ในกรณีที่เราแปลงข้อมูลแล้ว พบว่าปัญหาอัตตสหสัมพันธ์ยังคงอยู่ ในกรณีเช่นนี้เราต้องกระทำซ้ำในขั้นตอนที่ 1-3 อีกครั้ง จนกว่าปัญหาอัตตสหสัมพันธ์จะถูกขจัดไป เป็นที่น่าสังเกตว่าการกระทำวนซ้ำโดยวิธีของ Cochran-Orcutt Procedure เราจะกระทำวนซ้ำเพียง 1-2 ครั้งเท่านั้น ถ้าการวนซ้ำเกิดหลายครั้ง วิธีของ Cochran-Orcutt Procedure จะขาดความแม่นยำ จึงควรใช้วิธีอื่นที่มีความเหมาะสมกว่า เช่น วิธี First-Difference Procedure เป็นต้น

5. การพยากรณ์กรณีค่าความคลาดเคลื่อนมีความสัมพันธ์กัน

ได้กล่าวมาแล้วว่าวัตถุประสงค์ที่สำคัญของการวิเคราะห์ด้วยการถดถอย คือ การนำสมการถดถอยที่ได้ไปใช้ในการพยากรณ์ ในหัวข้อนี้จะกล่าวถึงการพยากรณ์ค่าตัวแปรตามเมื่อข้อมูลเกิดอัตตสหสัมพันธ์ในค่าความคลาดเคลื่อน โดยใช้สมการถดถอยเชิงเส้นอย่างง่าย และได้มีการใช้วิธีการแปลงข้อมูลโดยวิธี Cochran-Orcutt เพื่อประมาณค่าพารามิเตอร์

พิจารณาสมการการถดถอย

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

โดยที่

$$\varepsilon_t = \rho\varepsilon_{t-1} + U_t$$

และ

$$Y_t = \beta_0 - \beta_1 X_t - \rho\varepsilon_{t-1} + U_t$$

เมื่อต้องการพยากรณ์ข้อมูลอนุกรมเวลา ณ เวลา $t+1$ จะได้ว่า

$$Y_{t+1} = (\beta_0 - \beta_1 X_{t+1}) - \rho\varepsilon_t + U_{t+1}$$

จะเห็นได้ว่าการพยากรณ์ ณ เวลา $t+1$ ประกอบด้วย 3 ส่วน คือ

1. ค่าคาดหวังของ $\beta_0 + \beta_1 X_{t+1}$

2. ผลคูณของ ρ กับ ε_t

3. U_{t+1} โดยที่ $E(\mathrm{U}_{t+1}) = 0$

ดังนั้นในการพยากรณ์จึงต้องพิจารณาส่วนประกอบทั้ง 3 ส่วน ถ้าให้ F_{t+1} เป็นค่าพยากรณ์ ณ เวลา $t+1$ และเมื่อพิจารณาส่วนประกอบ 3 ส่วนในการพยากรณ์ที่กล่าวมา จะได้ว่า

$$\hat{Y}_{t+1} = b_0 + b_1 X_{t+1}$$

โดยที่

1. b_0, b_1 เป็นค่าประมาณของค่าสัมประสิทธิ์การถดถอยที่ได้จากการแปลง b_0' และ b_1'
2. r เป็นตัวประมาณของ r และ ε_t ประมาณได้ด้วย e_t

เมื่อ $$e_t = Y_t - (\mathrm{b}_0 + \mathrm{b}_1 \mathrm{X}_t)$$
$$= Y_t - \hat{Y}_t$$

3. $U_{t+1} \sim N(0, \sigma^2)$

ดังนั้นในการพยากรณ์ ณ เวลา $t+1$ จะอยู่ในรูปของ

$$F_{t+1} = \hat{Y}_{t+1} + re_t$$

ตอนที่ 6 การจำลองสถานการณ์ด้วยวิธีการมอนติคาร์โล

1. ความหมายของวิธีมอนติคาร์โล

เทคนิคมอนติคาร์โล เป็นส่วนหนึ่งของกระบวนการของชุดคำสั่งในการคำนวณที่ทำซ้ำ ๆ ของชุดตัวเลขเชิงสุ่ม เพื่อหาผลลัพธ์ของชุดตัวเลขนั้น ๆ เทคนิคมอนติคาร์โล มักใช้ในงานที่มีลักษณะการจำลองทางกายภาพ และทางคณิตศาสตร์ และเป็นการใช้ตัวเลขที่สุ่มมาด้วยเครื่องคอมพิวเตอร์ ซึ่งเรียกว่าตัวเลขสุ่มเทียม (Pseudo-Random Number) (Merriam–Webster, 1994, pp. 754–755)

เทคนิคมอนติคาร์โล เป็นรูปแบบการสร้างค่าเชิงสุ่มเพื่อการศึกษาในเชิงสถิติ ซึ่งเป็นค่าเชิงสุ่มที่มักจะได้จากการใช้เครื่องคอมพิวเตอร์ในการสร้างขึ้นมา เป็นเทคนิคที่เกี่ยวข้องกับชุดข้อมูลที่สร้างขึ้นมาหลากหลายชุดเพื่อใช้ในการศึกษาคุณลักษณะของตัวประมาณ เช่น ในกรณีที่นักสถิติที่ข้อมูลที่เกิดขึ้นในสถานการณ์จริงมาเพื่อทดสอบ และต่อมาสถานการณ์นั้น หรือสภาพแวดล้อมนั้น ๆ อาจจะเป็นเปลี่ยนไปในเวลาต่อมา นอกจากนั้นการเก็บข้อมูลจริงจำนวนมาก ๆ เพื่อทดสอบทฤษฎีจำเป็นที่ต้องใช้ต้นทุนสูงมาก ดังนั้นเพื่อทดสอบทฤษฎีจึงใช้วิธีการสร้างชุดข้อมูลตามสถานการณ์นั้น ๆ โดยใช้โปรแกรมอาร์ (Vogt, 2005, p. 196)

2. ความเป็นมาของวิธีมอนติคาร์โล

การจำลองสถานการณ์ด้วยคอมพิวเตอร์ จะต้องมีการคำนวณ ซึ่งมีข้อมูลทั้งที่เป็นข้อมูลนำเข้าและผลลัพธ์จากแบบจำลอง ดังนั้นการจัดเตรียมและการวิเคราะห์ข้อมูลรวมทั้งขั้นตอนต่าง ๆ ที่ใช้ในการจำลองสถานการณ์ จึงต้องอาศัยวิธีการต่าง ๆ ทางสถิติเข้าช่วย และวิธีการทางสถิติวิธีการหนึ่งซึ่งเป็นที่นิยมใช้กันมากและเกือบจะมีความจำเป็นในทุก ๆ การจำลองสถานการณ์ก็คือ การสุ่มตัวอย่างด้วยเทคนิคมอนติคาร์โล (Monte Carlo Sampling Technique) และเพื่อเน้นถึงความจำเป็นในการใช้เทคนิคดังกล่าว การจำลองสถานการณ์นี้จึงถูกเรียกว่า การจำลองสถานการณ์ด้วยเทคนิคมอนติคาร์โล ดังนั้น เทคนิคมอนติคาร์โล จึงเป็นวิธีหนึ่งของคณิตศาสตร์ที่ใช้คอมพิวเตอร์ช่วยในการจำลองสถานการณ์ (Simulation) โดยอาศัยตัวเลขสุ่ม (Random Number) มาสร้างตัวแปรให้เหมือนกับสถานการณ์จริง และมีการทดลองซ้ำหลาย ๆ ครั้ง เพื่อให้ได้ค่าที่แน่นอนที่ใช้เป็นข้อสรุปหรืออธิบายปรากฏการณ์ต่าง ๆ ในสถานการณ์จริง หรือช่วยหาคำตอบในเรื่องราวต่าง ๆ ที่ยังไม่แน่ใจในผลที่จะเกิดขึ้น (Hammersley & Handcomd, 1965)

เทคนิควิธีมอนติคาร์โล ได้มีการใช้มานานแล้ว เพียงแต่ไม่ได้เรียกว่า มอนติคาร์โล ต่อมาในราวปี ค.ศ. 1753 จอร์ส หลุยส์ เลคเลอร์ (Georges Lois Lectere & Comte de Buffon) ได้นำมาพัฒนาในทฤษฎีความน่าจะเป็นโดยการทดลองหาค่า π โดยการโยนเข็มที่มีความยาว k หน่วยอย่างสุ่มลงบนพื้นราบที่มีเส้นขนานอยู่ โดยให้ระยะห่างระหว่างเส้นขนานแต่ละเส้นห่างกัน d หน่วย และกำหนดให้ d มากกว่า k จะได้ความน่าจะเป็นที่เข็มจะตัดเส้นขนาน $P = \dfrac{2k}{d\pi}$ ซึ่งถ้าความน่าเป็น (p) เป็นค่าสุ่มก็จะหาค่า π ได้ ต่อมาในปี ค.ศ. 1980 กอสเซท (Gooset) ได้ศึกษาการแจกแจงความถี่ของความสูงของนักโทษอาชญากรรมจำนวน 3,000 คน โดยเทียบกับการแจกแจงความถี่ของกลุ่มตัวอย่างที่สุ่มมาครั้งละ 4 คน จำนวน 750 กลุ่มตัวอย่าง ผลการศึกษาพบว่าการแจกแจงความถี่ทั้งสองลักษณะเหมือนกัน กอสเซท ได้ตั้งชื่อการแจกแจงความถี่ที่ค้นพบนี้ว่า การแจกแจงค่าที (t–distribution) ซึ่งถือว่าเป็นจุดเริ่มต้นของเทคนิควิธีมอนติคาร์โล (Monte Carlo Method)

เทคนิคมอนติคาร์โล ได้รับการพัฒนาอย่างจริงจังในราวปี ค.ศ. 1944 ช่วงสงครามโลกครั้งที่ 2 อูลาม และออน นิวแมน (Ulam, & Von Neumann) เป็นผู้ตั้งชื่อ มอนติคาร์โล ซึ่งเป็นรหัสลับของงานที่ทำใน ลอส อาลามาส (Los Alamas) และได้นำเทคนิคนี้มาหาผลของการแพร่อย่างสุ่มของนิวตรอนในวัสดุเชื้อเพลิงที่เป็นการทดลองทางคณิตศาสตร์เพื่อหาผลของคำตอบก่อนที่จะทำการทดลองจริงซึ่งเป็นการหลีกเลี่ยงอันตรายและช่วยประหยัดค่าใช้จ่ายก่อนที่จะทดลองจริง หลังจากนั้นเทคนิคมอนติคาร์โล จึงมีการนำมาใช้อย่างกว้างขวางทั้งทางด้านฟิสิกส์ คณิตศาสตร์

สถิติ และการวิจัย นับได้ว่าเทคนิคมอนติคาร์โล มีประโยชน์อย่างมากในการขยายความรู้เชิงทฤษฎี (Robert & Casella, 2004, p. 696) แม้ว่าวิธีมอนติคาร์โลเป็นวิธีที่เกิดขึ้นนานมาแล้ว แต่เพิ่งจะมีการนำมาใช้ในการแก้ปัญหาจริงที่มีความซับซ้อนเมื่อไม่นานมานี้ เนื่องจากเทคโนโลยีด้านคอมพิวเตอร์มีความก้าวหน้าขึ้น จึงเอื้อต่อการใช้วิธีมอนติคาร์โล มากขึ้น วิธีมอนติคาร์โล แตกต่างจากวิธีการทาง Numerical Method ทั่วไปตรงที่ วิธีการทาง Numerical Method ทั่ว ๆ ไป จะเริ่มต้นในการแก้ปัญหาด้วยการสร้างแบบจำลองแล้วใช้ วิธีการทางคณิตศาสตร์ในการหาคำตอบของสมการแต่วิธีการมอนติคาร์โล อาจจะไม่จำเป็นต้องสร้างแบบจำลองทางคณิตศาสตร์ขึ้นมา อาจใช้วิธีจำลองพฤติกรรมของระบบขึ้นมาโดยตรง ซึ่งเป็นกระบวนการแบบสุ่ม (Stochastic Process) เหมือนอย่างเหตุการณ์ที่เกิดขึ้นในโลกความเป็นจริง การจำลองสถานการณ์สามารถทำซ้ำได้หลาย ๆ ครั้ง หลังจากนั้นจะทำการหาความแปรปรวนของคำตอบ และเพิ่มจำนวนการจำลองสถานการณ์เพื่อทำให้คำตอบมีความแปรปรวนลดลง อยู่ในขอบเขตตามที่ต้องการ การสร้างตัวแปรสุ่ม (Random Variables) ในกระบวนการของมอนติคาร์โล จะสร้างจาก Probability Density Function (ZP.D.F.) แบบต่าง ๆ ตามที่ต้องการ ตัวแปรสุ่มที่สร้างขึ้นจะเปรียบเสมือนข้อมูลที่เก็บได้จากโลกความจริงการสร้างตัวเลขสุ่มในทางปฏิบัติเดิม จะใช้ตารางตัวเลขสุ่ม แต่ในปัจจุบันเราสามารถใช้เครื่องคอมพิวเตอร์ซึ่งสามารถทำงานได้อย่างสะดวกขึ้นและรวดเร็ว โดยสรุปขั้นตอนการสร้างแบบจำลองสถานการณ์วิธีมอนติคาร์โล มีดังนี้ คือ

1. กำหนดปัญหาหรือระบบในสิ่งที่สนใจจะทำการจำลอง
2. ระบุองค์ประกอบของความไม่แน่นอนในปัญหานั้น
3. สร้างตารางแสดงการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่ต้องการ สำหรับใช้ในการจำลอง (พิจารณาจากข้อมูลที่ไปสำรวจหรือสังเกตมา)
4. หาการแจกแจงความน่าจะเป็น (Probability Distribution) ขององค์ประกอบที่มีความไม่แน่นอน
5. กำหนดค่าตัวเลขสุ่ม (Random Number) ที่ต้องใช้กับตัวแปรสุ่มให้สอดคล้องกับความน่าจะเป็นของตัวแปรสุ่ม
6. สร้างตัวแบบการจำลองทางคณิตศาสตร์ให้เข้ากับปัญหาตามวัตถุประสงค์ที่ตั้งไว้ ทำการทดสอบตัวแบบดังกล่าวว่าได้ผลตามเป้าหมายที่วางไว้หรือไม่
7. เมื่อผลการทดสอบเป็นไปตามเป้าหมายแล้วจะกำหนดจำนวนครั้งในการจำลอง
8. ทำการจำลองเพื่อหาค่าเฉลี่ยที่ต้องการ

คุณลักษณะสำคัญของแบบจำลองสถานการณ์

แบบจำลองสถานการณ์มีคุณลักษณะที่แตกต่างไปจากแบบจำลองชนิดอื่น ๆ ดังนี้ (กิตติ ภักดีวัฒนะกุล และพนิดา พานิชกุล, 2554, หน้า 252)

1. มีการตรวจสอบความถูกต้อง เพื่อไม่ให้เกิดข้อผิดพลาดในการนำค่าการทำนายไปใช้ประโยชน์

2. มีเหตุผลเป็นการตรวจสอบว่าผลที่ได้ต้องอยู่ในขอบเขตของผลลัพธ์ที่คาดคะเนไว้และแบบจำลองนั้นทำงานอย่างถูกต้องโดยสามารถนำผลลัพธ์นั้นมาวิเคราะห์ได้

3. ลดความเบี่ยงเบน โดยใช้ค่าสุ่มเดียวกันเพื่อลดความแปรผันและเพิ่มความถูกต้องเมื่อเปรียบเทียบกับองค์ประกอบที่ต่างกันได้

4. มีลักษณะเป็นการเลียนแบบสถานการณ์จริงมากกว่าเป็นการนำเสนอสถานการณ์จริง มีลักษณะเป็นการบรรยายหรือการคาดการณ์จริงที่จะเกิดขึ้นภายใต้เงื่อนไขต่าง ๆ กัน

5. เป็นแบบจำลองที่ใช้กับปัญหาที่มีความซับซ้อนสูง

การสร้างแบบจำลองสถานการณ์

การสร้างแบบจำลองสถานการณ์รวมถึงการสร้างแบบจำลองของระบบขึ้นมาจริงและกระทำการสร้าง แล้วทดลองซ้ำหลาย ๆ ครั้ง การสร้างแบ่งออกเป็น 7 ขั้นตอน ดังนี้ (กิตติ ภักดีวัฒนะกุล และพนิดา พานิชกุล, 2554, หน้า 254)

1. $E(b) = \beta$ นิยามปัญหา (Problem Definition) ต้องตรวจสอบปัญหาที่เกิดขึ้นและจัดแบ่งให้เป็นหมวดหมู่ นอกจากนี้ต้องมีการกำหนดขอบเขตของระบบ รวมทั้งจะต้องปรับให้รูปการของปัญหามีความชัดเจนและเข้าใจได้ง่ายขึ้น

2. สร้างแบบจำลองสถานการณ์ (Simulation Model Construction) กำหนดค่าตัวแปรและความสัมพันธ์ของตัวแปรเหล่านั้น รวมถึงการรวบรวมข้อมูลที่จำเป็นต่อการสร้างแบบจำลองสถานการณ์ ในขั้นตอนนี้มักนำ Flowchart มาใช้เพื่ออธิบายกระบวนการ จากนั้นจึงเขียนโปรแกรมตาม Flowchart ที่สร้างขึ้น

3. ทดสอบและตรวจสอบความถูกต้อง (Model Testing and Validation) เนื่องจากแบบจำลองสถานการณ์ จะต้องถูกนำไปศึกษาแทนเหตุการณ์จริง ดังนั้นจะต้องทดสอบ และค้นหาสิ่งผิดพลาดทั้งหมดเพื่อให้มั่นใจได้ว่าสามารถนำไปใช้แทนเหตุการณ์จริงได้อย่างสมบูรณ์

4. ออกแบบสถานการณ์เพื่อการทดลอง (Experimental Design) หลังจากที่แบบจำลองได้รับการพิสูจน์แล้ว จะออกแบบการทดลองเพื่อทำการจำลองสถานการณ์ขั้น 3 กรณี ได้แก่ Best-Case กรณี Worst-Case และกรณี Median–Case ซึ่งการทำเช่นนี้จะช่วยให้ผู้ตัดสินใจสามารถ

กำหนดขอบเขตของตัวแปรที่ใช้ในการทำงานของ Simulation ได้ และยังช่วยในการแก้ไขจุดบกพร่องในแบบจำลองสถานการณ์ที่สร้างเสร็จแล้วด้วย

5. การควบคุมการทดลอง (Experimental Conduction) เป็นการทดลองใส่ค่าตัวแปรจริงในแบบจำลองเพื่อแสดงสถานการณ์ตามตัวแปรที่ทดลองเปลี่ยนไป แล้วนำเสนอผลลัพธ์ออกมา

6. การประเมินผลลัพธ์จากการทดลอง (Result Evaluation) หากเป็นที่น่าพอใจจะนำไปใช้แก้ปัญหาทันที แต่ถ้าไม่พอใจอาจย้อนกลับไปปรับปรุงหรือสร้างแบบจำลองสถานการณ์ใหม่อีกครั้ง

7. การนำไปใช้แก้ปัญหาจริง (Implementation) นำไปใช้ได้เช่นเดียวกับแบบจำลองชนิดอื่น แต่จะดีกว่าตรงที่ผู้บริหารนั้นสามารถเลือกดูสถานการณ์ได้มากกว่า

3. ขั้นตอนของระเบียบเทคนิควิธีมอนติคาร์โล

หลักการสำคัญของเทคนิคมอนติคาร์โล (Monte Carlo Method) ก็คือ การนำเอาตัวเลขสุ่ม (Random Number) มาประยุกต์ในการแก้ปัญหาต่าง ๆ มีขั้นตอนที่สำคัญ ดังนี้

3.1 สร้างตัวเลขสุ่ม (Generate Random Number)

ในระยะแรก ๆ กระทำโดยการอาศัยเครื่องมือทางกายภาพ เช่น ล้อรูเล็ต ลูกเต๋า ไพ่กระดาษเขียนเบอร์ เป็นต้น เครื่องมือดังกล่าวใช้ได้เมื่อต้องการใช้ตัวเลขสุ่มจำนวนไม่มากนัก ต่อมาเมื่อมีความต้องการใช้ตัวเลขสุ่มจำนวนมาก ๆ ก็มีการหันมาใช้เครื่องมืออิเล็คทรอนิคส์ที่ใช้แพร่หลายที่สุด คือ เครื่องสร้างตัวเลขแบบสุ่มที่สร้างขึ้นโดย บริษัท แรนด์ (RAND) ที่ได้ตัวเลขแบบสุ่มจากเครื่องกำเนิดพัสส์อิเล็คทรอนิคส์ (Electromnic Pulse Generator) ซึ่งทำงานด้วยเสียงเครื่องดังกล่าวสามารถสร้างตัวเลขสุ่มได้เป็นล้านตัว ตัวเลขแบบสุ่มของบริษัท RAND นี้ มีการตีพิมพ์และลงไว้ในเทปแม่เหล็กเพื่อจำหน่าย

การสร้างหรือเลือกใช้ตัวเลขสุ่มกับเครื่องคอมพิวเตอร์นี้ มีปัญหา 2 ประการ คือ

1. เป็นการยากที่จะทำให้คอมพิวเตอร์สามารถเรียกใช้เมื่อมีความต้องการ และ2. เป็นการยากที่จะทำให้เครื่องมือดังกล่าว สร้างตัวเลขสุ่มชุดเดิม เมื่อต้องการใช้เปรียบเทียบวิธีการต่าง ๆ ภายใต้เงื่อนไขของระบบเลขสุ่มชุดเดียวกัน หรือถ้าจะเก็บเลขสุ่มเหล่านี้ไว้ในหน่วยความจำหรือจานแม่เหล็ก จะทำให้สูญเสียหน่วยความจำหรือเสียเวลาในการค้นหา ฉะนั้นการสร้างตัวเลขสุ่มในคอมพิวเตอร์จึงนิยมสร้างตัวเลขแบบสุ่มเทียมโดยอาศัยสูตรทางคณิตศาสตร์ วิธีที่ใช้กันมากมี 2 วิธีคือ

วิธีที่ 1 วิธีส่วนกลางกำลังสอง (Midsquare Method) เป็นวิธีใช้ในยุคแรก ๆ ของการสร้างตัวเลขสุ่มเทียม โดยมีขั้นตอนในการสร้างตัวเลขแบบสุ่ม ดังนี้ (ศิริรัตน์ วงศ์ปกรณ์กุล, 2539, หน้า 240-246)

1. เลือกตัวเลขขึ้นมาสี่หลัก

2. ยกกำลังสองของตัวเลขนั้น ถ้าตัวเลขที่ได้ไม่ครบแปดหลักให้เติมศูนย์ข้างหน้าให้ครบแปดหลัก

3. ใช้เลขที่หลักกลางที่ได้ในขั้นที่ 2 เป็นตัวเลขแบบสุ่ม

4. ยกกำลังสองของตัวเลขในข้อ 3

5. ทำซ้ำในขั้นที่ 3 และ 4 จะได้จำนวนตัวเลขสุ่มตามต้องการ

วิธีที่ 2 วิธีเศษเหลือ (Congruent Method) วิธีเศษเหลือที่นิยมใช้กันมากที่สุด คือ เศษเหลือของผลคูณ (Multiplicative Congruent Method) ซึ่งมีสูตร ดังนี้

$$X_{n+1} = ax_n(\mathrm{Mod})\,\mathrm{m}$$

โดยที่ a และ m ต้องเป็นตัวเลขที่ไม่เป็นค่าลบ การสร้างตัวเลขแบบสุ่มนี้เริ่มต้นด้วยตัวเลขเริ่มต้น ตัวเลขตัวต่อไปจะได้จากการคูณด้วยค่าคงที่ a และหารด้วยเศษเหลือจากการหาร คือ ตัวเลขที่ต้องการ

ในปัจจุบันมีการใช้วิธีการอีกหลายอย่างในการสร้างตัวเลขแบบสุ่มเทียม รวมทั้งมีโปรแกรมสำเร็จรูปให้เลือกใช้ โดยผู้ใช้ไม่ต้องสร้างโปรมแกรมเองแต่วิธีการและโปรแกรมที่ใช้ในการสร้างตัวเลขแบบสุ่มนั้น ต้องมีคุณสมบัติของตัวเลขสุ่มที่ดี ดังนี้

1. ตัวเลขสุ่มที่ได้จะต้องมีลักษณะของการกระจายความน่าจะเป็นแบบสม่ำเสมอ (Uniform Distribution)

2. ตัวเลขสุ่มที่ได้ต้องเป็นอิสระแก่กัน

3. อนุกรมของตัวเลขสุ่มที่ได้ต้องสามารถสร้างซ้ำเดิมได้

4. อนุกรมของตัวเลขสุ่มที่ได้ต้องไม่ซ้ำเดิมในช่วงที่ต้องการใช้ตัวเลขสุ่ม

5. ต้องใช้เวลาน้อยในการสร้างตัวเลขสุ่ม

6. ต้องใช้หน่วยความจำในคอมพิวเตอร์น้อย

3.2 การนำตัวเลขสุ่มมาประยุกต์ใช้กับปัญหาต่าง ๆ

เป็นข้อมูลของปัญหานั้น Swarup et al. (2001) เช่น สร้างตัวเลขสุ่มขึ้นมาจำนวนหนึ่งแล้วนำเลขสุ่มนั้นไปสร้างเป็นคะแนนผลการสอบของผู้เรียน บางครั้งตัวแปรของปัญหาที่จะศึกษาไม่ได้สร้างจากตัวเลขสุ่มโดยตรง แต่มีขั้นตอนที่ต้องอาศัยตัวเลขสุ่มเป็นพื้นฐานก็ได้

3.3 ทำการทดลองซ้ำหลาย ๆ ครั้ง

หลักการสำคัญประการหนึ่งของเทคนิควิธีมอนติคาร์โล ก็คือต้องมีการทดลองซ้ำหลาย ๆ ครั้ง เพื่อลดความคลาดเคลื่อนของคำตอบที่จะได้ (Hammersley & Handscomb, 1964) และสามารถสรุปเป็นความน่าจะเป็นของการเกิดเหตุการณ์ในปัญหานั้น ๆ

4 จุดเด่นของเทคนิควิธีมอนติคาร์โล

เทคนิคมอนติคาร์โล เป็นการใช้ตัวเลขสุ่มเป็นพื้นฐานการสร้างตัวแปรของปัญหาโดยอาศัยทฤษฎี สูตรหรือกฎเกณฑ์ต่าง ๆ ที่มีอยู่ และมีการทดลองซ้ำหลาย ๆ ครั้ง เพื่อลดความคลาดเคลื่อนต่าง ๆ ซึ่งนับว่ามีประโยชน์ที่สำคัญดังนี้

4.1. เทคนิควิธีมอนติคาร์โล สามารถควบคุมตัวแปรแทรกซ้อนและสามารถสังเกตได้อย่างสมบูรณ์ (Completely Controled and Completely Observed) (Payne et al., 1988) นอกจากนี้ยังสามารถทำการทดลองซ้ำภายใต้สภาพแวดล้อมเดิมหลาย ๆ ครั้งได้ ซึ่งในการทดลองจริงนั้นทำไม่ได้ เพราะไม่สามารถรักษาสภาพแวดล้อมให้เหมือนเดิมทุกอย่างได้ เมื่อเวลาเปลี่ยนไป

4.2 ในการใช้เทคนิควิธีมอนติคาร์โล ถ้ามีทฤษฎี สูตร หรือกฎเกณฑ์ต่าง ๆ ที่ถูกต้องรองรับในการสร้างตัวแปรของปัญหาในการทดลองแล้ว จะทำให้ผลที่ได้ถูกต้องแม่นยำกว่าเมื่อใช้ทดลองในสถานการณ์จริง ทั้งนี้เพราะสามารถลดตัวแปรแทรกซ้อนในเชิงจิตวิทยาได้

4.3 ลดการสิ้นเปลืองเวลา แรงงาน และค่าใช้จ่ายน้อยกว่า เมื่อเทียบกับการทดลองในสถานการณ์จริง

ตอนที่ 7 งานวิจัยที่เกี่ยวข้องกับ Reproducing Kernel Hilbert Space

Migule Piera, Emmanuel, and Fleury (2002) ได้ศึกษาการประยุกต์ทางวิศวกรรม ตัวอย่างของข้อมูลที่ได้รับมีคุณค่ามากและมีจำกัดในด้านของจำนวนเช่นในเรื่องเนื้อหาที่แตกต่างกัน วิธีการจำแนกประเภทข้อมูลที่ตั้งอยู่บนพื้นฐานของ RKHS ที่สามารถใช้ได้ในฟังก์ชั่น EVT เพื่อที่จะประมาณค่าของควอไทล์ที่หลากหลายและความน่าจะเป็นที่ลดน้อยลง สำหรับการประมาณค่าของควอไทล์ที่หลากหลาย โดยการใช้วิธีการของ RKHS ในการจำแนกประเภทที่เป็นไปได้ที่จะสร้างแผนของเป็นข้อมูลเวกเตอร์บนโปรแกรม R ในการประมาณค่าควอไทล์ที่สูงของการกระจายที่ปรับเปลี่ยนได้โดยค่าเฉลี่ยของ EVT สำหรับการประมาณค่าความน่าจะเป็นที่ลดน้อยลง สามารถประยุกต์ค่า EVT ที่หลากหลาย โดยข้อมูลที่ได้จากตัวอย่างที่ต้องการก่อนการสังเกตเหตุการณ์ที่เป็นด้านเดียวโดยค่าเฉลี่ย ของวิธีใหม่ตั้งอยู่บนพื้นฐานของการจำแนกประเภทโดย RKHS ซึ่งสามารถแก้ปัญหาและเพิ่มเหตุการณ์ที่เป็นด้านเดียวลงไปในการเก็บตัวอย่าง EVT ได้

Rosipal and Leonard (2002) ได้ศึกษาวิธีการของ Least Square Regression Model ที่ได้จากการพัฒนา Kernel Hilbert Space ถูกขยายโดย Kernel Partial Least Space (PLS) Regression Model ซึ่งในโมเดลของ PLS ถูกสร้างขึ้นโดยโมเดลความสัมพันธ์ระหว่างตัวแปรภายในและภายนอก ขณะที่ข้อมูลส่วนมากที่ศึกษาจะอยู่ในตัวแปรภายใน ซึ่งตรงกับข้อตกลงเบื้องต้น ในองค์ประกอบของโมเดล PCR และโมเดล PLS ที่ใช้กับการวิเคราะห์ในสถานการณ์ที่จำนวนของตัว

แปรอธิบายมีมากกว่าจำนวนของตัวแปรที่สังเกตได้ และหรือมีสหสัมพันธ์กันเองระหว่างตัวแปรอิสระมากกว่า 2 ตัวแปรขึ้นไปและอยู่ในระดับที่มีความสัมพันธ์กันสูง และโมเดลข้อมูลที่เป็นการถดถอยแบบไม่เป็นเชิงเส้นและมีความซับซ้อนสูง สามารถใช้โมเดล PLS และ PCR ในการวิเคราะห์ที่ได้ผลที่ดีทั้งสองโมเดลในระดับนัยสำคัญที่ต่างกัน

Preda (2005) ได้ศึกษาถึงรูปแบบและวิธีการของการถดถอยสำหรับฟังก์ชันข้อมูล การทำงานโดยการทำซ้ำ วิธีการใช้ช่องว่าง และได้พัฒนาการวิเคราะห์ความถอดถอยสำหรับฟังก์ชันของการสุ่มตัวแปร $x = \{x_t\}_{t \in T}$ จากวิธีการของ Kernel Hilbert Spaces ที่สามารถพัฒนานำมาปรับใช้ได้กับวิธีการของการวิเคราะห์การถดถอยได้อย่างดี

Oscar Gonzalez-Recio at el. (2008) ได้ศึกษาวิธีมาตรฐานของการประเมินพันธุกรรม (E-BLUP) 4 วิธี คือ F-metric Model, Kernel Regression, RKHS Regression, และ Bayesian regression ของพ่อพันธุ์ (ไก่) ซึ่งใช้อัตราการตายของไก่เป็นตัวแปรตอบสนอง ด้วยกรอบแนวคิดของเบส์ (Bayesian Framework) ข้อมูลสถิติอัตราการตายของไก่อายุมาก (14–42 วัน) 12, 167 ตัว จากพ่อพันธุ์ 200 ตัว ได้รับการตรวจแก้ล่วงหน้าสำหรับผลที่คงที่และผลเชิงสุ่ม ที่ถูกใช้ในตัวแบบสำหรับการประเมินพันธุกรรมและผลด้านการจับคู่ผสมพันธุ์ และสำหรับพ่อพันธุ์แต่ละตัวก็มีการคำนวณค่าเฉลี่ยของสถิติที่ได้รับการแก้ไขด้วย ในวิธีการ 3 วิธีที่ใช้ในการประเมินพันธุกรรม ปรากฏว่ามีข้อมูล SNP ที่เกี่ยวข้องกับอัตราการตาย 24 ชุด และมี 1,000 ชุดในวิธี Bayesian Regression โดยทำให้มีเครื่องหมายที่ Genome โดยรวม ในวิธี E-BLUP นั้น ค่าเฉลี่ยตอนท้าย (Posterior Mean) ของการสืบทอดสายพันธุ์ (Heritability) ของอัตราการตาย เท่ากับ 0.02 ซึ่งชี้ว่าเราสามารถที่จะทำให้การประเมินพันธุกรรมดีขึ้นได้ ซึ่งชี้ว่า RKHS เป็นวิธีที่ทำให้มีความแปรปรวนมากขึ้น วิธีนอนพาราเมตริก 2 วิธี (คือ Kernel และ RKHS Regression) เหมาะกับข้อมูลมากกว่า โดยมีค่าผลรวมกำลังสอง (Sum of Squares) ที่เหลือต่ำกว่าด้วย ส่วนความสามารถในการทำนาย ซึ่งวัดด้วยวิธี Cross Validation ชี้ให้เห็นถึงข้อดีของวิธี RKHS ซึ่งทำให้ความถูกต้องแม่นยำเพิ่มขึ้นมากจาก 25% เป็น 150% เมื่อเปรียบเทียบกับวิธีอื่น

Gianola and Van Kaam (2008) ใช้วิธี RKHS สำหรับทำนายค่าเชิงพันธุกรรมโดยรวมสำหรับลักษณะเชิงปริมาณ และใช้ประโยชน์จากข้อมูล Phenotype และ Genome ไปพร้อมกัน โดยมีข้อเสนอถึง ความจำเป็นต้องจัดกระทำกับข้อมูลแบบนอนพาราเมตริก เพื่อจับการกระทำร่วมกันที่ซับซ้อนหลายกิจกรรมที่เกิดขึ้นอย่างมีศักยภาพในตัวแบบ Genome โดยรวมที่เกี่ยวข้องกับตัวทำเครื่องหมาย (Markers) หรือชุดข้อมูลภาวะหลายรูปแบบหรือ SNP (Single-Nucleotide Polymorphism) หลังจากการตรวจทานด้วยวิธี RKHS แล้ว พบว่า ข้อกำหนดเฉพาะทางสถิติยอมรับ

การนำเสนอตัวแบบเชิงเส้นมาตรฐานแบบผลผสม ด้วยค่า Parameter ที่ราบรื่นที่ถูกจัดกระทำในฐานะที่เป็นองค์ประกอบของความแปรปรวน

Esperanze at el. (2008) ได้ศึกษาในหลาย ๆ ด้านของทางวิศวกรรมที่ศึกษาในเรื่องของระบบไบโอซิสเต็ม ในเรื่องของการเลือกใช้สถิติในการวิเคราะห์ข้อมูล และทดสอบสมมติฐานขั้นพื้นฐานที่จำเป็นสำหรับวิธีการพยากรณ์แบบปกติ (Conventional Forecasting Method) ในแต่ละสถานการณ์ที่ต้องหาวิธีและทางเลือกในเรื่องต่าง ๆ ที่ต้องนำสถิติมาใช้ในการพิจารณาวิเคราะห์ถึงข้อตกลงเบื้องต้นต่าง ๆ ที่ข้อมูลในทางวิศวกรรมไม่เป็นไปตามข้อตกลงเบื้องต้น การใช้ฟังก์ชันการประมาณค่าแบบนอนพาราเมตริก (Nonparametric Function Estimation) ประมาณค่าเป็นฟังก์ชันวิเคราะห์ข้อมูลในช่วงแคบ ๆ ซึ่งใช้ข้อมูลจากจุดเล็ก ๆ ทำให้เกิดจุดอ่อนของมูลในเรื่องของข้อตกลงเบื้องต้น เช่น ความต่อเนื่องและความแตกต่างของฟังก์ชันเป้าหมาย คือ การใช้ที่มากกว่าของนิยามของ ข้อตกลงเบื้องต้นของรูปแบบของฟังก์ชันเป้าหมายทั้งหมด ซึ่งผลการวิจัยพบว่า การนำกฎข้อตกลงพื้นฐานมาใช้ในการตัดสินใจแสดงให้เห็นว่าการประยุกต์ใช้ ฟังก์ชันการประมาณค่าแบบนอนพาราเมตริกซ์ (Nonparametric Estimate Method) ภายใต้กฎข้อตกลงเบื้องต้นต่าง ๆ ของสถิติในขั้นตอนแรกสามารถเชื่อมต่อกันได้อย่างดี ของการใช้สถิติสำหรับการประยุกต์ใช้ของการประมาณค่าแบบเคอร์เนล (Kernel Estimation) และการใช้สถิติในการประมาณค่าตัวแปรที่ศึกษา 1 ตัว และหลายตัวแปร และการวิเคราะห์ฟังก์ชันของความหนาแน่น รวมถึงการประมาณค่าการวิเคราะห์ถดถอย (Regression Estimation) และในบางโมเดลจะถูกประยุกต์ใช้ในต่างสถานการณ์ และการจำลองที่ต่างกัน

Okba Taouali et al. (2009) ได้ศึกษาเปรียบเทียบ โมเดล Volterra และ โมเดล RKHS ใน MISO โดย RKHS โมเดลซึ่งเป็นโมเดลการวิเคราะห์ที่ไม่เป็นเส้นตรง โดยใช้ทฤษฎีทางสถิติทั่วไปการหาข้อสรุปที่เป็นปกติ เป็นลักษณะเฉพาะของการรวมกันเข้าเป็นสมการเชิงเส้น ตามฟังก์ชันของเคอร์เนล (Kernels) ความซับซ้อนของ Valterra โมเดลต้องอาศัยระดับความจำในโมเดลซึ่งตรงข้ามกับโมเดล RKHS ที่จำนวนข้อมูลมาจากค่าที่สังเกตได้เพียงอย่างเดียว ในการปฏิบัติของทั้งสองโมเดลนี้มีการประเมินโดยการใช้ Monte Carle เป็นอันดับแรก โดยการใช้ตัวเลขสมมติ และการทดสอบทั้งสองวิธีนี้ยังคงเป็นวิธีการศึกษาสำหรับการสร้างโมเดลปฏิกิริยาเคมี และผลที่ได้จากการทดลองเป็นผลที่ดี

Geng, Minggen, and Zhang (2010) ได้พัฒนาวิธีการแบบ Homology Perturbation Method (HPM) ประยุกต์รวมกับวิธีการของ Reproducing Kernel Hilbert Space Method (RKHSM) และได้ผลดีประสบความสำเร็จสำหรับการแก้ปัญหาความเป็นเอกพจน์ของข้อมูลแบบไม่เป็นเชิงเส้น ซึ่งเป็นจุดเริ่มต้นของปัญหา ซึ่งรวมทั้งข้อมูลแบบเชิงเส้นทั่วไป และยังแสดงให้เห็น

ว่าวิธีการของ (RKHSM) เป็นเทคนิคที่มีความแม่นยำและเที่ยงตรงสำหรับแก้ปัญหาความเป็นเอกพจน์ของข้อมูลแบบไม่เป็นเชิงเส้น และยังมีประสิทธิภาพในการแก้ปัญหาข้อมูลที่ไม่เป็นเชิงเส้นอื่น ๆ ได้ดีด้วย

Bouboulis (2011) ได้ศึกษาพบว่า RKHS เป็นเครื่องมือที่ดีมีประโยชน์และทรงพลังของการวิเคราะห์ในหลาย ๆ ด้าน เช่น การวิเคราะห์ความแปรปรวนร่วมหลาย ๆ ตัวแปร และเป็นเครื่องมือในการเรียนรู้ ของข้อมูลแบบเศษส่วน และยังใช้เทคนิคการใช้ค่าเฉลี่ยสำหรับการคำนวณกับ Kernel ที่ซึ่งมีลักษณะเดียวกันกับ RKHS ทั่ว ๆ ไปได้ผลดีอีกด้วย

Geng (2011) ได้ศึกษาและใช้วิธีการของ Reproducing Kernel Method ในการแก้ปัญหาในเรื่องของข้อจำกัดหรือขอบเขตในการนำวิธีต่าง ๆ มาใช้ในการแก้ปัญหาของข้อมูลแบบเดิม ๆ ซึ่งพบว่าวิธีการของ Reproducing Kernel Method สามารถใช้ในการแก้ปัญหาในเรื่องของข้อจำกัดต่าง ๆ ของข้อมูลได้ดี

สรุปทิศทางของข้อค้นพบจากงานวิจัย

การใช้วิธีการรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ เป็นรูปแบบและวิธีการถดถอยสำหรับข้อมูลการทำซ้ำ และได้พัฒนาการวิเคราะห์การถอดถอยสำหรับฟังก์ชันของการสุ่มตัวแปร สามารถใช้ในการแก้ปัญหาในเรื่องของข้อจำกัดต่าง ๆ ของข้อมูลและเป็นเครื่องมือที่ใช้ในการวิเคราะห์ในอีกหลายด้าน ใช้ได้ผลที่ดีกับข้อมูลและวิธีการทางนอนพาราเมตริก และการใช้ในการวิเคราะห์ข้อมูลที่ไม่เป็นเส้นตรง เป็นต้น

งานวิจัยด้านแบนวิดจ์

Jones, Marron, and Sheather (1996) ได้ศึกษาและเปรียบเทียบวิธีการเลือกแบนวิดจ์ โดยแบ่งวิธีการเลือกแบนวิดจ์เป็น 2 ช่วงเวลา คือ 1) ช่วงเวลาที่ 1 วิธีการเลือกแบนวิดจ์ในช่วงเวลานี้ เช่น วิธีการเลือกแบนวิดจ์ Rule of Thumb วิธีการเลือกแบนวิดจ์ Least Square Cross-Validation และวิธีการเลือกแบนวิดจ์แบบ Biased Cross-Validation เป็นต้น 2) ช่วงเวลาที่ 2 วิธีการเลือกแบนวิดจ์ในช่วงเวลานี้ เช่น วิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in และวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap เป็นต้น โดยกำหนดสมมติฐานในการวิจัย 3 ข้อ คือ

1. วิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 จะให้ผลลัพธ์ที่ดีกว่าวิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 1
2. วิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 สามารถนำมาใช้กับโปรแกรมคอมพิวเตอร์ได้ง่ายกว่าวิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 1
3. วิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in เป็นวิธีการเลือกแบนวิดจ์ที่ดีที่สุดในวิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 จากวิธีการเลือกแบนวิดจ์ทั้งหมด

Jones, Marron, and Sheather (1996) ทำการเปรียบเทียบโดยใช้ข้อมูลจริงจาก The Australian Institute of Sport Data สามารถสรุปได้ว่า

1. วิธีการเลือกแบนวิดจ์แบบ Rule of Thumb จะให้ผลลัพธ์ที่ขาดลักษณะที่สำคัญของข้อมูลไปบางส่วน
2. วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะให้ผลลัพธ์ในทิศทางที่ด้อยกว่าและเชื่อถือไม่ค่อยได้
3. วิธีการเลือกแบนวิดจ์แบบ Biased Cross-Validation จะให้ผลลัพธ์ไม่แน่นอน

สรุปได้ว่าวิธีการเลือกแบนวิดจ์จากช่วงเวลาที่ 1 นั้น ไม่เหมาะสมที่จะนำไปใช้อย่างแพร่หลาย แต่วิธีการเลือกแบนวิดจ์ในช่วงเวลาที่ 2 นั้น เช่น วิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in ให้ผลลัพธ์ที่สม่ำเสมอและคงที่ แสดงให้เห็นว่าเหมาะสมที่จะนำไปใช้ในโปรแกรมทางคอมพิวเตอร์ นอกจากนี้ Jones, Marron and Sheather ได้ศึกษาโดยการจำลองข้อมูลจากการแจกแจงแบบปกติผสม (Mixture Normal Distribution) 15 ฟังก์ชัน โดยมีขนาดตัวอย่างเท่ากับ 100 และ 1,000 ได้ผลการศึกษา ดังนี้

1. การแจกแจงของค่าแบนวิดจ์จากการแบนวิดจ์แบบ Rule of Thumb มีค่าเฉลี่ยค่อนข้างมากแต่ค่าความแปรปรวนจะน้อยกว่าวิธีเลือกแบนวิดจ์อื่น ๆ เนื่องจากมาจากประมาณแบบสเกล จึงทำให้ลักษณะไม่เป็นไปโดยสุ่ม
2. การแจกแจงของค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าเฉลี่ยใกล้เคียงกับค่าแบนวิดจ์ของ MISE แต่มีค่าความแปรปรวนมากกว่าวิธีอื่นโดยลักษณะจะเป็นไปในทางต่ำกว่าค่าประมาณ
3. การแจกแจงของค่าแบนวิดจ์จากการเลือกแบนวิดจ์แบบ Biase Cross-Validation หารูปแบบได้ยาก เนื่องจากโค้งที่ได้มีลักษณะที่ไม่แน่นอน เมื่อขนาดตัวอย่างเท่ากับ 100 จะมีค่าเฉลี่ยมาก ในขณะที่ขนาดตัวอย่างเท่ากับ 1,000 ค่าเฉลี่ยจะมีค่าไม่แน่นอน บางครั้งมีค่ามากและบางครั้งมีค่าเข้าใกล้ค่าแบนวิดจ์ที่ประมาณได้
4. ลักษณะของค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in จะมีลักษณะคล้ายกับค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Rule of Thumb และวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation การกระจายจะมีจุดศูนย์กลางใกล้กับค่าแบนวิดจ์ที่เหมาะสม แต่การกระจายจะน้อยกว่าค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าความเอนเอียงมากกว่าเนื่องจากมีตัวแปรมากกว่า

5. ลักษณะของค่าแบนวิดจ์ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Smoothed Bootstrap โดยทั่วไปจะมีลักษณะคล้ายกับค่าแบนวิดจ์จากวิธีการเลือกแบนวิดจ์แบบ Solve the Equation Plug-in แต่ค่าที่ได้จะมีค่ามากกว่า

Hansen (2004) ได้ศึกษาวิธีการเลือกแบนวิดจ์แบบ Plug-in โดยใช้ค่า AMISE เป็นเกณฑ์ในการพิจารณา ทำการจำลองข้อมูลจากการแจกแจงปกติผสม 9 การแจกแจง กำหนดขนาดตัวอย่างเท่ากับ 10, 30, 60 และ 120 และใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ทำการวนซ้ำจำนวน 1,000 รอบ ในแต่ละตัวแบบและขนาดตัวอย่าง ได้ผลการวิจัย ดังนี้

1. ค่า MISE และค่า AMISE มีรูปร่างคล้ายกันในหลาย ๆ กรณี โค้งการแจกแจงของข้อมูลจะใกล้เคียงกัน โดยเฉพาะเมื่อค่าแบนวิดจ์มีขนาดเล็ก สำหรับในกรณีที่ค่าแบนวิดจ์มีขนาดใหญ่ ค่า MISE จะแสดงลักษณะที่สำคัญผิดพลาด

2. ค่าแบนวิดจ์ที่ทำให้ค่า MISE และ AMISE น้อยที่สุดจะมีค่าใกล้เคียงกัน

3. โค้งข้อมูลที่ได้จากค่าแบนวิดจ์ที่ใช้ค่า MISE เป็นเกณฑ์จะผันแปรตามการแจกแจง

Mugdadi and Ahman (2004) อธิบายขั้นตอนวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation และวิธีการเลือกค่าแบนวิดจ์แบบ Kernel Contrast และทำการเปรียบเทียบทั้ง 2 วิธี โดยใช้วิธีการจำลองแบบมอนติคาร์โล และขนาดตัวอย่างมี 2 ขนาด คือ 25 และ 50 และกำหนดให้ตัวแปรสุ่มมีรูปแบบ $g(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_m) = \sum_{i=1}^{m} X_i$ โดยที่ $m = 2$ โดยศึกษาฟังก์ชันเคอร์เนล 3 ฟังก์ชัน คือ 1) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน 2) ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov Kernel Function) และ 3) ฟังก์ชันเคอร์เนลแบบไบเวท (Biweight Kernel Function) ทำการจำลองข้อมูลจากการแจกแจง 3 การแจกแจง คือ 1) การแจกแจงแบบปกติ 2) การแจกแจงแบบเอกซ์โพเนนเชียล และ 3) การแจกแจงแบบโคชี ทำการเปรียบเทียบโดยนำค่าความแปรปรวนและค่าความเอนเอียงของค่าแบบวิดจ์ที่ประมาณมาพิจารณา สามารถสรุปผลการวิจัยได้ดังนี้

1. ค่าความแปรปรวนของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast มีค่าน้อยกว่าค่าความแปรปรวนของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เมื่อจำลองโดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบไบเวท

2. ในกรณีอื่น ๆ ค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะมีค่าน้อยกว่าค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation กรณีที่ข้อมูลมีการแจกแจงแบบ

ปกติโดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบไบเวท และกรณีที่ข้อมูลมีการแจกแจงแบบเอกซ์โพเนนเชียล ในฟังก์ชันเคอร์เนลแบบเกาส์เซียน

3. ค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าน้อยกว่าค่าความแปรปรวนของค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast กรณีที่ข้อมูลมีการแจกแจงแบบโคชีและการแจกแจงแบบเอกซ์โพเนนเชียล เมื่อใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบไบเวท

4. ค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast มีค่าน้อยกว่าค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เมื่อข้อมูลมีการแจกแจงแบบโคชี ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ และฟังก์ชันเคอร์เนลแบบไบเวท

5. ค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation มีค่าน้อยกว่าค่าสัมบูรณ์ของค่าความเอนเอียงของ $\hat{h}$ จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast ในกรณีที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนทั้ง 2 ฟังก์ชัน

6. ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะมีค่าน้อยกว่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เมื่อทำการจำลองข้อมูลจากการแจกแจงแบบปกติ และการแจกแจงแบบเอกซ์โพเนนเชียล แต่ในทางกลับกันเมื่อทำการจำลองข้อมูลจากการแจกแจงแบบโคชี ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ที่ได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation จะมีค่าน้อยกว่าค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) ของวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast

7. ในทุก ๆ กรณีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) จะลดลงเมื่อขนาดของตัวอย่างเพิ่มขึ้น

นอกจากนี้ Mugdadi and Ahman (2004) ได้นำข้อมูลจริงโดยใช้ข้อมูลที่เกี่ยวข้องกับการประกันภัยจากบริษัทออโต้ การประกันภัย ประเทศสหรัฐอเมริกา เดือนกันยายน ในปี ค.ศ.2001 มาวิเคราะห์ โดยใช้ฟังก์ชันของตัวแปรสุ่ม คือ $g(\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_m) = \sum_{i=1}^{m} X_i$ โดยที่ $m = 2$ ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ได้ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation เท่ากับ 0.29 และได้ค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast เท่ากับ 0.42 จากนั้น Mugdadi และ Ahman ยังนำข้อมูลจริงข้างต้นมาศึกษาในวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast โดยใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียนโดยมี

ค่าเฉลี่ยเท่ากับ 0 และความแปรปรวนที่แตกต่างกัน 4 ค่า คือ 1, 4, 9 และ 16 โดยมีสัมประสิทธิ์ความแตกต่าง (Contrast Coefficient) แทนด้วย p_1, p_2, p_3 และ p_4 ตามลำดับ โดยสรุปผลการทดลองได้ดังนี้

1. วิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation ไม่เหมาะสมสำหรับการประมาณค่าความหนาแน่น ในกรณีที่ m = 1

2. เมื่อสัมประสิทธิ์ความแตกต่างมีค่าเท่ากับ $\frac{1}{2}$ และ $\frac{1}{2}$ จะได้การประมาณความหนาแน่นเช่นเดียวกันกับกรณีที่สัมประสิทธิ์ความแตกต่างมีค่าเท่ากับ $\frac{1}{3}$ และ $\frac{2}{3}$ แต่จะคล้ายกับกรณีที่สัมประสิทธิ์ความแตกต่างมีค่าเท่ากับ $\frac{1}{10}$ และ $\frac{9}{10}$ แต่ไม่เหมือนกันทุกประการ

3. การประมาณความหนาแน่นที่ใช้ 2 ฟังก์ชันเคอร์เนล หรือ 4 ฟังก์ชันเคอร์เนล เมื่อฟังก์ชันมีค่าถ่วงน้ำหนักเดียวกันจะได้โค้งที่เหมือนกัน

4. เมื่อใช้ค่าสัมประสิทธิ์ความแตกต่าง ค่าแบนวิดจ์ที่ได้จากการประมาณจะมีค่าเล็กกว่าวิธีการเลือกแบนวิดจ์แบบ Least Square Cross-Validation แต่ถ้าใช้ฟังก์ชันเคอร์เนล 4 ฟังก์ชันในการประมาณความหนาแน่น วิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่าแบนวิดจ์และประมาณค่าความหนาแน่นได้ปรับเรียบมากกว่า

สรุปว่าตัวเลือกของแบนวิดจ์โดยใช้ข้อมูลจริงจะสม่ำเสมอ และจากการจำลองข้อมูลจะเห็นว่าค่าแบนวิดจ์ที่ประมาณได้จากวิธีการเลือกแบนวิดจ์แบบ Kernel Contrast จะให้ค่าแบนวิดจ์และการประมาณความหนาแน่นที่ปรับเรียบมากกว่า

สรุปทิศทางของข้อค้นพบจากงานวิจัย

การพิจารณาเลือกแบนวิดจ์ที่เหมาะสมสำหรับการวิจัย รวมถึงรูปแบบของแบนวิดจ์ที่นำไปใช้กับฟังก์ชันเคอร์เนลแบบต่าง ๆ วิธีการเลือกแบนวิดจ์ในรูปแบบต่าง ๆ ข้อตกลงเบื้องต้นของรูปแบบ และเกณฑ์การเลือกใช้แบนวิดจ์ในงานวิจัย

จากการศึกษางานวิจัยแสดงว่าฟังก์ชันเคอร์เนลฮิลเบอร์ทสเปซ (Reproducing Kernel Hilbert Space) เป็นวิธีการวิเคราะห์ข้อมูล ที่มีความสะดวกการลดขั้นตอนการคิดคำนวณ ตลอดจนสามารถนำมาปรับใช้กับข้อมูลในลักษณะที่ไม่เป็นเส้นตรง (Nonlinear) และแบบนอนพาราเมตริก (Nonparametric) ให้สามารถนำมาวิเคราะห์ได้อย่างดี ลดขั้นตอนกระบวนการวิเคราะห์ และได้ผลเป็นอย่างดีและน่าเชื่อถือตลอดจนเป็นเครื่องมือที่ดีมีประโยชน์และทรงพลังของการวิเคราะห์ในหลายด้าน เช่น การวิเคราะห์ความแปรปรวนร่วมหลาย ๆ ตัวแปร การประยุกต์ไปสู่การวิเคราะห์และแก้ปัญหาในด้านของวิศวกรรม ทางการแพทย์ ทางชีววิทยาในเรื่องของการวิเคราะห์ยีนส์ และ

สามารถปรับประยุกต์โมเดลไปสู่การใช้ฟังก์ชั่นทางคณิตศาสตร์ที่มาสร้างโมเดลในด้านของฟังก์ชันเอกซ์โพเนนเชียล, ควอไทล์ และฟังก์ชันลอการิทึม รวมถึงอนุกรมเวลาและยังจัดกระทำได้ทั้งข้อมูลจริงและการจำลองข้อมูลซึ่งจะนำไปสู่การประยุกต์ใช้และพัฒนาการใช้วิธีการวัดที่มีความเหมาะสมถูกต้องและแม่นยำ และเป็นเครื่องมือในการเรียนรู้ เป็นเทคนิคที่มีความแม่นยำและเที่ยงตรงสำหรับแก้ปัญหาความเป็นเอกพจน์ของข้อมูลแบบไม่เป็นเชิงเส้น และยังมีประสิทธิภาพในการแก้ปัญหาข้อมูลที่ไม่เป็นเชิงเส้นอื่น ๆ ได้ดีด้วย เป็นวิธีการที่สามารถสร้างแบบจำลองทางสถิติที่น่าสนใจการวิเคราะห์ผลจากค่าที่วัดได้ของข้อมูลที่หลากหลายและการกระจัดกระจายและข้อมูลอื่น ๆ ไปพร้อมกันได้มีกรอบแนวคิดสำหรับการประมาณค่าฟังก์ชันที่มีความยืดหยุ่นและสร้างตัวแบบทางสถิติที่มีข้อมูลทั้งโดยตรงและโดยอ้อม รวมถึงการวิเคราะห์ฟังก์ชั่นแบบ RKHS เป็นพื้นฐานสำหรับการประมาณค่าความคลาดเคลื่อน และวิธีการต่าง ๆ ในการปรับข้อมูลเข้าสู่ระบบ และสามารถจัดการกับปัญหาและการกระจายต่าง ๆ ของข้อมูลที่หลากหลาย เช่น ทฤษฎีของเกาส์ใช้สกุลเลขยกกำลังทั่วไป (General Exponential Families) การประมาณค่าที่มีประสิทธิภาพ (Robust Estimation) การสังเกตแบบช่วง (Interval Observations) ได้เป็นอย่างดีอีกด้วย

ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับการพยากรณ์ราคาข้าวเปลือกเจ้า

จิตราภรณ์ ฟั่นศิริ (2547) ศึกษาการพยากรณ์ราคาส่งออกข้าวด้วยวิธี ARIMA โดยมีวัตถุประสงค์เพื่อศึกษารูปแบบและพยากรณ์ราคาส่งออกข้าวของไทย ในการวิเคราะห์จะใช้ข้อมูลราคาส่งออกข้าวเป็นรายเดือนในช่วงเดือนมกราคม พ.ศ. 2531 ถึงเดือนธันวาคม พ.ศ. 2546 จำนวน 192 เดือน ข้อมูลจากกรมการค้าต่างประเทศ วิธีการศึกษาจะทดสอบความนิ่งของข้อมูลโดยใช้วิธีการทดสอบ Unit Root และกำหนดตัวแบบด้วยวิธีของบอกซ์-เจนกินส์ ผลการทดสอบพบว่าข้อมูลราคาส่งออกข้าวมีลักษณะไม่นิ่งจึงทำการหาผลต่างอันดับ 1 และจากการพิจารณาค่าสหสัมพันธ์ในตัวเองจะได้ตัวแบบที่เหมาะสมขึ้นอยู่กับค่า AR(1) และ AR(19) โดยมีค่าสัมประสิทธิ์เท่ากับ 0.360 และ 0.228 ตามลำดับ และมีนัยสำคัญทางสถิติที่ระดับ .01 สำหรับการตรวจสอบความถูกต้องพบว่าค่าประมาณความคลาดเคลื่อนมีลักษณะเป็นเชิงสุ่ม ที่ระดับนัยสำคัญทางสถิติ .1 จากค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสองและค่าสัมประสิทธิ์ Thiel ที่มีค่าต่ำสุด จะได้ว่าตัวแบบที่มีความเหมาะสมที่ดีที่สุดซึ่งมีสมการพยากรณ์ คือ $\nabla \ln P_1 = -0.004190 + 0.360211\nabla$ $In P_{t-1} + 0.227957\nabla In P_{t-19} + e_t$ เมื่อเทียบกับตัวแบบอื่น ๆ ดังนั้นในการศึกษาครั้งนี้จึงนำตัวแบบ AR (1) และ AR (19) ไปพยากรณ์ราคาส่งออกข้าวในช่วงเดือนมกราคม ถึงเดือนเมษายน พ.ศ. 2547 ได้ค่าเท่ากับ 205, 204, 202, และ 201 เหรียญสหรัฐต่อตัน ตามลำดับ

พรหมภรณ์ แสงภัทรเนตร (2548) ได้ศึกษาเรื่อง การพยากรณ์ราคาข้าวภายในประเทศ วิเคราะห์การเคลื่อนไหวของพื้นที่ปลูกผลผลิตและราคาข้าว ได้แก่ข้าวเปลือกเจ้านาปี 5%, 10%, 15%, 25% และหอมมะลิ 100% โดยใช้ข้อมูลราคาข้าวแต่ละชนิด เป็นรายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2527 จนถึงเดือนกุมภาพันธ์พ.ศ. 2548 และใช้เทคนิคการพยากรณ์เชิงปริมาณ ได้แก่วิธี Winters, วิธีBox–Jenkins , วิธีแยกส่วนประกอบอนุกรมเวลา และวิธีวิเคราะห์การถดถอย และใช้ค่า MAPE, MAD และMSD ในการเปรียบเทียบวิธีการพยากรณ์ ซึ่งจากผลการศึกษาพบว่า วิธีแยกส่วนประกอบเหมาะกับการพยากรณ์ราคาข้าวเปลือกเจ้านาปี 5% ,15%, 25% และหอมมะลิ 100% ส่วนวิธี Box–Jenkins เหมาะกับการพยากรณ์ราคาข้าวเปลือกเจ้านาปี 10% ทั้งนี้ได้เสนอแนะแนวทางแก้ไขเกี่ยวกับเรื่องการพยากรณ์ด้วยวิธีแยกส่วนประกอบอนุกรมเวลาว่าควรจะต้องมีการคำนึงถึงค่าความผันแปรของวัฏจักรด้วย อีกทั้งยังมีปัจจัยอื่น ๆ ที่อาจจะมีผลกระทบต่อราคาข้าวซึ่งไม่ได้นำมาวิเคราะห์ในการศึกษา เช่น ต้นทุนการผลิต ราคาน้ำมันเป็นต้นและข้อมูลใหม่ ๆ ที่เพิ่มขึ้นเมื่อเวลาผ่านไป ดังนั้นจึงต้องมีการปรับปรุงตัวแบบด้วยการเพิ่มปัจจัยเหล่านี้ในการปรับปรุงตัวแบบ รวมถึงการนำข้อมูลใหม่ที่เพิ่มขึ้นเข้ามาร่วมตรวจสอบความเหมาะสมใหม่อีกครั้ง เพื่อเพิ่มความแม่นยำในการนำมาใช้ประโยชน์

เบญจมาศ ธัญน้อม (2549) ศึกษาการพยากรณ์ราคาสัญญาล่วงหน้าข้าวขาว 5% ด้วยวิธี ARIMA เพื่อการพยากรณ์ราคาสัญญาล่วงหน้าข้าวขาว 5% โดยวิธี ARIMA ซึ่งใช้ข้อมูลรายวันของราคาสัญญาล่วงหน้า 3 สัญญา ดังนี้ ข้าวขาว 5% ของเดือนมกราคม พ.ศ. 2549 จำนวน 101 ตัวอย่าง ข้าวขาว 5% ของเดือนกุมภาพันธ์ พ.ศ. 2549 จำนวน 101 ตัวอย่าง และข้าวขาว 5% ของเดือนมีนาคม พ.ศ. 2549 จำนวน 79 ตัวอย่าง ได้จำนวนข้อมูลทั้งสิ้น 281 ตัวอย่าง มีลักษณะเป็นอนุกรมเวลาจึงทำการทดสอบด้วยวิธีการ Unit Root Test และกำหนดตัวแบบด้วยวิธีบอกซ์-เจนกินส์ ผลการศึกษาพบว่า ราคาสัญญาล่วงหน้าข้าวขาว 5% มีลักษณะนิ่งที่ผลต่างอันดับที่ 1 และจากการพิจารณาค่าสหสัมพันธ์ในตัวเองจะได้ตัวแบบ AR(7) และ MA(18) เป็นตัวแบบของเดือนมกราคม ตัวแบบ AR(8) และ MA(8) เป็นตัวแบบของเดือนกุมภาพันธ์ และตัวแบบ AR(8) และ MA(8) เป็นตัวแบบของเดือนมีนาคม จากการประมาณค่าของตัวแบบพบว่าทุกตัวแบบมีค่าสถิติแตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ .01 จากค่าสถิติของ Q พบว่าค่าความคลาดเคลื่อนที่ประมาณมีลักษณะเป็นเชิงสุ่ม เมื่อนำสมการไปพยากรณ์ราคาสัญญาล่วงหน้าข้าวขาว 5% ในเดือนมกราคม ในอีก 3 ช่วงเวลาถัดไปจะได้ค่าพยากรณ์ 11.2715, 11.2650 และ 11.2676 บาทต่อกิโลกรัม พยากรณ์ราคาสัญญาล่วงหน้าข้าวขาว 5% ในเดือนกุมภาพันธ์ ในอีก 3 ช่วงเวลาถัดไปจะได้ค่าพยากรณ์ 11.1337, 11.1505 และ 11.1334 บาทต่อกิโลกรัม และพยากรณ์ราคาสัญญาล่วงหน้าข้าวขาว 5% ในเดือน

มีนาคม ในอีก 3 ช่วงเวลาถัดไปจะได้ค่าพยากรณ์ 11.1337, 11.1505 และ 11.1334 บาทต่อกิโลกรัม ตามลำดับ

อัมรา เวียงวีระ (2550) ศึกษาราคาข้าวในตลาดโลกและประเทศไทย พบว่าการผลิตและการบริโภคข้าวส่วนใหญ่อยู่ในทวีปเอเชีย เนื่องจากมีสภาพแวดล้อมที่เหมาะสมต่อการผลิตข้าวและมีพฤติกรรมการบริโภคข้าวเป็นอาหารหลัก นอกจากนี้จากที่มีการขยายตัวของประชากรอย่างรวดเร็วในภูมิภาคดังกล่าว ส่งผลให้มีการสนับสนุนการผลิตข้าวเพื่อให้เพียงพอต่อความต้องการใช้ภายในประเทศ และก่อให้เกิดความมั่นคงด้านอาหาร (Food Security) ในประเทศต่าง ๆ นอกจากนั้นยังสามารถอาศัยการส่งออกข้าวเป็นแหล่งรายได้ที่สำคัญของประเทศ สภาวะการผลิตข้าวของโลกมีการเปลี่ยนแปลง เพิ่มและลดตามภาวการณ์ผลิตของประเทศผู้ส่งออกและนำเข้าข้าวที่สำคัญในตลาดโลก ผลผลิตข้าวของโลกในช่วงปีการผลิต พ.ศ. 2546 ถึง พ.ศ. 2550 มีการขยายตัวเพิ่มขึ้นโดยตลอด โดยมีอัตราการเจริญเติบโตของผลผลิตเฉลี่ยร้อยละ 2.376 ขณะที่ความต้องการใช้ในประเทศที่เพิ่มขึ้นในอัตราเฉลี่ยร้อยละ 0.616 ในขณะที่มีการส่งออกและนำเข้าเพิ่มขึ้นในอัตราเฉลี่ยร้อยละ 6.406 มีผลให้ปริมาณข้าวคงเหลือสะสมในสต็อกปลายปีมีแนวโน้มลดลง ส่วนการผลิตในช่วงปี พ.ศ. 2541 ถึงปี พ.ศ. 2550 ประเทศไทยมีพื้นที่เพาะปลูกข้าวประมาณปีละ 57.5 ล้านไร่ ผลผลิตข้าวเปลือกประมาณ 19 ถึง 27 ล้านตัน ผลผลิตข้าวส่วนใหญ่ได้มาจากการปลูกข้าวในฤดูนาปีประมาณร้อยละ 80 ของผลผลิตข้าวเปลือกทั้งหมด โดยมีผลผลิตต่อไร่มีแนวโน้มเพิ่มขึ้นเพียงเล็กน้อยโดยเฉลี่ยประมาณ 389 กิโลกรัม ส่วนผลผลิตในฤดูนาปรังค่อนข้างคงที่ และผลผลิตต่อไร่มีแนวโน้มลดลงโดยเฉลี่ยประมาณ 685 กิโลกรัม

ไพสิฐ ศิริบวร (2551) ได้ศึกษาเรื่องแนวโน้มราคาน้ำมันดิบ โดยศึกษาถึงราคาน้ำมันดิบในอนาคต เพื่อเป็นแนวทางประกอบการวางแผน การตัดสินใจต่าง ๆ ทางธุรกิจ และการกำหนดนโยบายสำหรับหน่วยงานที่เกี่ยวข้อง โดยใช้ข้อมูลของราคาน้ำมันดิบเฉลี่ยรายเดือนระหว่างเดือนมกราคม พ.ศ. 2543 ถึงเดือนกันยายน พ.ศ. 2551 และใช้เทคนิคการพยากรณ์อนุกรมเวลา ได้แก่ วิธีแยกส่วนประกอบอนุกรมเวลาที่ได้รับอิทธิพลจากแนวโน้มและฤดูกาล, วิธีแยกส่วนประกอบอนุกรมเวลาที่ได้รับอิทธิพลจากฤดูกาล, วิธีHolt’s Linear Method, วิธี Winters และวิธี Brown ในการวิเคราะห์ข้อมูลและใช้ค่า MSD (Mean Square Deviation) เป็นค่าวัดความถูกต้องของการพยากรณ์ที่ได้จากแต่ละวิธีซึ่งจากผลการศึกษาพบว่า วิธี Brown เป็นวิธีที่เหมาะสมที่สุดในการพยากรณ์ราคาน้ำมันดิบ และราคาน้ำมันดิบมีแนวโน้ม ที่ลดต่ำลงเมื่อเทียบกับ ราคาน้ำมันดิบในเดือนสิงหาคมและกันยายน พ.ศ.2551 ซึ่งในอนาคตไม่ว่าราคาน้ำมัน จะเพิ่มสูงขึ้นหรือลดต่ำลงก็ตาม หน่วยงานที่เกี่ยวข้องควรให้ความสำคัญในการวางแผนการใช้พลังงานอย่างมีประสิทธิภาพรวมไปถึง

การใช้พลังงานเลือกเช่นพลังงานแสงอาทิตย์พลังงานลม หรือพลังงานนิวเคลียร์ควบคู่กันไปอย่างเหมาะสม

อัมรินทร์ ก้อนแพง และสมจิตร อาจอินทร์ (2554) ได้พยากรณ์ราคาข้าวเปลือก 4 ชนิด คือ ข้าวเหนียว ข้าวเปลือกเจ้านาปี ข้าวหอมมะลิ และข้าวเจ้านาปี และนาปรัง ความชื้น 14-15% พบว่าวิธีโครงข่ายประสาทเทียมเหมาะสำหรับการพยากรณ์ข้าวเปลือกมากกว่าวิธีการปรับให้เรียบแบบเอกซ์โปเนนเชียลด้วยแนวโน้ม เนื่องจากวิธีโครงข่ายประสาทเทียมมีค่า MAE RMSE และ MAPE ต่ำกว่า สำหรับการศึกษาของ ณัฐนัยน์ เงินเกื้อกูล และพีรยุทธ์ ชาญเศรษฐิกุล (2533, หน้า 1-4) ที่ได้พยากรณ์ราคาข้าวสารและข้าวเปลือกเพื่อตัดสินใจเลือกผลิตข้าวให้ได้กำไรสูงสุด ด้วยวิธีการพยากรณ์ 4 วิธี คือ วิธีถัวเฉลี่ยเคลื่อนที่อย่างง่าย ถัวเฉลี่ยเคลื่อนที่แบบถ่วงน้ำหนัก วิธีปรับให้เรียบแบบเอกซ์โปเนนเชียลโดยรวมองค์ประกอบแนวโน้มและฤดูกาล การวิเคราะห์การถดถอย พบว่าข้าวเปลือกซึ่งได้แก่ ชนิดข้าวสุพรรณ ข้าวปทุม ข้าวหอมมะลิ ข้าวเหนียวยาว ข้าวเหนียวสั้น ควรเลือกวิธีการวิเคราะห์การถดถอย สำหรับข้าวสารซึ่งได้แก่ ชนิดข้าวขาว 100% ข้าวขาว 5% ข้าวปทุม ข้าวเหนียวยาว และข้าวเหนียวสั้น ควรเลือกวิธีการวิเคราะห์การถดถอย แต่ข้าวหอมมะลิ ควรเลือกวิธีถัวเฉลี่ยเคลื่อนที่ถ่วงน้ำหนัก สำหรับชาห่อซึ่งได้แก่ชนิดข้าวขาว 100% ข้าวปทุม ข้าวหอมมะลิ ควรเลือกวิธีการวิเคราะห์การถดถอย และสำหรับปลาย ซึ่งได้แก่ชนิดข้าวขาว ข้าวปทุม ข้าวหอมมะลิ ข้าวเหนียว ควรเลือกวิธีการวิเคราะห์การถดถอย

สรุปทิศทางของข้อค้นพบจากงานวิจัย

รูปแบบและการพยากรณ์ราคาส่งออกข้าวของไทย ใช้การพยากรณ์ราคาส่งออกข้าวด้วยวิธี ARIMA ในการวิเคราะห์จะใช้ข้อมูลราคาส่งออกข้าวเป็นรายเดือน และวิธีการวิเคราะห์การถดถอยในการพยากรณ์พืชผลทางการเกษตร และวิธีการศึกษาจะทดสอบความนิ่งของข้อมูลโดยใช้วิธีการทดสอบ Unit Root และกำหนดตัวแบบด้วยวิธีของบอกซ์-เจนกินส์ และใช้เทคนิคการพยากรณ์อนุกรมเวลา ได้แก่ วิธีแยกส่วนประกอบอนุกรมเวลาที่ได้รับอิทธิพลจากแนวโน้มและฤดูกาล, วิธีแยกส่วนประกอบอนุกรมเวลาที่ได้รับอิทธิพลจากฤดูกาล วิธี Holt’s Linear Method, วิธี Winters เป็นค่าวัดความถูกต้องของการพยากรณ์ที่ได้จากแต่ละวิธี และการหาสมการและตัวแบบที่เหมาะสมที่ใช้พยากรณ์พืชผลทางการเกษตร เพื่อพัฒนาและวางแผนผลผลิตต่อไป

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยนี้เป็นการวิจัยเพื่อพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ สำหรับการพยากรณ์ราคาข้าวเปลือกเจ้าของไทยล่วงหน้า เพื่อเปรียบเทียบประสิทธิภาพของวิธีการรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่ กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ (2x6x9) ได้แก่ การใช้แบนวิดจ์ 2 วิธี คือ วิธีที่ 1 Rules of Thumb และวิธีที่ 2 Siliverman’s Rules of Thumb ขนาดตัวอย่าง 6 ขนาด ($n=$5, 10, 15, 30, 50, 100) และค่าสัมประสิทธิ์สหสัมพันธ์ 9 ค่า ($\rho=$0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) โดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) และเพื่อพยากรณ์ราคาของข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA มีขั้นการวิจัย ดังนี้

ตอนที่ 1 การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ

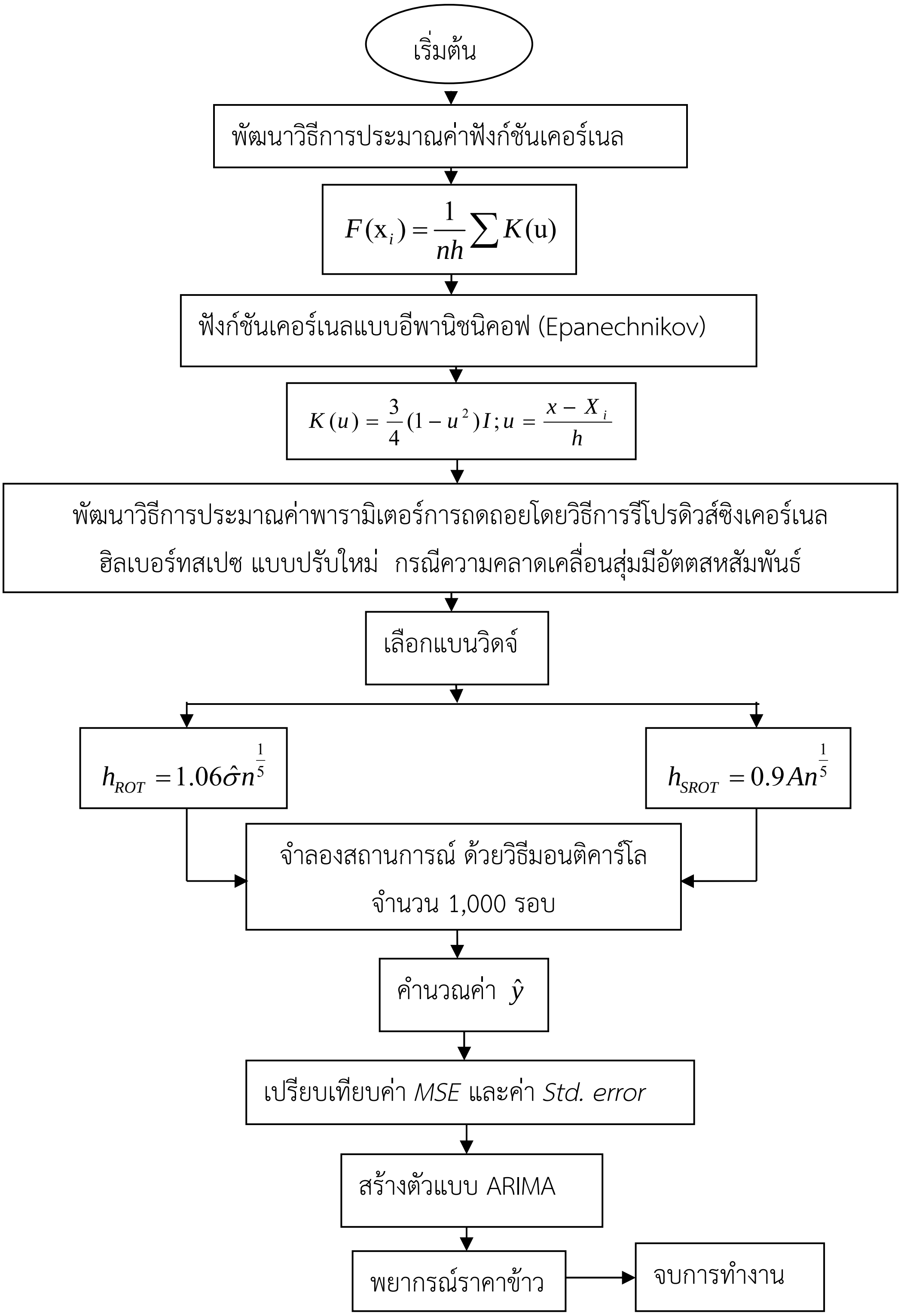
ตอนที่ 2 การเปรียบเทียบประสิทธิภาพของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ภายใต้สถานการณ์ 108 โดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique)

ตอนที่ 3 การพยากรณ์ราคาของข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA

ตอนที่ 1 การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์

การดำเนินการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ ดำเนินการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับการประมาณค่าสัมประสิทธิ์การถดถอยกรณีค่าความคลาดเคลื่อนมีอัตตสหสัมพันธ์ เพื่อรวบรวมเป็นแนวคิดและวิธีการต่าง ๆ ในการนำไปใช้ในการพัฒนาวิธีการประมาณค่าสัมประสิทธิ์การถดถอยแบบใหม่ การศึกษาถึงแนวคิดและทฤษฎีการอนุมานค่าของสถิติที่เกี่ยวข้องกับการคำนวณทางคณิตศาสตร์ (Mathmatical Expecfed Value) ศึกษาที่ได้มาถึงวิธีการประมาณค่าด้วยการใช้วิธีการของเคอร์เนลในการปรับข้อมูลให้เรียบภายใต้กรอบของฮิลเบอร์สเปซ ตามวิธีการของ Ferraty (2007) และงานวิจัยอื่น ๆ ที่เกี่ยวข้องอันจะนำไปสู่แนวทางการพัฒนาวิธีการประมาณค่าสัมประสิทธิ์

การถดถอยโดยวิธีการประมาณค่าแบบรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ในกรณีความคลาดเคลื่อนมีอัตตสหสัมพันธ์ ซึ่งสรุปเป็นแบบการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยแบบวิธีการของรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ การดำเนินการพัฒนาวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่ที่พัฒนาขึ้น ดำเนินตามวัตถุประสงค์การวิจัยที่ต้องการพัฒนาวิธีการประมาณค่าความคลาดเคลื่อนดังกล่าวมาใช้กับข้อมูลเชิงปริมาณที่มีลักษณะเป็น Nonlinear แบบ Nonparametric ซึ่งผู้วิจัยดำเนินการตามลำดับ ดังนี้



ภาพที่ 3-1 แผนผังของขั้นตอนวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่ การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีการรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ กรณีความคลาดเคลื่อนมีอัตตสหสัมพันธ์

การประมาณค่าพารามิเตอร์การถดถอยโดยวิธีการรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)

1. กำหนดสมการ

$$y = F(x) + \varepsilon_i$$

2. กำหนดฟังก์ชันเคอร์เนล
3. Generate X_i แบบ N~(0,1)
4. คำนวณซ้ำจากตัวประมาณค่าเคอร์เนลในกรอบของ RKHS เมื่อ

$$F(x_i) = \frac{1}{nh}\sum K(u)$$

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}u^2); u = \frac{x - X_i}{h}$$

$$K(u) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}(\frac{x - X_i}{h})^2)$$

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ และ

$h_{SROT} = 0.9A(n^{\frac{1}{5}})$

$$F(x_i) = \frac{1}{nh}\sum\frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2h^2}(x^2 - 2xX_i + X_i^2)$$

$$\hat{y} = -\frac{1}{2nh^3\sqrt{2\pi}}\sum(x^2 - 2xX_i + X_i^2) + \varepsilon_i$$

5. ประมาณค่าพารามิเตอร์การถดถอย

การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยแบบปรับใหม่ โดยใช้วิธีของฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov)

การประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่

1. กำหนดสมการ

$$y = F(x) + \varepsilon_i$$

2. กำหนดฟังก์ชันเคอร์เนล
3. Generate X_i แบบ Cau~(0,1)
4. คำนวณซ้ำจากตัวประมาณค่าเคอร์เนลในกรอบของ RKHS

4.1 ข้อมูลอนุกรมเวลา กรณีมีอัตตสหสัมพันธ์ กำหนดด้วย

$$\varepsilon_i = \rho\varepsilon_{i-1} + a_i$$

4.2 ปรับสมการ

$$y = F(x) + \rho\varepsilon_{i-1} + a_i$$

เมื่อ

$$F(x_i) = \frac{1}{nh}\sum K(u)$$

$$K(u) = \frac{3}{4}(1 - u^2)I; u = \frac{x - X_i}{h}$$

$$K(u) = \frac{3}{4}(1 - (\frac{x - X_i}{h})^2)I$$

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ และ

$h_{SROT} = 0.9A(n^{\frac{1}{5}})$

$$F(x) = \frac{1}{nh}\sum\frac{3}{4}(1 - (\frac{x - X_i}{h})^2)I$$

สมการที่ปรับแล้ว

$$\hat{y} = \frac{3}{4nh^3}\sum(h^2 - (x - x_i)^2)I + \rho\varepsilon_{i-1} + a_i$$

5. ประมาณค่าพารามิเตอร์การถดถอย

ภาพที่ 3-2 แผนผังของการพัฒนาวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS)

ตอนที่ 2 การเปรียบเทียบค่าความคลาดเคลื่อนจากการพยากรณ์ข้อมูลล่วงหน้าโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)

ขั้นตอนนี้เป็นการจำลองเพื่อทดสอบวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบิร์ทสเปซแบบปรับใหม่ กับวิธีการรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) โดยการจำลองด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) ด้วยโปรแกรมอาร์ มีหลักการ ดังนี้

ขั้นตอนดำเนินการวิจัย

1. สร้างข้อมูลของตัวแปรสุ่ม X_1 และ X_2 ให้มีการแจกแจงตามที่กำหนด

2. กำหนดฟังก์ชันตามสถานการณ์กำหนด

2.1 ฟังก์ชันเคอร์เนลแบบเกาส์เซียน (Gaussian)

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\mathrm{u}^2\right)$$

2.2 ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov)

$$K(u) = \frac{3}{4}\left(1-\mathrm{u}^2\right)\mathrm{I}\left(|\mathrm{u}| \le 1\right)$$

3. ประมาณค่าแบนวิดจ์จากแบนวิดจ์ 2 วิธี ซึ่งแต่ละขั้นตอนมีรายละเอียด ดังนี้

3.1 สร้างข้อมูลของตัวแปรสุ่ม X_1 และ X_2 ให้มีการแจกแจงตามที่กำหนด ดังนี้

3.1.1 ผู้วิจัยจะสร้างข้อมูลของตัวแปรสุ่ม X ให้มีการแจกแจงและขนาดที่กำหนดคือ

3.1.1.1 กำหนดให้ $X_1 \sim N(0,1)$ และ $X_2 \sim N(0,1)$

3.1.1.2 กำหนดให้ $X_1 \sim Cau(0,1)$ และ $X_2 \sim Cau(0,1)$

3.1.2 สร้างฟังก์ชันของตัวแปรสุ่ม เนื่องจากกำหนดให้ฟังก์ชันของตัวแปรสุ่มมีลักษณะ $g(X_1, X_2, \ldots X_n) = \sum_{i=1}^{m} X_i$

3.2 กำหนดฟังก์ชันตามสถานการณ์ที่กำหนด ดังนี้

3.2.1 วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบิร์ทสเปซ ของ Ferraty (2007) เป็นฟังก์ชันเคอร์เนลแบบเกาส์เซียน

3.2.2 วิธีการประมาณค่าพารามิเตอร์การถดถอย โดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบิร์ทสเปซแบบปรับใหม่ เป็นฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ

ในแต่ละสถานการณ์ที่กำหนดในข้อที่ 3.2.1 สร้างตัวแปรสุ่มตามการแจกแจงแบบปกติมาตรฐาน

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2} \right], -\infty < x < \infty$$

สถานการณ์ที่กำหนดในข้อที่ 3.2.2 สร้างตัวแปรสุ่มตามการแจกแจงแบบโคชี (Cauchy Distribution)

$$\mathrm{f(x)} = \frac{1}{\pi\beta[1+(\frac{x-\theta}{\beta})^2]}, -\alpha < \theta < \alpha$$

4. ประมาณค่าแบนวิดจ์จากแบนวิดจ์ 2 วิธี ในข้อที่ 3.2.1 และ 3.2.2 ดังนี้

$$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$$

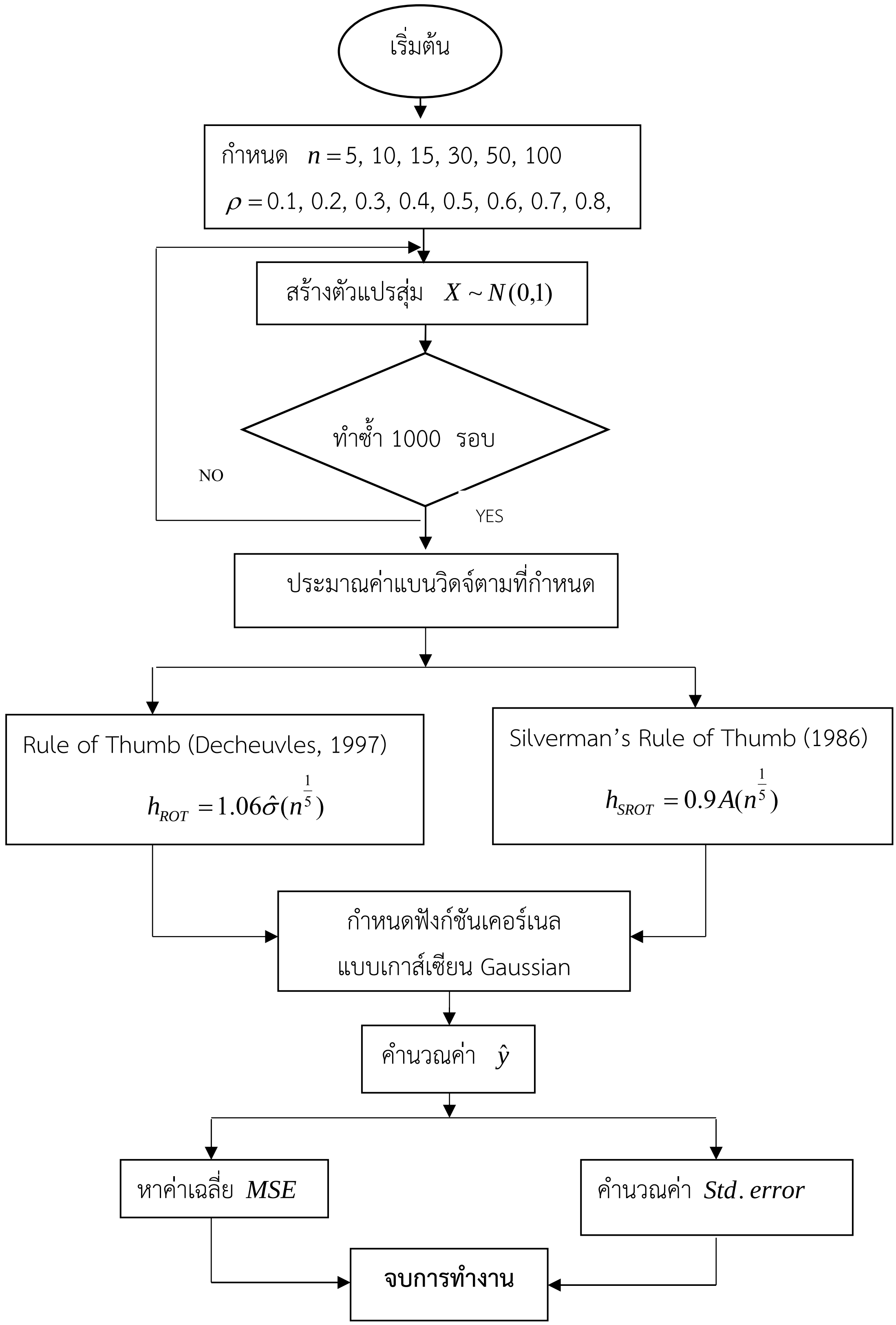
และ $$h_{SROT} = 0.9A(n^{\frac{1}{5}})$$

5. กำหนดค่าระดับความสัมพันธ์ของอัตตสหสัมพันธ์ (0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

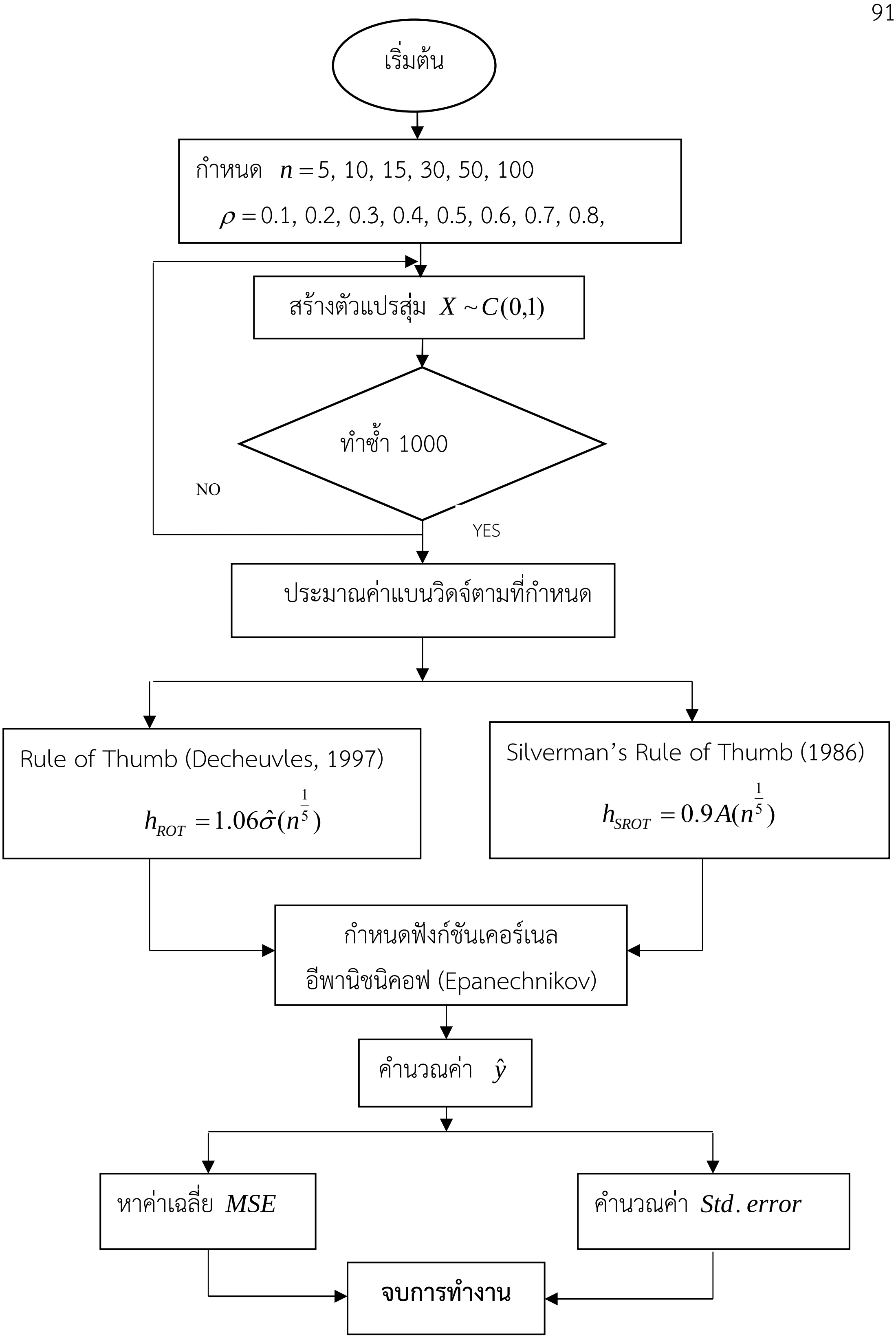
6. ทำซ้ำในแต่ละสถานการณ์จำนวน 1,000 ครั้งโดยการเก็บค่าพารามิเตอร์ในการถดถอยที่คำนวณได้จากวิธีการของรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบใหม่ที่พัฒนาขึ้น วิธีการรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) เพื่อนำไปคำนวณหาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Standard error: $Std.\,error$) ของแต่ละวิธีและในแต่ละสถานการณ์

7. วิเคราะห์หาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Standard Error: $Std.\,error$) ของทั้งสองวิธีในแต่ละสถานการณ์เพื่อนำแต่ละวิธีและแต่ละสถานการณ์ ซึ่งสามารถสรุปวิธีรวบรวมข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) ได้ดังภาพที่ 3-3 และ 3-4

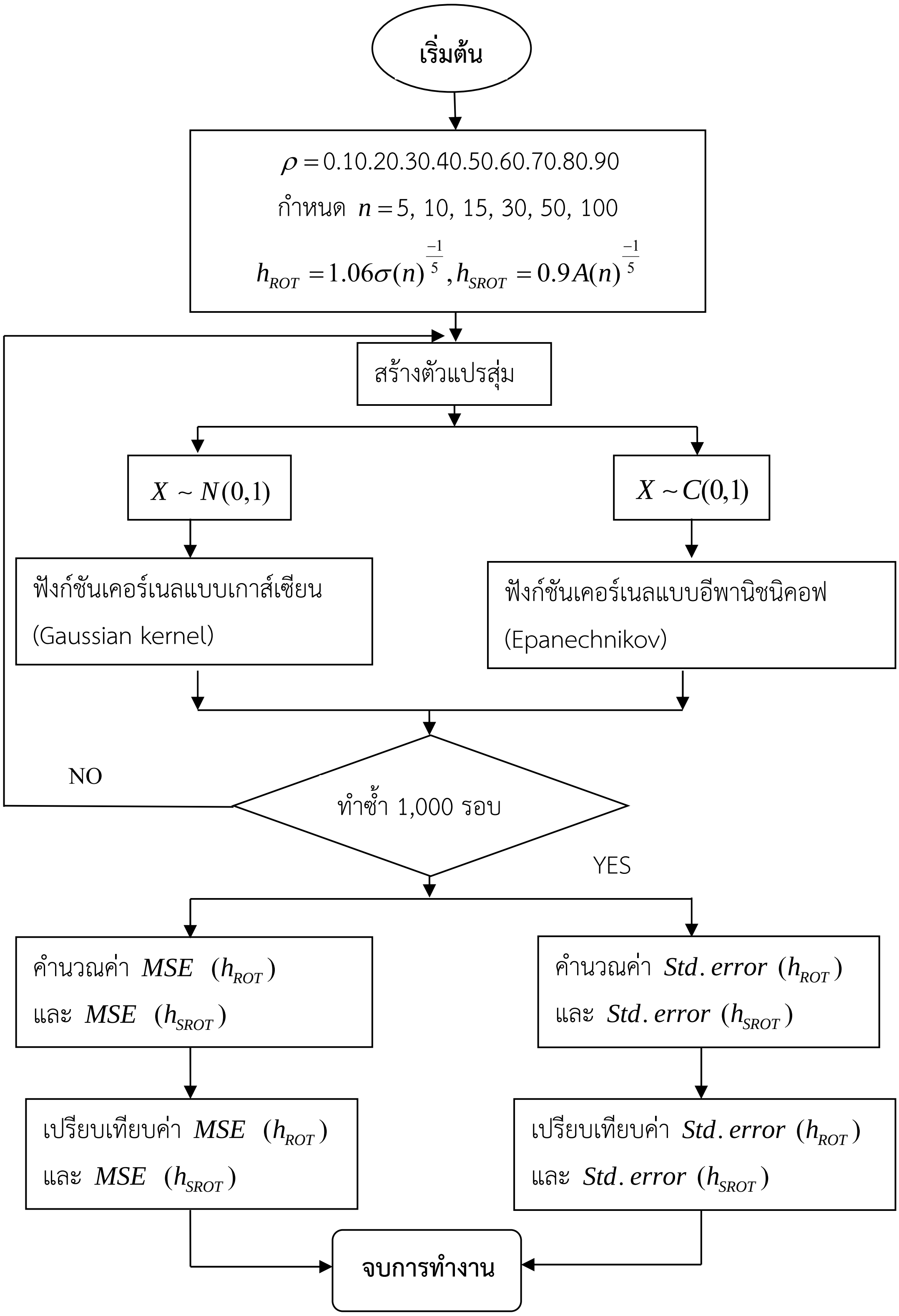
8. เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Standard error: $Std.\,error$) ของทั้งสองวิธีวิธีใดให้ค่า MSE และ $Std.\,error$ น้อยกว่าจะเป็นวิธีการประมาณค่าพารามิเตอร์การถดถอยในกรณีข้อมูลมีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์ที่มีประสิทธิภาพมากกว่า ดังภาพที่ 3-5



ภาพที่ 3-3 แผนผังของการพัฒนาวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)



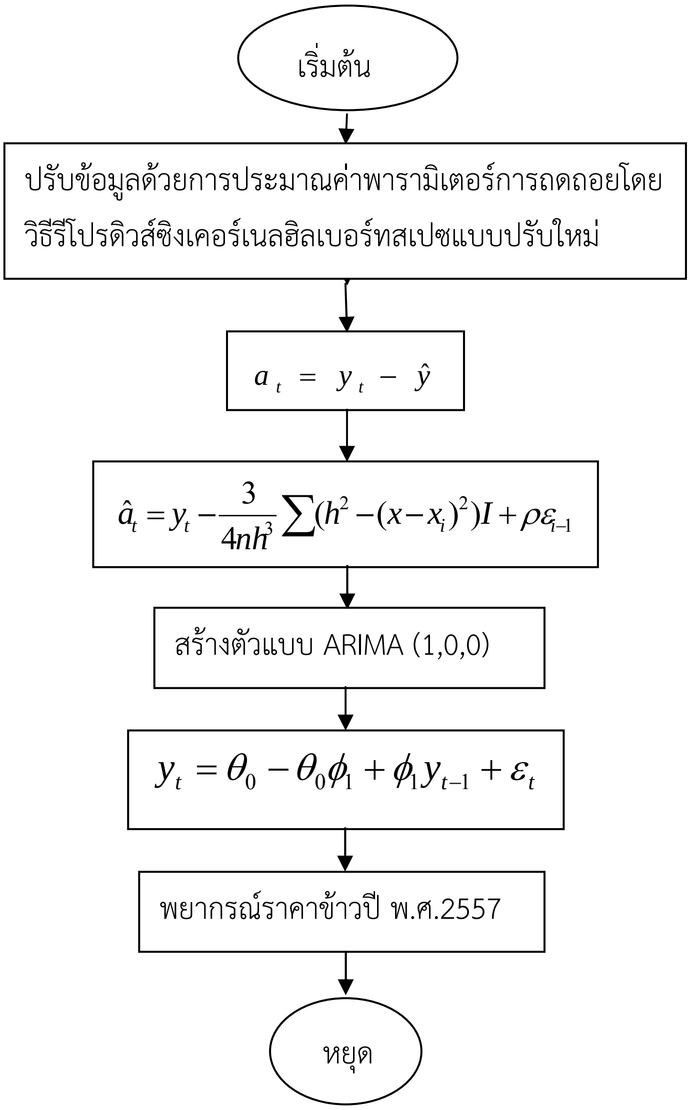
ภาพที่ 3-4 แผนผังของการพัฒนาวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่



ภาพที่ 3-5 แผนผังของการเปรียบเทียบวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่ กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)

ตอนที่ 3 การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA

การพยากรณ์ราคาของข้าวเปลือกเจ้าไทยโดยวิธีการประมาณค่าพารามิเตอร์การถดถอย โดยวิธี รีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ในขั้นตอนนี้ผู้วิจัยนำวิธีการประมาณค่าพารามิเตอร์ถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่มาประยุกต์ใช้กับการพยากรณ์ด้วยตัวแบบ ARIMA (1, 0, 0) ของราคาข้าวเปลือกเจ้า 15% ของไทย โดยเป็นข้อมูลทุติยภูมิ เพื่อเป็นแนวทางในการแก้ปัญหาของการเกิดความคลาดเคลื่อนของข้อมูลที่เกิดจากการพยากรณ์ข้อมูลผลผลิตทางการเกษตรล่วงหน้าเมื่อข้อมูลไม่เป็นเชิงเส้น มีขั้นตอน ดังนี้



ภาพที่ 3-6 แผนผังของการพัฒนาวิธีการพยากรณ์ราคาข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA

การพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยในปี พ.ศ. 2557 ด้วยตัวแบบ ARIMA มีวิธีการศึกษา ดังนี้

1. กลุ่มตัวอย่าง
2. เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล
3. วิธีการเก็บรวบรวมข้อมูล
4. ขั้นตอนและการวิเคราะห์ข้อมูล

1 กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นข้อมูลทุติยภูมิราคาข้าวเปลือกเจ้า 15% ปี พ.ศ. 2552-2556 (สำนักงานเศรษฐกิจการเกษตร, 2556) ที่ได้มีการจัดเก็บไว้แล้ว

2. เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล คือ แบบบันทึกข้อมูล (รายละเอียดในภาคผนวก ง) ประกอบด้วยราคาข้าวเปลือกเจ้าปี พ.ศ. 2552-2556 (สำนักงานเศรษฐกิจการเกษตร, 2556)

3. วิธีการเก็บรวบรวมข้อมูล

วิธีการดำเนินการเก็บรวบรวมข้อมูลราคาข้าวเปลือกเจ้า 15% มีวิธีดำเนินการ ดังนี้

3.1 ติดต่อสำนักงานเศรษฐกิจทางการเกษตร กระทรวงเกษตรและสหกรณ์ เพื่อขอข้อมูลราคาข้าวเปลือกเจ้า 15%

3.2 นำแบบบันทึกข้อมูลบันทึกข้อมูลราคาข้าวเปลือกเจ้า 15% ปี พ.ศ. 2552-2556 (สำนักงานเศรษฐกิจการเกษตร, 2556)

3.3 เก็บรวบรวมข้อมูลที่เกี่ยวกับปัจจัยต่าง ๆ ทั้งทางด้านการส่งเสริมการผลิต ด้านราคา และอุปสรรคที่มีต่อผลผลิต และราคา

3.4 นำข้อมูลที่เก็บรวบรวมได้มาดำเนินการวิเคราะห์ข้อมูล

4. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

วิเคราะห์อนุกรมเวลาด้วยวิธีของบ๊อกซ์และเจนกินส์ร่วมกับตัวแปรอิสระ (Andrews et al., 2013, pp. 6-49) โดยใช้ตัวแบบ ARIMA เมื่อปรับค่าพารามิเตอร์ด้วยฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov) โดยการแจกแจงแบบโคชี การดำเนินการวิเคราะห์ข้อมูลเป็นการวิเคราะห์หาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Standard error: $Std.\,error$)

4.1 วิเคราะห์หาค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE)

จากการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ล่วงหน้าโดยวิธีการประมาค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ คำนวณค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Square Error: MSE) ดังนี้

$$MSE(\hat{\mathrm{f}}(\mathrm{x})) = \frac{\sum_{i=1}^{1000} [\,\hat{f}_h(\mathrm{x}) - \mathrm{f}(\mathrm{x})]^2}{1000}$$

4.2 วิเคราะห์หาค่าความคลาดเคลื่อนมาตรฐาน (Standard Error: $Std.\,error$) จากการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ล่วงหน้าโดยวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่

คำนวณค่าความคลาดเคลื่อนมาตรฐาน (Standard Error: $Std.\,error$) ดังนี้

$$Std.\,error = \frac{\sigma}{\sqrt{n}}$$

4.3 คำนวณค่าสถิติพื้นฐาน

ค่าเฉลี่ย (mean)

$$\bar{x} = \frac{\sum x_i}{n}$$

ส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)

$$sd = \sqrt{\frac{\sum (\mathrm{x}_i - \bar{\mathrm{x}})^2}{n-1}}$$

4.4 พยากรณ์ราคาผลผลิตข้าวเปลือกเจ้า 15% ของไทยด้วยตัวแบบ ARIMA

4.5 นำผลการพยากรณ์ราคาผลผลิตข้าวเปลือกเจ้า 15% ของไทยด้วยตัวแบบ ARIMA ในข้อ 4.4 เทียบกับราคาผลผลิตข้าวเปลือกเจ้า 15% ของไทยในปี พ.ศ. 2557

ในการพยากรณ์ราคาผลผลิตข้าวเปลือกเจ้า 15% ของไทยในปี พ.ศ. 2557 ด้วยตัวแบบ ARIMA มีขั้นตอน ดังนี้

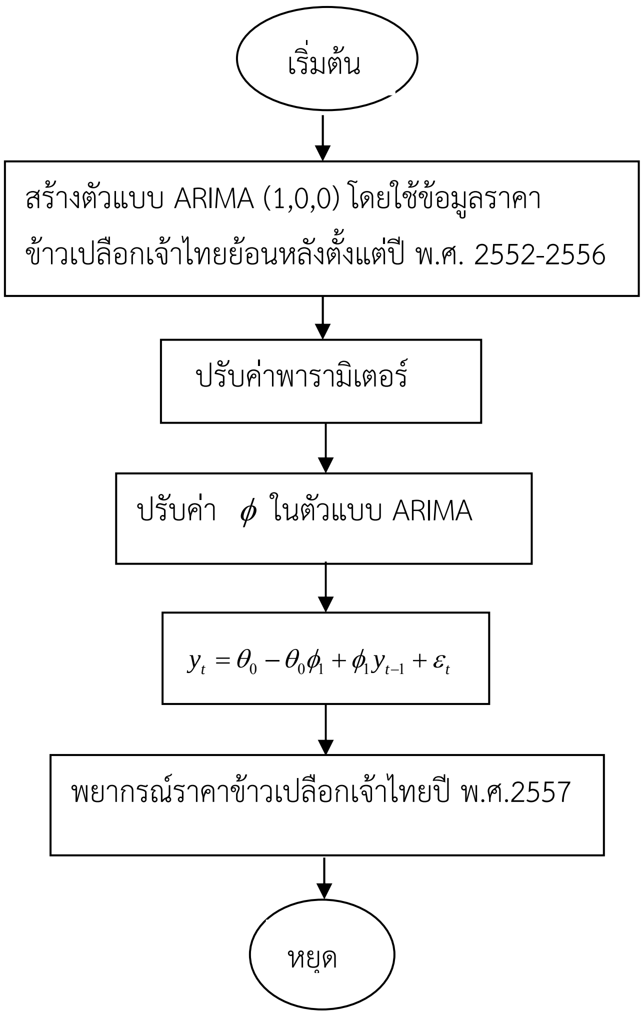
1. นำข้อมูลราคาข้าวเปลือกเจ้า 15% ของไทยตั้งแต่ปี พ.ศ. 2552-2556 ที่ปรับแล้วด้วยวิธีการประมาณค่าพารามิเตอร์การถดถอย โดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ สร้างตัวแบบ ARIMA (p,d,q)
2. สร้างตัวแบบ ARIMA (1,0,0)

3. ปรับค่าพารามิเตอร์

4. ได้สมการ $y_t = \theta_0 - \theta_0\phi_1 + \phi_1 y_{t-1} + \varepsilon_t$

5. พยากรณ์ราคาข้าวเปลือกเจ้าของไทย ปี พ.ศ. 2557

แสดงแผนผังขั้นตอนการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ. 2557 แสดงได้ดังแผนภาพที่ 3-7



ภาพที่ 3-7 แผนผังของขั้นตอนการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทย ปี พ.ศ. 2557

บทที่ 4
ผลการวิจัย

การวิจัยเรื่อง การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่ เสนอผลการวิเคราะห์ข้อมูลเป็น 3 ตอนดังนี้

ตอนที่ 1 ผลการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์

ตอนที่ 2 ผลการเปรียบเทียบประสิทธิภาพของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)

ตอนที่ 3 ผลการพยากรณ์ราคาข้าวเปลือกเจ้าของไทยล่วงหน้าด้วยตัวแบบ ARIMA

ในการนำเสนอผลส่วนนี้ กำหนดความหมายของคำย่อและสัญลักษณ์ ดังนี้

h	แทน	แบนวิดจ์ มีค่าเป็นจำนวนจริงบวก
h_{SROT}	แทน	แบนวิดจ์ที่ได้จากวิธี (Silverman’s Rules of Thumb)
h_{ROT}	แทน	แบนวิดจ์ที่ได้จากวิธี (Rules of Thumb)
$K(u)$	แทน	ฟังก์ชันเคอร์เนล (Kernel Function)
MSE	แทน	ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง
$Std.error$	แทน	ค่าความคลาดเคลื่อนมาตรฐาน
n	แทน	ขนาดตัวอย่าง
ρ	แทน	ระดับสัมประสิทธิ์สหสัมพันธ์
I	แทน	เมทริกเอกลักษณ์ที่มีมิติ nxn

ตอนที่ 1 ผลการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์

ในการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์

กำหนดสมการ $y = F(x) + \varepsilon_i$ ....(4.1)

เมื่อข้อมูลเป็น AR(1) ค่าความคลาดเคลื่อน $\varepsilon_i = \rho\varepsilon_{i-1} + a_i$ (SEBER & WILD, 1988, p. 18)

$$y = F(x) + \rho\varepsilon_{i-1} + a_i \quad (4.2)$$

$$F(x) = \frac{1}{nh}\sum K(u) \quad (4.3)$$

การพัฒนาวิธีการประมาณค่าแบบปรับใหม่โดยใช้ฟังก์ชันเคอร์เนล ของอีพานิชนิคอฟ (Epanechnikov Kernel Function)

เมื่อ $K(u) = \frac{3}{4}(1-u^2)I$ ....(4.4)

เมื่อ $u = \frac{x-x_i}{h}$ ....(4.5)

แทนค่า $u = \frac{x-x_i}{h}$ ลงในสมการที่ (4.4)

$$K(u) = \frac{3}{4}\left(1-\left(\frac{x-x_i}{h}\right)^2\right)I \quad (4.6)$$

$$= \frac{3}{4}\left(1-\frac{(x-x_i)^2}{h^2}\right)I \quad (4.7)$$

$$K(u) = \frac{3}{4}\left(\frac{h^2-(x-x_i)^2}{h^2}\right)I \quad (4.8)$$

แทนค่า $K(u)$ ในสมการที่ 4.8 ลงในสมการที่ 4.3

$$F(x) = \frac{1}{nh}\sum\frac{3}{4}\left(\frac{h^2-(x-x_i)^2}{h^2}\right)I \quad (4.9)$$

$$= \frac{3}{4nh}\sum\left(\frac{h^2-(x-x_i)^2}{h^2}\right)I \quad (4.10)$$

$$= \frac{3}{4nh}\sum\frac{1}{h^2}\left(h^2-(x-x_i)^2\right)I \quad (4.11)$$

$$F(x) = \frac{3}{4nh^3}\sum(h^2 - (x - x_i)^2)I \quad (4.12)$$

นำค่า $F(x)$ ในสมการที่ 4.12 แทนลงในสมการที่ 4.2 ได้สมการที่ปรับใหม่

$$\hat{y} = \frac{3}{4nh^3}\sum(h^2 - (x - x_i)^2)I + \rho\varepsilon_{i-1} + a_i \quad (4.13)$$

แบนวิดจ์ของฟังก์ชันเคอร์เนลที่ศึกษามี 2 วิธี

เมื่อ h เป็นค่าแบนวิดจ์ของฟังก์ชันเคอร์เนลที่ศึกษามี 2 วิธี

วิธีที่ 1 วิธี Rules of Thumb (Deheuvels, 1977, p. 36, Silverman, 1986, p. 44)

เมื่อ $$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}}) \quad (4.14)$$

แทนค่า h ในสมการที่ 4.14 ลงในสมการ (4.13)

$$\hat{y} = \frac{3}{4n(1.06\hat{\sigma}n^{\frac{1}{5}})^3}\sum(1.06\hat{\sigma}n^{\frac{1}{5}})^2 - (x - x_i)^2)I + \rho\varepsilon_{i-1} + a_i \quad (4.15)$$

$$= \frac{3}{4n(1.06)^3\hat{\sigma}^3 n^{\frac{3}{5}}}\sum((1.06)^2\hat{\sigma}^2 n^{\frac{2}{5}} - (x - x_i)^2)I + \rho\varepsilon_{i-1} + a_i \quad (4.16)$$

$$\hat{y} = \frac{3}{4(1.06)^3\hat{\sigma}^3 n^{\frac{8}{5}}}\sum((1.06)^2\hat{\sigma}^2 n^{\frac{2}{5}} - (x - x_i)^2)I + \rho\varepsilon_{i-1} + a_i \quad (4.17)$$

วิธีที่ 2 วิธี Silverman Rules of Thumb (Silverman, 1986, p. 44)

เมื่อ $$h_{SROT} = 0.9A(n^{\frac{1}{5}}) \quad (4.18)$$

แทนค่า h ในสมการที่ 4.18 ลงในสมการ (4.13)

$$\hat{y} = \frac{3}{4n(0.9\,\mathrm{A}n^{\frac{1}{5}})^3}\sum((0.9\,\mathrm{A}n^{\frac{1}{5}})^2 - (\mathrm{x} - \mathrm{x}_i)^2)I + \rho\varepsilon_{i-1} + a_i \quad (4.19)$$

$$= \frac{3}{4n(0.9)^3\,\mathrm{A}^3\mathrm{n}^{\frac{3}{5}}}\sum((0.9)^2\,\mathrm{A}^2\mathrm{n}^{\frac{2}{5}} - (\mathrm{x} - \mathrm{x}_i)^2)I + \rho\varepsilon_{i-1} + a_i \quad (4.20)$$

$$\hat{y} = \frac{3}{4n(0.9)^3\,\mathrm{A}^3\mathrm{n}^{\frac{3}{5}}}\sum((0.9)^2\,\mathrm{A}^2\mathrm{n}^{\frac{2}{5}} - (\mathrm{x} - \mathrm{x}_i)^2)I + \rho\varepsilon_{i-1} + a_i \quad (4.21)$$

ตอนที่ 2 ผลการเปรียบเทียบประสิทธิภาพของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)

การวิเคราะห์เปรียบเทียบประสิทธิภาพของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) โดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ภายใต้สถานการณ์ 108 สถานการณ์ (2x6X9) ได้แก่ แบนวิดจ์ 2 วิธี คือ วิธีที่ 1 วิธี Rules of Thumb วิธีที่ 2 วิธี Siliverman’s Rule of Thumb ขนาดตัวอย่างเท่ากับ 5, 10, 15, 30, 50 และ 100 และค่าสัมประสิทธิ์สหสัมพันธ์ (ρ) เท่ากับ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8 และ 0.9 โดยมีการทำซ้ำใน แต่ละสถานการณ์ 1,000 ครั้ง โดยใช้ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง และค่าความคลาดเคลื่อนมาตรฐาน เป็นเกณฑ์ในการเปรียบเทียบ

ผลการเปรียบเทียบประสิทธิภาพของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)

ตารางที่ 4-1 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) เมื่อ n เท่ากับ 5 ที่มีแบนวิดจ์แบบ Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

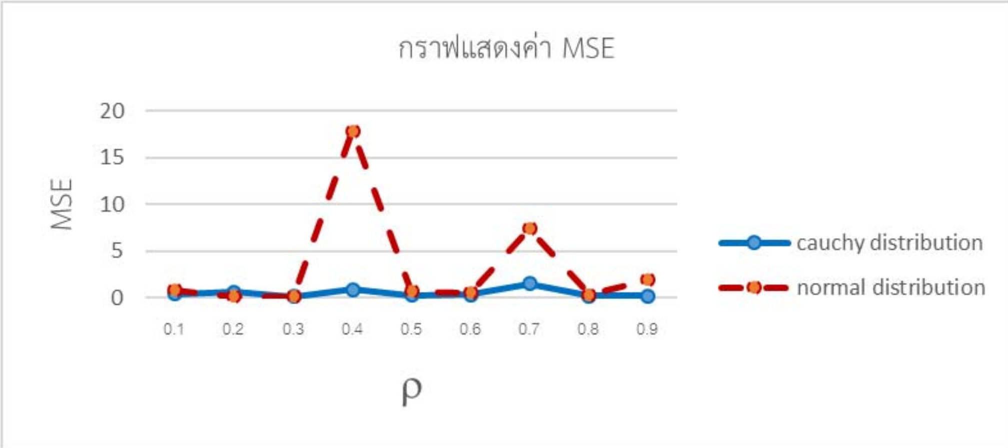
$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.1	0.3947371	0.5913857	0.7539949	0.5685084
0.2	0.6122232	0.7824469	0.1272428	0.3567111
0.3	0.09307431	0.3050808	0.06947197	0.2635754
0.4	0.8607768	0.9277806	17.87866	4.228316
0.5	0.2228366	0.4720557	0.6145895	0.7839576

ตารางที่ 4-1 (ต่อ)

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.\,error$	MSE	$Std.\,error$
0.6	0.3367094	0.5802667	0.4982718	0.7058837
0.7	1.482699	1.217661	7.411393	2.722387
0.8	0.1808493	0.4252638	0.3094456	0.5562783
0.9	0.1644902	0.4055739	1.89925	1.378133

จากตารางที่ 4-1 ผลการแสดงค่า MSE และ $Std.\,error$ เมื่อ $n = 5$ แบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ โดย $\rho = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE และค่า $Std.\,error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) โดยเมื่อการแจกแจงแบบโคชี (Cauchy) ที่ $\rho = 0.3$ มีค่าของ MSE ต่ำที่สุดที่ 0.09307431 และค่า $Std.\,error$ ต่ำที่สุดที่ 0.3050808 และที่ $\rho = 0.7$ จะมีค่า MSE และค่า $Std.\,error$ สูงที่สุด ที่ 1.482699 และ 1.217661 ตามลำดับ ยกเว้นกรณี $\rho = 0.2$-0.3 ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE และค่า $Std.\,error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ที่แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 5$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) 7 กรณี คือ $\rho = 0.1$, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9



ภาพที่ 4-1 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 5$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 5$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) 7 กรณี คือ $\rho = 0.1$, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

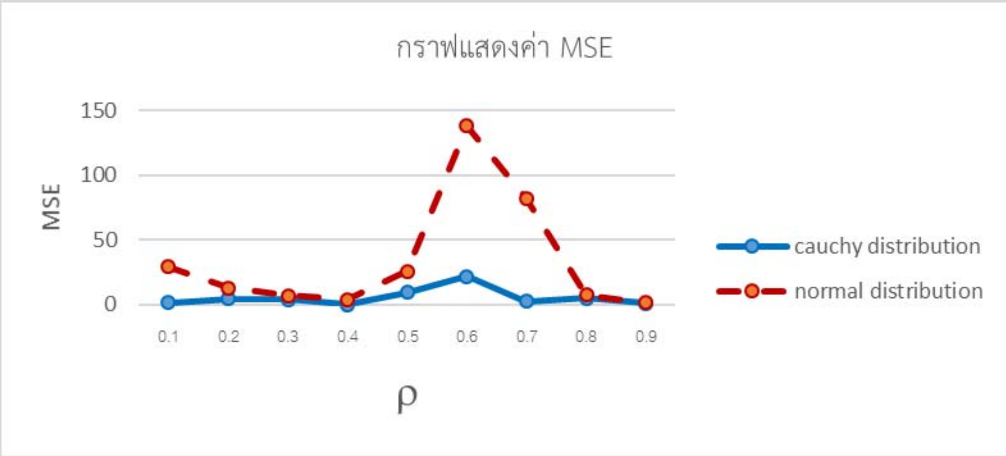
ภาพที่ 4-2 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 5$

ตารางที่ 4-2 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) เมื่อ n เท่ากับ 10 ที่มีแบนวิดจ์แบบ Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ การแจกแจงแบบแบบโคชี (Cauchy)		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.\,error$	MSE	$Std.\,error$
0.1	1.664089	1.289996	29.1842	5.402241
0.2	4.699224	2.167769	12.76805	3.573241
0.3	3.791083	1.94707	7.116949	2.667761
0.4	0.1735095	0.4165447	4.071698	2.017845
0.5	9.565195	3.092765	25.96916	5.095994
0.6	22.02206	4.692766	138.1992	11.75582
0.7	2.416365	1.554466	82.48132	9.081923
0.8	4.975977	2.23069	7.594851	2.755876
0.9	1.071894	1.035323	1.644086	1.282219

จากตารางที่ 4-2 ผลการแสดงค่า MSE และ $Std.\,error$ เมื่อ $n = 10$ แบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ โดย $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE และค่า $Std.\,error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบโคชี (Cauchy) ที่ $\rho =$ 0.4 มีค่าของ MSE ต่ำที่สุดที่ 0.1735095และค่า $Std.\,error$ ต่ำที่สุดที่ 0.4165447 และที่ $\rho = 0.6$ จะมีค่า MSE และค่า $Std.\,error$ สูงที่สุดที่ 22.02206 และ 4.692766 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 10$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



ภาพที่ 4-3 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 10$

กราฟแสดงการเปรียบเทียบค่า $Std.\,error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 10$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ

ภาพที่ 4-4 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 10$

ตารางที่ 4-3 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) เมื่อ n เท่ากับ 15 ที่มีแบนวิดจ์แบบ Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) การแจกแจงแบบปกติ (Normal)

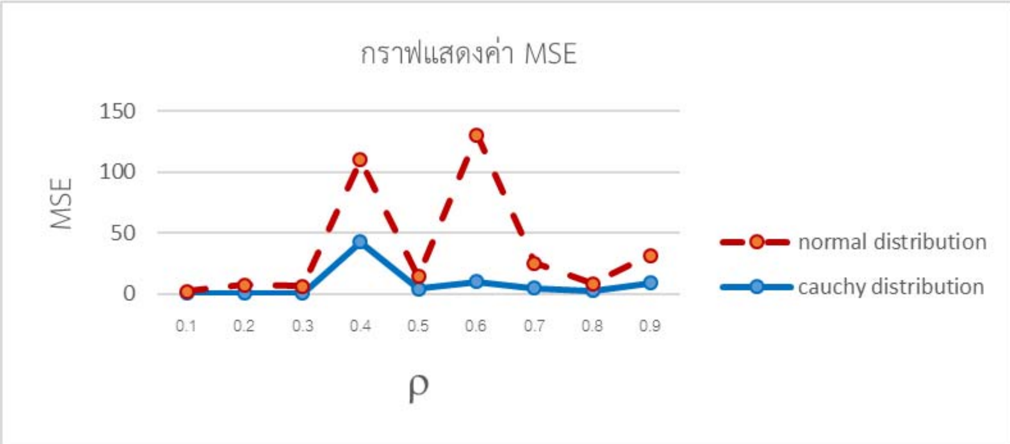
$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)	
ρ	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.1	0.6582165	0.8113054	1.566023	1.251408
0.2	0.8156667	0.9031427	6.762281	2.600439
0.3	0.5590166	0.7476742	5.918349	2.432766
0.4	42.49066	6.518486	67.64166	8.224455
0.5	4.391799	2.095662	9.736672	3.120364
0.6	10.10806	3.179317	120.4375	10.9744

ตารางที่ 4-3 (ต่อ)

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.7	4.685716	2.164651	20.63742	4.542843
0.8	2.956788	1.719531	5.384255	2.3204
0.9	9.062405	3.010383	22.45317	4.738477

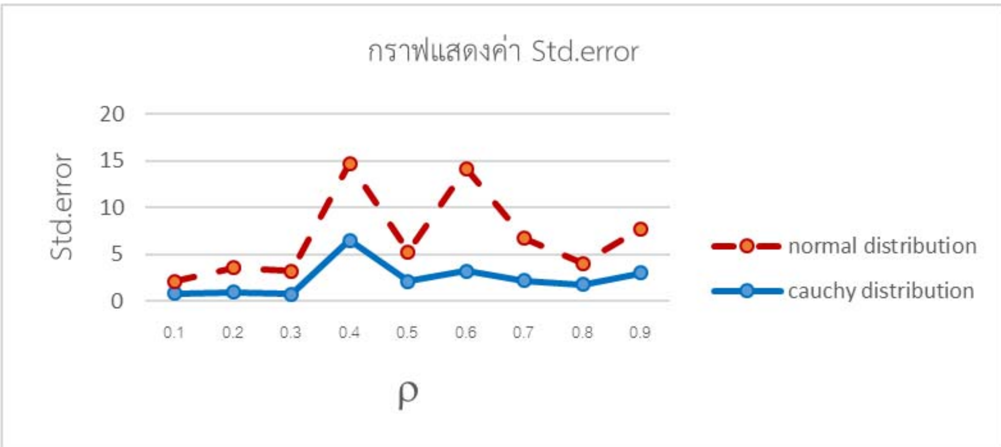
จากตารางที่ 4-3 ผลการแสดงค่า MSE และ $Std.error$ เมื่อ $n = 15$ แบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ โดย $\rho = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE และค่า $Std.error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบโคชี (Cauchy) ที่ $\rho = 0.3$ มีค่าของ MSE ต่ำที่สุดที่ 0.5590166 และค่า $Std.error$ ต่ำที่สุดที่ 0.7476742 และที่ $\rho = 0.4$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 42.49066 และ 6.518486 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 15$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



ภาพที่ 4-5 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 15$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 15$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนล ฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



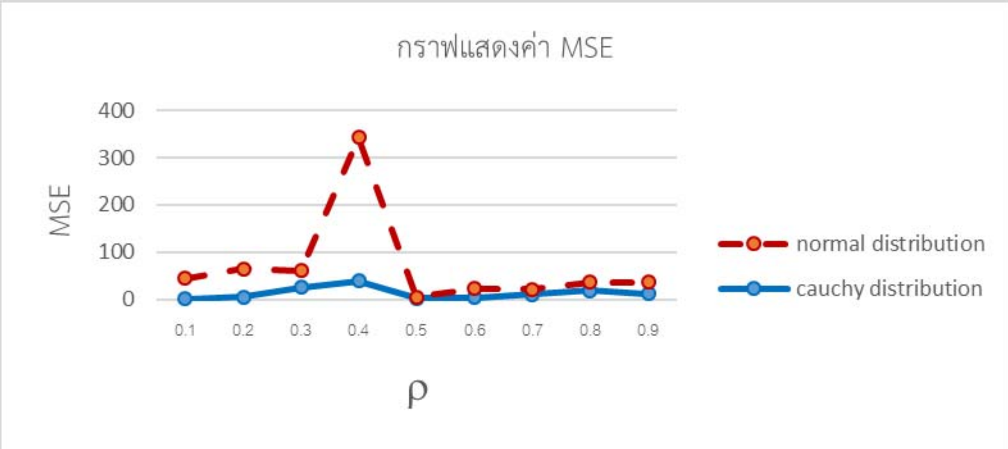
ภาพที่ 4-6 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 15$

ตารางที่ 4-4 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\, error$) เมื่อ n เท่ากับ 30 มีแบนวิดจ์แบบ Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.\, error$	MSE	$Std.\, error$
0.1	1.318778	1.14838	42.85186	6.546133
0.2	5.027516	2.242212	59.48774	7.71283
0.3	25.20197	5.020156	35.73703	5.978046
0.4	38.69058	6.220175	305.08	17.46654
0.5	2.02698	1.423721	3.165722	1.779248
0.6	3.326179	1.823781	18.92993	4.350854
0.7	10.89838	3.301269	10.72347	3.274672
0.8	17.84559	4.224404	18.21116	4.267454
0.9	11.23695	3.352156	24.21221	4.920591

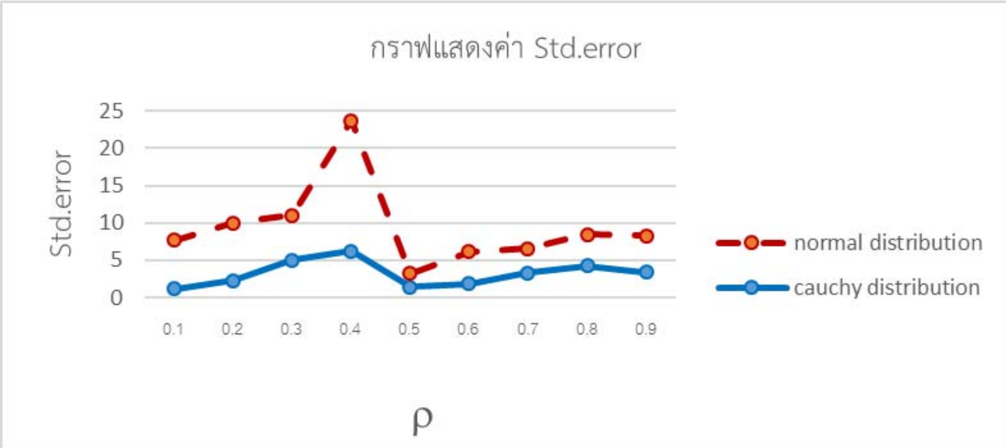
จากตารางที่ 4-4 ผลการแสดงค่า MSE และ $Std.\, error$ เมื่อ $n = 30$ แบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ โดย $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE และค่า $Std.\, error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบโคชี (Cauchy) ที่ $\rho = 0.1$ มีค่าของ MSE ต่ำที่สุดที่ 1.318778 และค่า $Std.\, error$ ต่ำที่สุดที่ 1.14838 และที่ $\rho = 0.4$ จะมีค่า MSE และค่า $Std.\, error$ สูงที่สุดที่ 38.69058 และ 6.220175 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 30$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



ภาพที่ 4-7 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 30$

กราฟแสดงการเปรียบเทียบค่า $Std.\,error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 30$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ $Std.\,error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



ภาพที่ 4-8 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 30$

ตารางที่ 4-5 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) เมื่อ n เท่ากับ 50 มีแบนวิดจ์แบบ Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

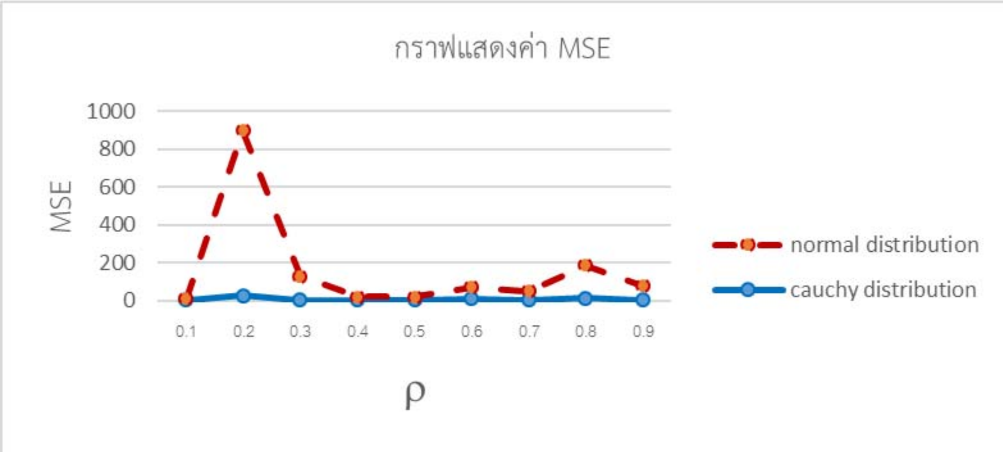
	$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$			
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.1	0.4523936	0.6726021	8.066216	2.840108
0.2	24.98548	4.998548	872.6413	29.5405
0.3	2.0848	1.443884	123.8391	11.1283
0.4	2.522758	1.588319	15.87558	3.984417
0.5	1.808578	1.344834	18.39301	4.288707
0.6	7.499222	2.738471	61.65818	7.852272

ตารางที่ 4-5 (ต่อ)

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.7	1.309465	1.144319	48.66674	6.976155
0.8	13.13541	3.62428	171.6599	13.1019
0.9	1.126489	1.061362	75.58655	8.694053

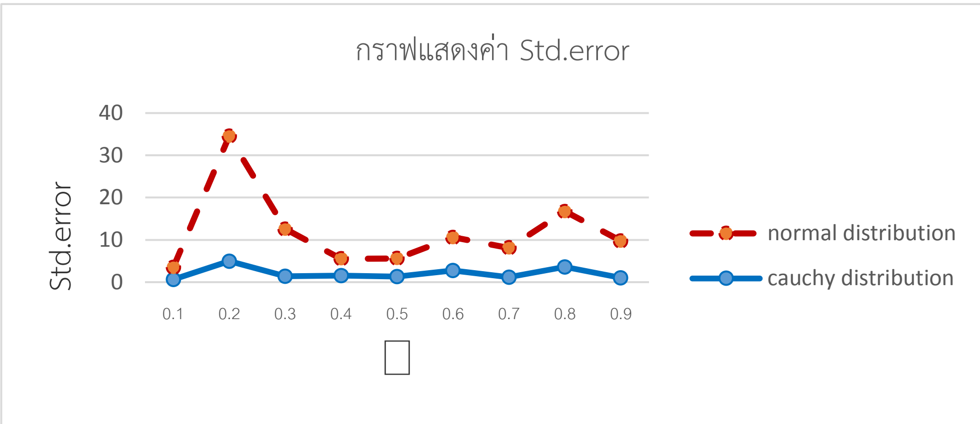
จากตารางที่ 4-5 ผลการแสดงค่า MSE และ $Std.error$ เมื่อ $n = 50$ แบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ โดย $\rho = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7,0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE และค่า $Std.error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบโคชี (Cauchy) ที่ $\rho = 0.1$ มีค่าของ MSE ต่ำที่สุดที่ 0.4523936 และค่า $Std.error$ ต่ำที่สุดที่ 0.6726021 และที่ $\rho = 0.2$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 24.98548 และ 4.998548 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 50$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



ภาพที่ 4-9 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 50$

กราฟแสดงการเปรียบเทียบค่า $Std.\,error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 50$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ $Std.\,error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



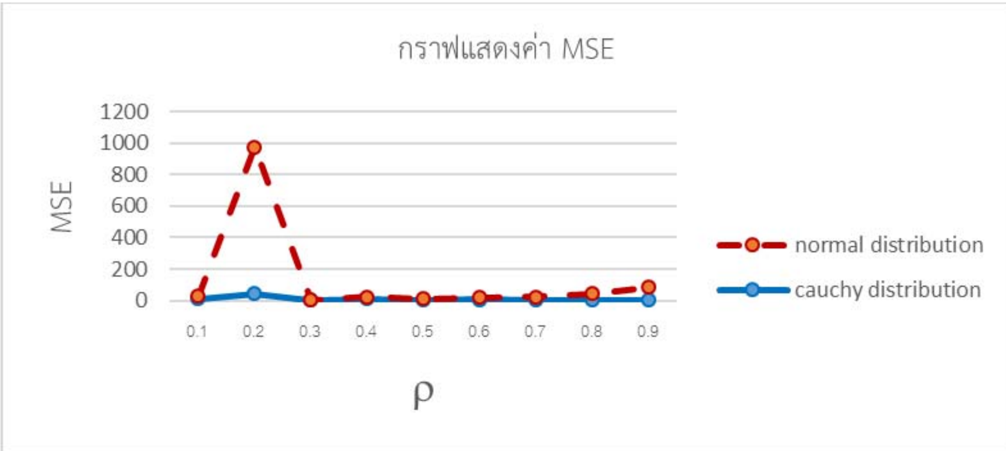
ภาพที่ 4-10 การเปรียบเทียบค่า $Std.\,error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 50$

ตารางที่ 4-6 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง MSE และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) เมื่อ n เท่ากับ 100 มีแบนวิดจ์แบบ Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.1	11.76728	3.430347	16.40854	4.050745
0.2	41.94124	6.476206	936.0035	30.59418
0.3	1.174519	1.083752	1.181845	1.087127
0.4	11.54988	3.398511	10.30073	3.209476
0.5	1.362921	1.167442	7.282934	2.698691
0.6	5.59004	2.364326	14.00966	3.742949
0.7	0.7780196	0.8820542	21.08249	4.591568
0.8	1.322509	1.150004	42.72618	6.536527
0.9	0.7656246	0.8749998	81.41796	9.02319

จากตารางที่ 4-6 ผลการแสดงค่า MSE และ $Std.error$ เมื่อ $n = 100$ แบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ โดย $\rho = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE และค่า $Std.error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบโคชี (Cauchy) ที่ $\rho = 0.9$ มีค่าของ MSE ต่ำที่สุดที่ 0.7656246 และค่า $Std.error$ ต่ำที่สุดที่ 0.8749998 และที่ $\rho = 0.2$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 41.94124 และ 6.476206 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 100$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ



ภาพที่ 4-11 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 100$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 100$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ในทุกค่าของ ρ

ภาพที่ 4-12 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 100$

จากการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล (Monte Carlo Technique) ของการหาประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 54 สถานการณ์ ได้แก่ การใช้แบนวิดจ์แบบ Rules of Thumb (h_{ROT}) ขนาดตัวอย่าง 6 ขนาด (n=5, 10, 15, 30, 50, 100) และค่าสัมประสิทธิ์สหสัมพันธ์ 9 ค่า (ρ=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

สรุปได้ว่า

1. วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี มีแบนวิดจ์แบบ h_{ROT} ใน 54 สถานการณ์ ปรากฏว่ามี 52 สถานการณ์ที่ n = 10, 15, 30, 50 และ 100 ในทุกค่าของ ρ ดังตารางที่ 4-2 ถึง 4-6 ที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ โดยเมื่อ $n = 5$ ดังตารางที่ 4-1 ปรากฏว่าวิธีการประมาณค่าพารามิเตอร์ การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ที่ ρ=0.1, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 ยกเว้นกรณีที่ ρ=0.2, 0.3 ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ฟังก์ชันเคอร์เนล

แบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี ซึ่งเป็นไปตามสมมุติฐานข้อที่ 1

2. วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ การแจกแจงแบบโคชี มีแบนวิดจ์แบบ h_{ROT} ใน 54 สถานการณ์ ปรากฏว่ามี 52 สถานการณ์ที่ n = 10, 15, 30, 50 และ 100 ในทุกค่าของ ρ ดังตารางที่ 4-2 ถึง 4-6 ที่มีค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ โดยเมื่อ $n = 5$ ดังตารางที่ 4-1 ปรากฏว่า วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ที่ $\rho = 0.1$, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 ยกเว้นกรณีที่ $\rho = 0.2$, 0.3 ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ มีค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ที่ใช้ฟังก์ชันเคอร์เนล แบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี ซึ่งเป็นไปตามสมมุติฐานข้อที่ 2

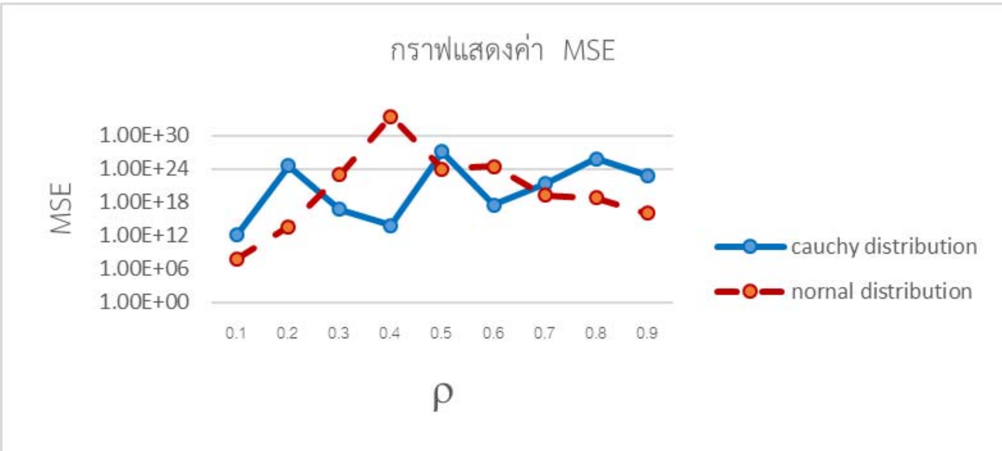
ตารางที่ 4-7 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) เมื่อ n เท่ากับ 5 ที่มีแบนวิดจ์แบบ Silverman's Rules of Thumb จำแนกตาม การแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

$h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.1	1.59x10^{12}	1.30x10^{6}	7.00x10^{7}	8.34x10^{3}
0.2	3.18x10^{24}	1.78x10^{12}	3.98x10^{13}	6.31x10^{6}
0.3	5.11x10^{16}	2.30x10^{8}	8.67x10^{22}	2.94x10^{11}
0.4	5.68x10^{13}	7.54x10^{6}	2.97x10^{33}	1.72x10^{12}
0.5	1.36x10^{27}	3.69x10^{13}	9.82x10^{23}	9.91x10^{11}
0.6	3.50x10^{17}	5.92x10^{8}	2.60x10^{24}	1.61x10^{12}
0.7	2.17x10^{21}	4.70x10^{10}	1.93x10^{19}	4.39x10^{9}
0.8	6.04x10^{25}	7.77x10^{12}	5.83x10^{18}	2.41x10^{9}
0.9	7.20x10^{22}	2.68x10^{11}	1.08x10^{16}	1.04x10^{8}

จากตารางที่ 4-7 ผลการแสดงค่า MSE และ $Std.error$ เมื่อ $n = 5$ แบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ โดย $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) มีค่าของ MSE และค่า $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) ที่ $\rho =$ 0.3, 0.4, 0.6 โดยเมื่อการแจกแจงแบบปกติ (Normal) ที่ $\rho =$ 0.1 มีค่าของ MSE ต่ำที่สุดที่ 7.00x10^{7} และค่า $Std.error$ ต่ำที่สุดที่ 8.34x10^{3} และที่ $\rho =$ 0.4 จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 2.97x10^{33} และ 1.72x10^{12} ตามลำดับ แต่ในกรณีที่

$\rho = 0.1, 0.2, 0.5, 0.7, 0.8, 0.9$ ที่ใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่า MSE และค่า $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 5$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) 6 กรณี คือ $\rho = 0.1, 0.2, 0.5, 0.7, 0.8, 0.9$



ภาพที่ 4-13 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 5$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 5$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) 6 กรณี คือ $\rho = 0.1, 0.2, 0.5, 0.7, 0.8, 0.9$

ภาพที่ 4-14 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.96\,A(n^{\frac{1}{5}})$ ตาม การแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n=5$

ตารางที่ 4-8 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อน มาตรฐาน ($Std.error$) เมื่อ n เท่ากับ 10 ที่มีแบนวิดจ์แบบ Silverman's Rules of Thumb จำแนกตาม การแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

$h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.1	3.42x10^{6}	1850.00	211.8464	1.46E+01
0.2	3.31x10^{5}	574.9508	2700.00	51.97183
0.3	2.72x10^{6}	1650.536	256.7443	16.02324
0.4	1.66x10^{6}	1289.882	31.60	5.617159
0.5	1.45x10^{6}	1202.444	73.20	8.554825
0.6	1.33x10^{7}	3648.124	437.00	20.91463

ตารางที่ 4-8 (ต่อ)

$h_{SROT} = 0.9A(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)	
ρ	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.7	1.03×10^{8}	10132.9	3.78×10^{5}	614.4602
0.8	2.76×10^{4}	166.075	2780.00	52.7158
0.9	4.19×10^{6}	2047.716	1480.00	38.40806

จากตารางที่ 4-8 ผลการแสดงค่า MSE และ $Std.error$ เมื่อ $n = 10$ แบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ โดย $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE และค่า $Std.error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบปกติ (Normal) ที่ $\rho = 0.4$ มีค่าของ MSE ต่ำที่สุดที่ 31.60 และค่า $Std.error$ ต่ำที่สุดที่ 5.61759 และที่ $\rho = 0.7$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 3.78×10^{5} และ 614.4602 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 10$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ

ภาพที่ 4-15 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 10$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 10$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ

ภาพที่ 4-16 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 10$

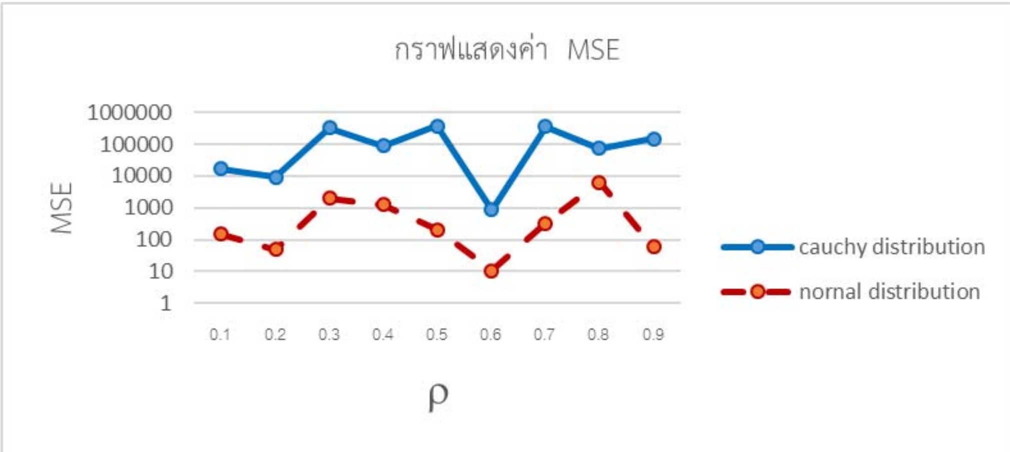
ตารางที่ 4-9 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) เมื่อ n เท่ากับ 15 ที่มีแบนวิดจ์แบบ Silverman's Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

$h_{SROT} = 0.9A(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.\,error$	MSE	$Std.\,error$
0.1	16790.5	129.5782	151.5478	12.31048
0.2	9.05×10^3	95.13253	48.20	6.945996
0.3	3.26×10^5	571.2711	1960.00	44.2269
0.4	8.98×10^4	299.5972	1270.00	35.65924
0.5	3.83×10^5	618.9955	204.00	14.267
0.6	8.86×10^2	29.76327	10.30	3.216892
0.7	3.67×10^5	605.7881	333.00	18.23463
0.8	7.47×10^4	273.2994	6280.00	79.21766
0.9	1.47×10^5	383.9889	59.80	7.733792

จากตารางที่ 4-9 ผลการแสดงค่า MSE และ $Std.\,error$ เมื่อ $n = 15$ แบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ โดย $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE และค่า $Std.\,error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบปกติ (Normal) ที่ $\rho = 0.6$ มีค่าของ

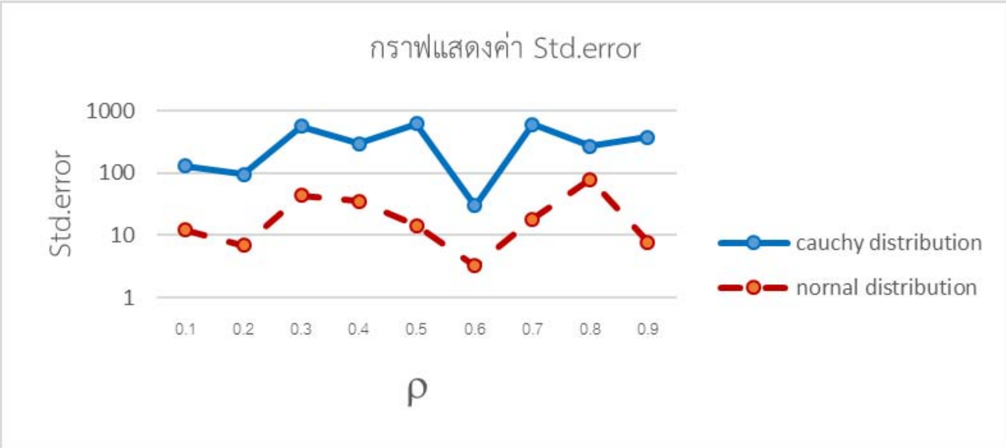
MSE ต่ำที่สุดที่ 10.30 และค่า $Std.error$ ต่ำที่สุดที่ 3.216892 และที่ $\rho = 0.8$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 6280.00 และ 79.21766 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 15$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ



ภาพที่ 4-17 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 15$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 15$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ



ภาพที่ 4-18 การเปรียบเทียบค่า $Std.\,error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 15$

ตารางที่ 4-10 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) เมื่อ n เท่ากับ 30 ที่มีแบนวิดจ์แบบ Silverman’s Rules of Thumb จำแนกตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal)

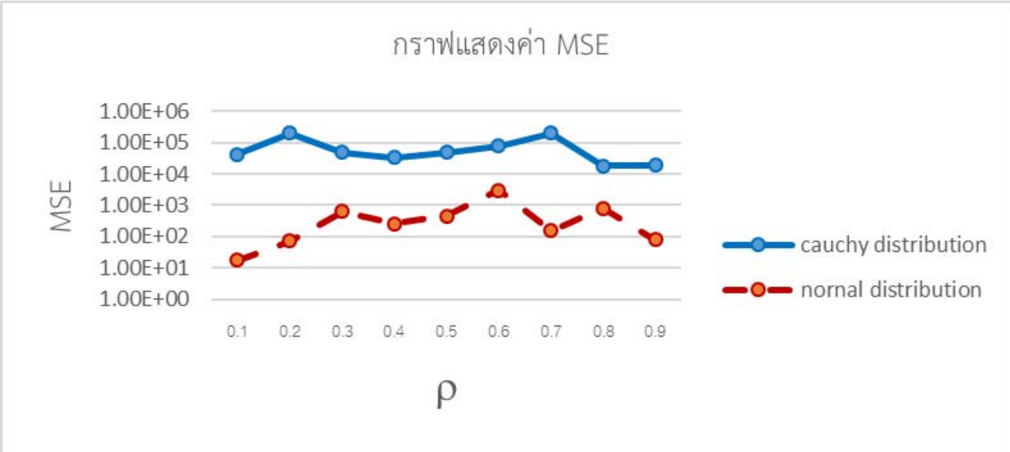
	$h_{SROT} = 0.9A(n^{\frac{1}{5}})$			
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.\,error$	MSE	$Std.\,error$
0.1	42558.58	206.2973	17.51259	4.184805
0.2	2.05×10^5	452.7375	73.70	8.587233
0.3	5.01×10^4	223.7958	637.00	25.23162
0.4	3.43×10^4	185.2277	264.00	16.24477
0.5	4.96×10^4	222.6261	464.00	21.53365

ตารางที่ 4-10 (ต่อ)

$h_{SROT} = 0.9A(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
ρ	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (Normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.6	7.74×10^4	278.2094	3010.00	54.89506
0.7	1.99×10^5	446.2979	155.00	12.4386
0.8	1.83×10^4	135.1675	793.00	28.1517
0.9	1.93×10^4	139.0552	80.60	8.979027

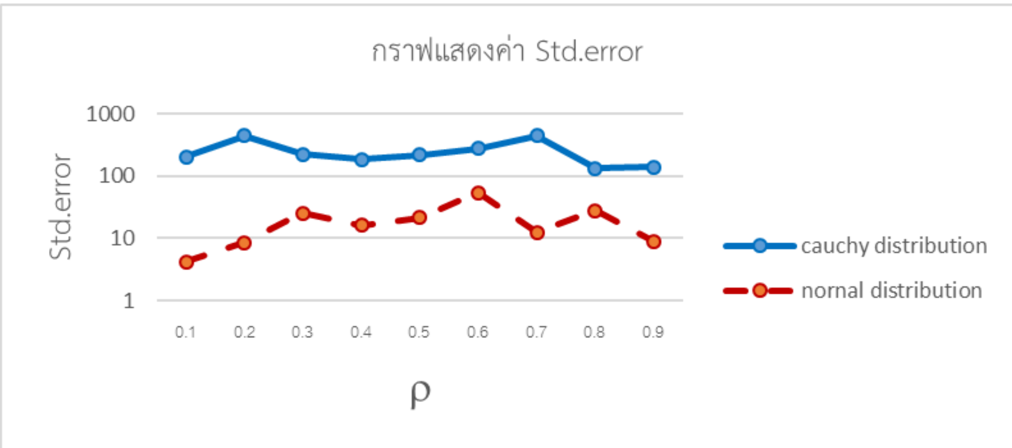
จากตารางที่ 4-10 ผลการแสดงค่า MSE และ $Std.error$ เมื่อ $n = 30$ แบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ โดย $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE และค่า $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบปกติ (Normal) ที่ $\rho = 0.2$ มีค่าของ MSE ต่ำที่สุดที่ 73.70 และค่า $Std.error$ ต่ำที่สุดที่ 8.587233 และที่ $\rho = 0.6$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 3010.00 และ54.89506 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 30$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ



ภาพที่ 4-19 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 30$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 30$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ



ภาพที่ 4-20 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 30$

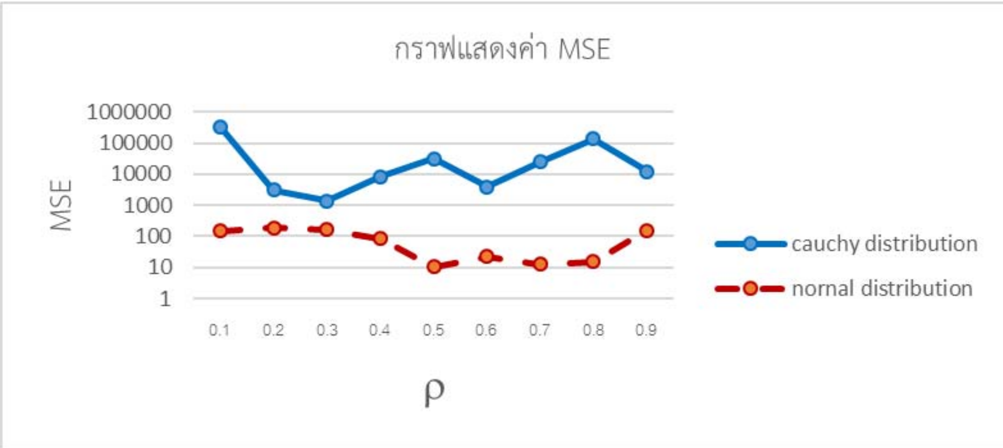
ตารางที่ 4-11 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) เมื่อ n เท่ากับ 50 ที่มีแบนวิดจ์แบบ Silverman’s Rules of Thumb จำแนกตามการแจกแจงแบบโคซี (Cauchy) และการแจกแจงแบบปกติ (Normal)

$h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
ρ	การแจกแจงแบบแบบโคซี (Cauchy)		การแจกแจงแบบปกติ (normal)	
	MSE	$Std.\ error$	MSE	$Std.\ error$
0.1	318×10^5	564.1293	150.5159	12.26849
0.2	3.14×10^3	56.0247	185.00	13.58648
0.3	1.38×10^3	37.11695	164.00	12.79659
0.4	8.55×10^3	92.48325	82.60	9.087965
0.5	3.12×10^4	176.5048	10.30	3.21155
0.6	3.91×10^3	62.56735	22.50	4.743227
0.7	2.49×10^4	157.8609	12.90	3.585561
0.8	1.38×10^5	371.2015	15.20	3.904113
0.9	1.21×10^4	109.8761	156.00	12.4841

จากตารางที่ 4-11 ผลการแสดงค่า MSE และ $Std.\ error$ เมื่อ $n = 50$ แบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ โดย $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE และค่า $Std.\ error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคซี (Cauchy) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบปกติ (Normal) ที่ $\rho = 0.5$

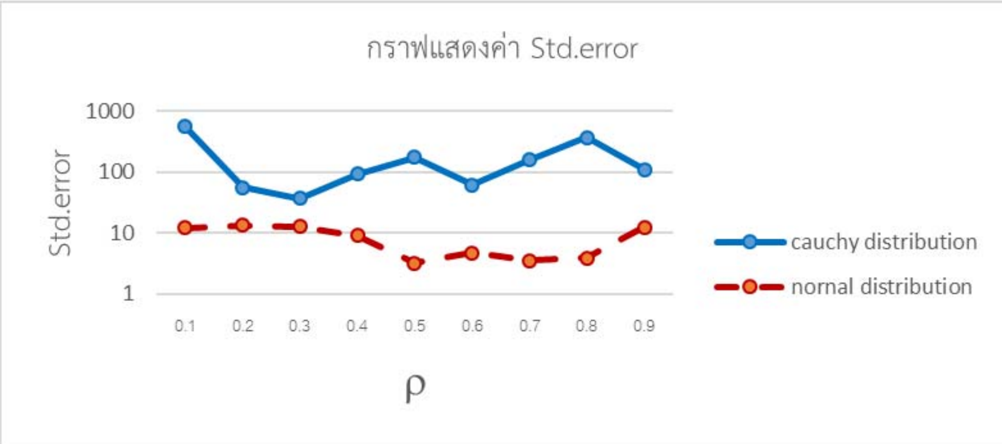
มีค่าของ MSE ต่ำที่สุดที่ 10.30 และค่า $Std.error$ ต่ำที่สุดที่ 3.21155 และที่ $\rho = 0.2$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 185.00 และ 13.58648 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 50$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ



ภาพที่ 4-21 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 50$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 50$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ



ภาพที่ 4-22 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 50$

ตารางที่ 4-12 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) เมื่อ n เท่ากับ 100 ที่มีแบนวิดจ์แบบ Silverman's Rules of Thumb จำแนกตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy)

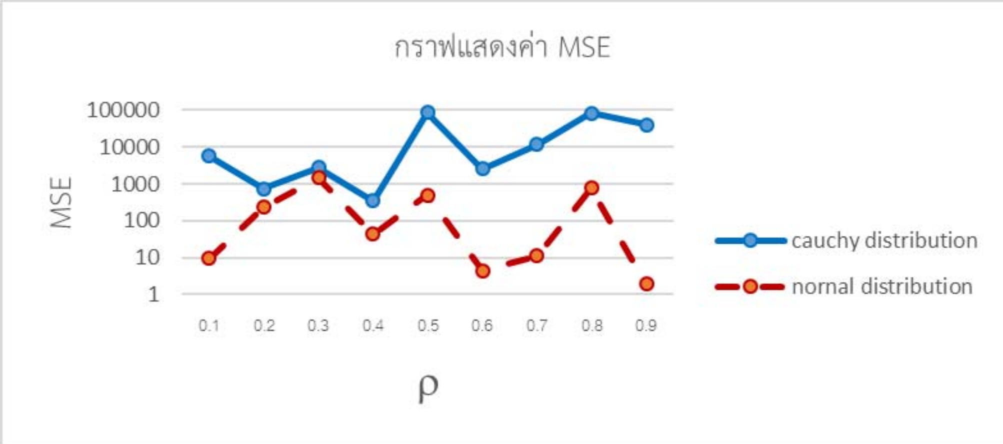
$h_{SROT} = 0.9A(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์ ρ	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)	
	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.1	5674.845	75.33157	9.59208	3.097108
0.2	$7.35\text{x}10^{2}$	27.11286	232.00	15.22515
0.3	$2.77\text{x}10^{3}$	52.62469	1460.00	38.18212
0.4	$3.41\text{x}10^{2}$	18.46832	43.70	6.610801
0.5	$8.63\text{x}10^{4}$	293.692	496.00	22.27686

ตารางที่ 4-12 (ต่อ)

$h_{SROT} = 0.9A(n^{\frac{1}{5}})$				
ค่าสัมประสิทธิ์	วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่		วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007)	
ρ	การแจกแจงแบบแบบโคชี (Cauchy)		การแจกแจงแบบปกติ (normal)	
	MSE	$Std.error$	MSE	$Std.error$
0.6	2.54x10^{3}	50.41717	4.36	2.087637
0.7	1.18x10^{4}	108.7378	11.30	3.357755
0.8	8.11x10^{4}	284.8054	772.00	27.78739
0.9	3.97x10^{4}	199.3697	1.88	1.372823

จากตารางที่ 4.12 ผลการแสดงค่า MSE และ $Std.error$ เมื่อ $n = 100$ แบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ โดย $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 พบว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE และค่า $Std.error$ ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ โดยเมื่อการแจกแจงแบบปกติ (Normal) ที่ $\rho = 0.6$ มีค่าของ MSE ต่ำที่สุดที่ 4.36 และค่า $Std.error$ ต่ำที่สุดที่ 2.087637 และที่ $\rho = 0.3$ จะมีค่า MSE และค่า $Std.error$ สูงที่สุดที่ 1460.00 และ 38.18212 ตามลำดับ แสดงได้ดังกราฟ

กราฟแสดงการเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 100$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ MSE ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ



ภาพที่ 4-23 การเปรียบเทียบค่า MSE กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 100$

กราฟแสดงการเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy) กรณี $n = 100$ พบว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ (Normal) มีค่าของ $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี (Cauchy) ในทุกค่าของ ρ

ภาพที่ 4-24 การเปรียบเทียบค่า $Std.error$ กรณีแบนวิดจ์แบบ $h_{SROT} = 0.9A(\mathrm{n}^{\frac{1}{5}})$ ตามการแจกแจงแบบโคชี (Cauchy) และการแจกแจงแบบปกติ (Normal) กรณี $n = 100$

จากการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล (Monte Carlo Technique) ของการหาประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 54 สถานการณ์ ได้แก่ การใช้แบนวิดจ์แบบ Silverman's Rules thumb (h_{SROT}) ขนาดตัวอย่าง 6 ขนาด (n=5, 10, 15, 30, 50, 100) และค่าสัมประสิทธิ์สหสัมพันธ์ 9 ค่า (ρ=0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9)

สรุปได้ว่า

1. วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ มีแบนวิดจ์แบบ h_{SROT} ใน 54 สถานการณ์ ปรากฏว่ามี 51 สถานการณ์ที่ n = 10, 15, 30, 50 และ 100 ในทุกค่าของ ρ ดังตารางที่ 4-8 ถึง 4-12 ที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี โดยเมื่อ $n=5$ ดังตารางที่ 4-7 ปรากฏว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ต่ำกว่าวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ที่ $\rho=0.1$, 0.4, 0.5, 0.7, 0.8, 0.9 ยกเว้นกรณีที่ $\rho=0.2$, 0.3, 0.6 ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบ ปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) ต่ำกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ ซึ่งเป็นไปตามสมมุติฐานข้อที่ 1

2. วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ มีแบนวิดจ์แบบ h_{SROT} ใน 54 สถานการณ์ ปรากฏว่ามี 51 สถานการณ์ที่ n = 10, 15, 30, 50 และ 100 ในทุกค่าของ ρ ดังตารางที่ 4-8 ถึง 4-12 ที่มีค่าเฉลี่ยความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี โดยเมื่อ $n=5$ ดังตารางที่ 4-7 ปรากฏว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ที่ มีค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ต่ำกว่า วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิง

เคอร์เนลฮิลเบอร์ทสเปซแบบ ปรับใหม่ที่ $\rho = 0.1, 0.4, 0.5, 0.7, 0.8, 0.9$ ยกเว้นกรณีที่ $\rho = 0.2, 0.3, 0.6$ ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชีที่มีค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ ซึ่งเป็นไปตามสมมุติฐานข้อที่ 2

ตอนที่ 3 ผลการการพยากรณ์ราคาข้าวเปลือกเจ้าของไทยล่วงหน้าด้วยตัวแบบ ARIMA

จากผลการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ในตอนที่ 2 ปรากฏว่าที่วิธีการประมาณค่าพารามิเตอร์การถดถอยด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่มีการแจกแจงแบบโคชี และแบนวิดจ์แบบ Rule of thumb (h_{ROT}) ที่ $n = 50$ และ $\rho = 0.2$ มีค่า MSE และค่า $Std.error$ ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่มีการแจกแจงแบบปกติ ซึ่งที่ค่า $\rho = 0.2$ เป็นค่าสหสัมพันธ์ที่สอดคล้องกับข้อมูลจริงของราคาข้าวเปลือกเจ้า 15% ของไทย ดังภาคผนวกที่ ง ที่ใช้เป็นข้อมูลย้อนหลัง 5 ปีในการสร้างตัวแบบการพยากรณ์ จึงใช้วิธีการประมาณค่าพารามิเตอร์การถดถอยด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่

ผลของพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยโดยใช้วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ จากข้อมูลราคาข้าวย้อนหลัง 5 ปี (มกราคมปี พ.ศ. 2552 ถึง เดือนธันวาคม พ.ศ. 2556) ได้ผลการวิเคราะห์โดยตัวแบบ ARIMA (p,d,q) มีสมบัติทั่วไป คือ ARIMA (1, 0, 0) ดังนี้

1. ประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ ด้วยสมการ

$$\hat{y} = \frac{3}{4nh^3} \sum (h^2 - (x - x_i)^2) I + \rho \varepsilon_{i-1} + a_i$$

2. พยากรณ์ราคาผลผลิตข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าด้วยตัวแบบ ARIMA จากข้อมูลจริงของราคาข้าวเปลือกเจ้า 15% ของไทย ที่ใช้ข้อมูลย้อนหลัง 5 ปีในการสร้างตัวแบบการพยากรณ์ ดังภาคผนวก ง ได้ผลการวิเคราะห์โดยตัวแบบ ARIMA (p, d, q) สามารถสร้างตัวแบบพยากรณ์ได้ 4 ตัวแบบ คือ ARIMA (2, 0, 0), ARIMA (0, 0, 2), ARIMA (0, 0, 1)

และ ARIMA (1, 0, 0) ปรากฏว่า ตัวแบบ ARIMA (1, 0, 0) ให้ค่า R-Squared สูงที่สุดที่ 0.603 จึงใช้ตัวแบบ ARIMA (1, 0, 0) เป็นตัวแบบการพยากรณ์ มีสมการดังนี้

สมการทั่วไป $(1-\phi_1 B)(y_t-\theta_0)=\varepsilon_t$

$$y_t-\theta_0-\phi_1 B y_t+\theta_0\phi_1 B=\varepsilon_t$$

$$y_t-\theta_0-\phi_1 y_{t-1}+\theta_0\phi_1 B=\varepsilon_t$$

$$y_t=\theta_0-\theta_0\phi_1+\phi_1 y_{t-1}+\varepsilon_t$$

เมื่อ y_t คือ ค่าของข้อมูลอนุกรมเวลาที่ต้องการพยากรณ์ ณ เวลาที่ t

y_{t-1} คือ ค่าของข้อมูลอนุกรมเวลาที่ต้องการพยากรณ์ ณ เวลาที่ $t-1$

B คือ ตัวดำเนินการย้อนหลังเวลา (Backward Shift Operator)

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลาที่ t

θ_0 คือ พารามิเตอร์แสดง Constant Term

ϕ_1 คือ พารามิเตอร์แสดงสัมประสิทธิ์การถดถอยในตัวเองแบบ Nonseasonal อันดับที่ 1 (Nonseasonal Auto-Regressive Process of Order 1)

เมื่อแทนค่า $\hat{\theta}_0=9465.825$ $\hat{\phi}_1=0.848$

$$\hat{y}_t=9465.825-9465.825x0.848+0.848y_{t-1}+\varepsilon_t$$

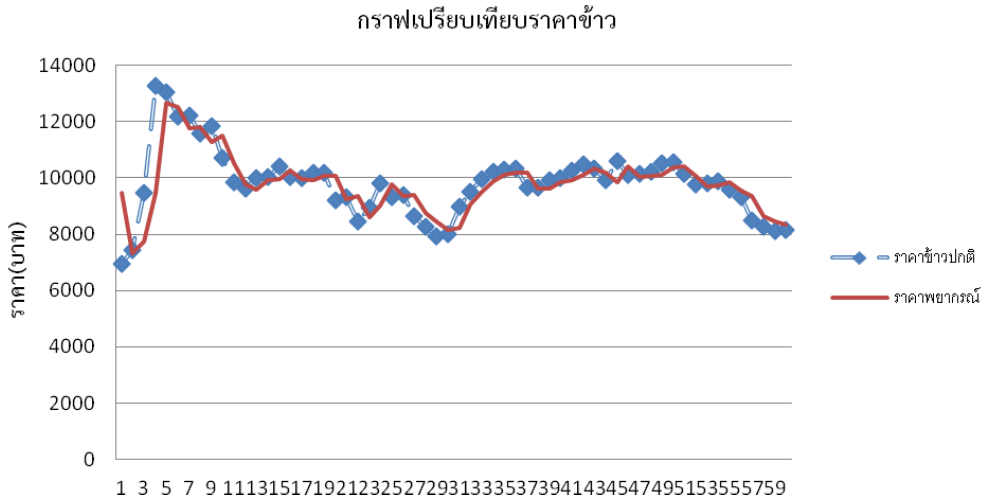
ได้สมการ $\hat{y}_t=1438.8054+0.848y_{t-1}+\varepsilon_t$

3. นำผลการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยที่ได้จากตัวแบบ ARIMA (p, d, q) ที่มีสมบัติทั่วไปคือ ARIMA (1, 0, 0) ไปเทียบกับราคาข้าวเปลือกเจ้าของไทยปี พ.ศ. 2557 ดังตารางที่ 4-14

ตารางที่ 4-13 ผลการเปรียบเทียบราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ. 2552-2556 กับค่าพยากรณ์ราคาด้วยตัวแบบ ARIMA(1, 0, 0)

ราคาข้าวจริง (บาท)	ราคาข้าวพยากรณ์ (บาท)	ราคาข้าวจริง (บาท)	ราคาข้าวพยากรณ์ (บาท)	ราคาข้าวจริง (บาท)	ราคาข้าวพยากรณ์ (บาท)
6955	9257.58	9318	9209.3	10242	9918.995
7439	7327.02	8460	9308.24	10467	10124.3
9453	7732.83	8948	8588.86	10326	10315.18
13259	9421.43	9825	8998.02	9914	10195.56
13042	12612.5	9326	9733.32	10584	9846.037
12162	12430.56	9400	9314.95	10123	10414.43
12203	11692.74	8636	9376.99	10160	10023.34
11592	11727.11	8257	8736.43	10221	10054.73
11857	11214.83	7916	8418.66	10526	10106.48
10692	11437.02	7985	8132.76	10559	10365.23
9843	10460.24	8993	8209.559	10158	10393.23
9618	9748.41	9523	9064.701	9753	10053.04
10005	9559.77	9949	9514.33	9799	9709.451
10029	9884.24	10217	9875.729	9897	9748.476
10392	9904.36	10305	10103.09	9578	9831.614
10047	10208.71	10322	10177.74	9329	9560.989
10003	9919.45	9641	10192.17	8480	9349.749
10192	9882.56	9645	9614.435	8280	8629.495
12182	10041.03	9921	9617.829	8126	8459.824
9200	10032.64	10000	9851.975	8134	8329.177

กราฟแสดงการเปรียบเทียบราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ. 2552-2556 กับค่าพยากรณ์ราคาด้วยตัวแบบ ARIMA (1, 0, 0)

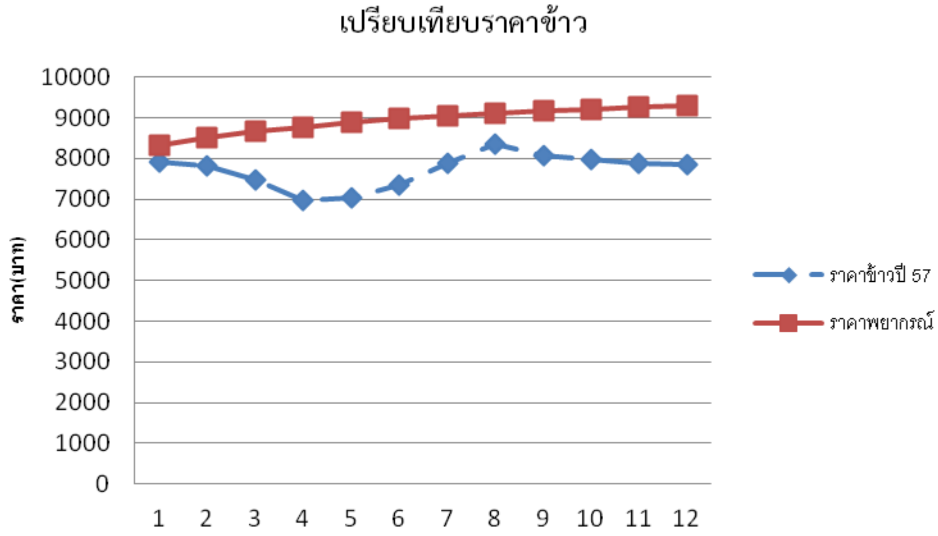


ภาพที่ 4-25 การเปรียบเทียบราคาข้าวเปลือกเจ้า 15% ของไทยปี 2552-2556 กับค่าพยากรณ์ราคาด้วยตัวแบบ ARIMA

ตารางที่ 4-14 ผลการเปรียบเทียบราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ. 2557 กับค่าพยากรณ์ราคาด้วยตัวแบบ ARIMA (1, 0, 0)

เดือนที่	ราคาจริง	ราคาพยากรณ์	เดือนที่	ราคาจริง	ราคาพยากรณ์
1	7914	8335.964	7	7870	9044.621
2	7827	8507.301	8	8339	9108.494
3	7482	8652.656	9	8055	9162.681
4	6982	8775.968	10	7980	9208.651
5	7034	8880.581	11	7878	9247.65
6	7337	8969.33	12	7862	9280.735

กราฟแสดงการเปรียบเทียบราคาข้าวเปลือกเจ้า 15% ของไทยปี พ.ศ. 2557 กับค่าพยากรณ์ราคาด้วยตัวแบบ ARIMA (1, 0, 0)



ภาพที่ 4-26 การเปรียบเทียบราคาข้าวปี 2557 กับค่าพยากรณ์ราคาโดยตัวแบบ ARIMA

การพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยในปี พ.ศ. 2557 ด้วยตัวแบบ ARIMA โดยใช้วิธีประมาณค่าพารามิเตอร์ด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ใช้การแจกแจงแบบโคชี ตัวแบบ ARIMA ประกอบด้วยตัวแบบ ARIMA (p, d, q) และตัวแปรราคาข้าว (a_t) ตัวแบบที่ได้ คือ ARIMA (1, 0, 0) สามารถเขียนเป็นสมการได้ ดังนี้

$$a_t = y_t - \frac{3}{4nh^3}\sum(h^2-(x-x_i)^2)I + \rho\varepsilon_{i-1}$$

$$\hat{a}_t = y_t - (0.0002151) + (-0.252)(1.89478x10^{-14})$$

$$\hat{a}_t = y_t - 0.0002151$$

a_t คือ ราคาข้าวที่ปรับ

y_t คือ ราคาข้าวปีปัจจุบัน

ε_{i-1} คือ ความคลาดเคลื่อนจากการพยากรณ์ ณ เวลาที่ $t-1$

แสดงผลการวิเคราะห์ได้ดังตารางที่ 4-15

ตารางที่ 4-15 ตัวแปร Estimate ค่าคลาดเคลื่อนมาตรฐาน t และ p-value

ตัวแปร		ค่า Estimate	ค่าคลาดเคลื่อนมาตรฐาน	t	p-value
a_t	Lag 1	.848	.069	12.293	.000

จากตารางที่ 4-15 ผลการตรวจสอบความเหมาะสมของตัวแบบ ARIMA (1, 0, 0) ชี้ให้เห็นว่า ค่าพารามิเตอร์มีค่าประมาณ .848 ค่าคลาดเคลื่อนมาตรฐาน .069 ค่า t เท่ากับ 12.293 และค่า p-value เท่ากับ .000 สามารถสรุปได้ว่า ตัวแบบ ARIMA (1, 0, 0) มีความเหมาะสมในการพยากรณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงดังตารางที่ 4-16

ตารางที่ 4-16 Model Fit Statistics และ Ljung-box

Model Fit Statistics		Ljung-Box Q		
R-squared	RMSE	Statistics	df	p-value
.603	764.718	17.831	17	.40

จากดังตารางที่ 4-16 การตรวจสอบความเหมาะสมของตัวแบบ ARIMA (1, 0, 0) โดยใช้ Model Fit Statistics และ Ljung-box ของตัวแบบพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยในปี พ.ศ. 2557 โดยใช้ฟังก์ชันรีโปรดิวซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ สมการ คือ $\hat{a}_t = y_t - \frac{3}{4nh^3}\sum(h^2-(x-x_i)^2)I + \rho\varepsilon_{i-1}$ ได้ $\hat{a}_t = y_t - 0.0002151$ แต่สมการนี้เป็นการปรับข้อมูลให้เรียบแล้ว จึงมีการปรับสมการพยากรณ์เป็น ดังนี้

$$y_t = \theta_0 - \theta_0\phi_1 + \phi_1 y_{t-1} + \varepsilon_t$$
$$\hat{y}_t = 1438.8054 + 0.848 y_{t-1} + \varepsilon_t$$

ผลการตรวจสอบความเหมาะสมของตัวแบบ ARIMA (1, 0, 0) ชี้ให้เห็นว่าสัมประสิทธิ์การกำหนดสามารถอธิบายความแปรผันได้ 60.30 % และเมื่อใช้สมการนี้ในการพยากรณ์ราคาข้าวมีค่า RMSE เท่ากับ 764.718 บาทหรือโดยประมาณ 765 บาทต่อตัน และจากสถิติ Ljung-Box ที่มีค่า p-value เท่ากับ .40 สามารถสรุปได้ว่า ตัวแบบ ARIMA (1, 0, 0) มีความเหมาะสมในการพยากรณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เป็นไปตามสมมติฐานข้อ 3

บทที่ 5
สรุปและอภิปรายผล

การวิจัยนี้เป็นการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ กรณีความคลาดเคลื่อนมีอัตตสหสัมพันธ์ เป็นการปรับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ตามสูตรของ Ferraty (2007) โดยใช้หลักการของฟังก์ชันเคอร์เนล การเลือกใช้แบนวิดจ์ และการแจกแจงของตัวแปรสุ่ม ร่วมกับตัวแบบ ARIMA เพื่อให้ได้ตัวแบบที่เหมาะสม โดยมีวัตถุประสงค์เพื่อ 1) พัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ สำหรับการพยากรณ์ราคาข้าวเปลือกเจ้าของไทยล่วงหน้า 2) เพื่อเปรียบเทียบประสิทธิภาพของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ (RKHS) แบบปรับใหม่กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ (2x6x9) ได้แก่ การใช้แบนวิดจ์ 2 วิธีคือ วิธีที่ 1 Rules of Thumb และวิธีที่ 2 Silverman’s Rules of Thumb ขนาดตัวอย่าง 6 ขนาด ($n=$5, 10, 15, 30, 50, 100) และค่าสัมประสิทธิ์สหสัมพันธ์ 9 ค่า ($\rho=$0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) และโดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) 3) เพื่อพยากรณ์ราคาของข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA โดยสามารถสรุปผลการพัฒนาได้ ดังนี้

สรุปผลการวิจัย

1. ผลการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่

การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ กรณีความคลาดเคลื่อนมีอัตตสหสัมพันธ์กัน ที่นำเสนอในงานวิจัยนี้เป็นการประมาณค่าพารามิเตอร์การถดถอยแบบปรับใหม่ที่ประยุกต์และพัฒนามาจากสูตรของ Ferraty (2007) และงานวิจัยของ Heng Lian (2007) ทำให้ได้วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ได้ดังสมการ

$$\hat{y}=\frac{3}{4nh^3}\sum(h^2-(x-x_i)^2)I+\rho\varepsilon_{i-1}+a_i$$

2. ผลการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)

ผลการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ กับวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) โดยวิธีการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล (Monte Carlo Technique) ภายใต้สถานการณ์ 108 สถานการณ์ ดังนี้

2.1 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) แบนวิดจ์แบบ h_{ROT} ใน 54 สถานการณ์ ปรากฏว่า มี 52 สถานการณ์ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ คือในสถานการณ์ที่ n =10, 15, 30, 50 และ 100 ในทุกค่าของ ρ ยกเว้นกรณีที่ $n=5$ ที่ $\rho=0.1$, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) แต่ที่กรณี $\rho=0.2$, 0.3 ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) สูงกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007)

2.2 ผลการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) แบนวิดจ์แบบ h_{SROT} ใน 54 สถานการณ์ปรากฏว่ามี 51 สถานการณ์ที่วิธีรีโปรดิวส์ซิงเคอร์เนล ฮิลเบอร์ทสเปซของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี คือ ในสถานการณ์ที่ n =10, 15, 30, 50 และ 100 ในทุกค่าของ ρ ยกเว้นกรณีที่ $n=5$ ที่ $\rho=0.2$, 0.3, 0.6

ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) สูงกว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่

3. ผลการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าด้วยตัวแบบ ARIMA

ผลการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าโดยวิธีการประมาณค่าพารามิเตอร์การถดถอยด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ สามารถพยากรณ์ได้ด้วยตัวแบบ ARIMA (p, d, q) โดยใช้วิธีประมาณค่าพารามิเตอร์ด้วยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบิร์ทสเปซ แบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ แบนวิดจ์แบบ h_{ROT} การแจกแจงแบบโคชี ด้วยตัวแบบ ARIMA ที่ใช้ในการพยากรณ์ คือ ARIMA (1, 0, 0) สามารถเขียนเป็นสมการได้ ดังนี้

$$\hat{y}_t = 1438.8054 + 0.848 y_{t-1} + \varepsilon_t$$

ผลการตรวจสอบความเหมาะสมของตัวแบบ ARIMA (1, 0, 0) ชี้ให้เห็นว่าสัมประสิทธิ์การกำหนดสามารถอธิบายความแปรผันได้ 60.30 % และเมื่อใช้สมการนี้ในการพยากรณ์ราคาข้าวเปลือกเจ้ามีค่า RMSE เท่ากับ 764.718 บาท หรือโดยประมาณ 765 บาท และจากสถิติ Ljung-Box ที่มีค่า p-value เท่ากับ 0.40 สามารถสรุปได้ว่า ตัวแบบ ARIMA (1, 0, 0) มีความเหมาะสมในการพยากรณ์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อภิปรายผล

1. การประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov) ตามการใช้แบนวิดจ์ 2 วิธี คือ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ และ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ที่ $n = 5$, 10, 15, 30, 50, 100 และ ($\rho = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9) ปรากฏว่า การใช้แบนวิดจ์แบบ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ มีค่าของ MSE และค่า $Std.\ error$ ต่ำกว่าการใช้แบนวิดจ์แบบ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$ ทั้งการแจกแจงแบบปกติ (Normal) และการแจกแจงแบบโคชี (Cauchy)

2. การประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ โดยฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี และวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) โดยฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ ใช้แบนวิดจ์แบบ h_{ROT} เมื่อ $n=5$, 10, 15, 30, 50 และ 100 ปรากฏว่า วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ต่ำกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติในทุกค่าของ ρ ยกเว้นกรณีที่ $n=5$ และค่า $\rho=0.6$ ที่วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) สูงกว่าวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ

การประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ (Epanechnikov) ซึ่งเป็นฟังก์ชันเคอร์เนลที่ประสิทธิ์ภาพสัมพัทธ์สูงที่สุด (1.00) ที่ใช้แบนวิดจ์แบบ $h_{ROT}=1.06\hat{\sigma}(n^{\frac{1}{5}})$ การแจกแจงแบบโคชีจะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ลดลง ซึ่งสอดคล้องกับงานวิจัยของ Mugdadi และ Ahman (2004) โดยเมื่อมีขนาดกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) จะลดลงในทุกกรณี Preda (2005) ได้ศึกษาถึงรูปแบบและวิธีการของการถดถอยสำหรับฟังก์ชันข้อมูลการทำงานโดยการทำซ้ำ วิธีการใช้ช่องว่าง และได้พัฒนาการวิเคราะห์ความถอดถอยสำหรับฟังก์ชันของการสุ่มตัวแปร $x=\{x_t\}_{t\in T}$ จากวิธีการของ Kernel Hilbert Spaces ที่สามารถพัฒนานำมาปรับใช้ได้กับวิธีการของการวิเคราะห์การถดถอยได้อย่างดี Oscar Gonzalez-Recio, at el. (2008) ได้ศึกษาวิธีมาตรฐานของการประเมินพันธุกรรม (E-BLUP) 4 วิธี คือ F-metric Model, Kernel Regression, RKHS Regression, และ Bayesian Regression ของพ่อพันธุ์ (ไก่) ซึ่งใช้อัตราการตายของไก่เป็นตัวแปรตอบสนอง ด้วยกรอบแนวคิดของเบส์ (Bayesian framework) ข้อมูลสถิติอัตราการตายของไก่อายุมาก (14–42 วัน) 12, 167 ตัวจากพ่อพันธุ์ 200 ตัว ได้รับการตรวจแก้ล่วงหน้าสำหรับผลที่คงที่และผลเชิงสุ่ม ที่ถูกใช้ในตัวแบบสำหรับ

การประเมินพันธุกรรมและผลด้านการจับคู่ผสมพันธุ์ และสำหรับพ่อพันธุ์แต่ละตัวก็มีการคำนวณค่าเฉลี่ยของสถิติที่ได้รับการแก้ไขด้วย ในวิธีการ 3 วิธีที่ใช้ในการประเมินพันธุกรรม ปรากฏว่ามีข้อมูล SNP ที่เกี่ยวข้องกับอัตราการตาย 24 ชุด และมี 1,000 ชุดในวิธี Bayesian regression โดยทำให้มีเครื่องหมายที่ Genome โดยรวม ในวิธี E-BLUP นั้น ค่าเฉลี่ยตอนท้าย (Posterior Mean) ของการสืบทอดสายพันธุ์ (Heritability) ของอัตราการตาย เท่ากับ 0.02 ซึ่งชี้ว่าสามารถที่จะทำให้การประเมินพันธุกรรมดีขึ้นได้ ซึ่ง RKHS เป็นวิธีที่ทำให้มีความแปรปรวนมากขึ้น มีค่าผลรวมกำลังสอง (Sum of Squares) ที่เหลือต่ำ ส่วนความสามารถในการทำนาย ซึ่งวัดด้วยวิธี Cross Validation ชี้ให้เห็นถึงข้อดีของวิธี RKHS ซึ่งทำให้ความถูกต้องแม่นยำเพิ่มขึ้นมากจาก 25 % เป็น 150% และ Gianola and Van Kaam (2008) ใช้วิธี RKHS สำหรับทำนายค่าเชิงพันธุกรรมโดยรวมสำหรับลักษณะเชิงปริมาณ และใช้ประโยชน์จากข้อมูล Phenotype และ Genome ไปพร้อมกัน โดยมีข้อเสนอถึงความจำเป็นต้องจัดกระทำกับข้อมูลแบบนอนพาราเมตริก เพื่อจับการกระทำร่วมกันที่ซับซ้อนหลายกิจกรรมที่เกิดขึ้นอย่างมีศักยภาพในตัวแบบ Genome โดยรวม ที่เกี่ยวข้องกับตัวทำเครื่องหมาย (Markers) หรือชุดข้อมูลภาวะหลายรูปแบบหรือ SNP (Single-Nucleotide Polymorphism) หลังจากการตรวจทานด้วยวิธี RKHS แล้ว พบว่า ข้อกำหนดเฉพาะทางสถิติยอมรับการนำเสนอตัวแบบเชิงเส้นมาตรฐานแบบผลผสมด้วยค่าพารามิเตอร์ ที่ราบรื่นที่ถูกจัดกระทำในฐานะที่เป็นองค์ประกอบของความแปรปรวน และ Esperanze Ayuga Te’llz at el. (2008) ได้ศึกษาในหลาย ๆ ด้านของทางวิศวกรรมที่ศึกษาในเรื่องของระบบไบโอซิสเต็ม ในเรื่องของการเลือกใช้สถิติในการวิเคราะห์ข้อมูล และทดสอบสมมติฐานขั้นพื้นฐานที่จำเป็นสำหรับวิธีการพยากรณ์แบบปกติ (Conventional Forecasting Method) ในแต่ละสถานการณ์ที่ต้องหาวิธีและทางเลือกในเรื่องต่าง ๆ ที่ต้องนำสถิติมาใช้ในการพิจารณาวิเคราะห์ถึงข้อตกลงเบื้องต้นต่าง ๆ ที่ข้อมูลในทางวิศวกรรมไม่เป็นไปตามข้อตกลงเบื้องต้น การใช้ฟังก์ชันการประมาณค่าแบบนอนพาราเมตริก (Nonparametric Function Estimation) ประมาณค่าเป็นฟังก์ชันวิเคราะห์ข้อมูลในช่วงแคบ ๆ ซึ่งใช้ข้อมูลจากจุดเล็กทำให้เกิดจุดอ่อนของมูลในเรื่องของข้อตกลงเบื้องต้น เช่น ความต่อเนื่องและความแตกต่างของฟังก์ชันเป้าหมาย คือ การใช้ที่มากกว่านิยามของข้อตกลงเบื้องต้น ของรูปแบบของฟังก์ชันเป้าหมายทั้งหมด ซึ่งผลการวิจัยพบว่า การนำกฎข้อตกลงพื้นฐานมาใช้ในการตัดสินใจแสดงให้เห็นว่าการประยุกต์ฟังก์ชันการประมาณค่าแบบนอนพาราเมตริก (Nonparametric Estimate Method)

ภายใต้กฎข้อตกลงเบื้องต้นต่าง ๆ ของสถิติในขั้นตอนแรกสามารถเชื่อมต่อกันได้อย่างดีของการประยุกต์สถิติในการประมาณค่าแบบเคอร์เนล (Kernel Estimation)

3. การพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยด้วยตัวแบบ ARIMA ผลการตรวจสอบความเหมาะสมของตัวแบบ ARIMA (1, 0, 0) ชี้ให้เห็นว่า สัมประสิทธิ์การกำหนดสามารถอธิบายความแปรผันได้ 60.30 % และเมื่อใช้สมการนี้ในการพยากรณ์ราคาข้าวมีค่า RMSE เท่ากับ 764.718 บาท หรือโดยประมาณ 765 บาท และจากสถิติ Ljung-Box ที่มีค่า p-value เท่ากับ .40 สามารถสรุปได้ว่าตัวแบบ ARIMA (1, 0, 0) มีความเหมาะสมในการพยากรณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับ จิตราภรณ์ ฟั่นศิริ (2547) ศึกษาการพยากรณ์ราคาส่งออกข้าวด้วยวิธี ARIMA โดยมีวัตถุประสงค์เพื่อศึกษารูปแบบและพยากรณ์ราคาส่งออกข้าวของไทย ในการวิเคราะห์จะใช้ข้อมูลราคาส่งออกข้าวเป็นรายเดือนในช่วงเดือนมกราคม พ.ศ. 2531 ถึง เดือนธันวาคม พ.ศ. 2546 จำนวน 192 เดือน ข้อมูลจากกรมการค้าต่างประเทศวิธีการศึกษาจะทดสอบความนิ่งของข้อมูลโดยใช้วิธีการทดสอบ Unit Root และกำหนดตัวแบบด้วยวิธีของบอกซ์-เจนกินส์ ผลการทดสอบพบว่าข้อมูลราคาส่งออกข้าวมีลักษณะไม่นิ่งจึงทำการหาผลต่างอันดับ 1 และจากการพิจารณาค่าสหสัมพันธ์ในตัวเองจะได้ตัวแบบที่เหมาะสมขึ้นอยู่กับค่า AR(1) และ AR(19) โดยมีค่าสัมประสิทธิ์เท่ากับ 0.360 และ 0.228 ตามลำดับ และมีนัยสำคัญทางสถิติที่ระดับ .01 สำหรับการตรวจสอบความถูกต้องพบว่าค่าประมาณความคลาดเคลื่อนมีลักษณะเป็นเชิงสุ่ม ที่ระดับนัยสำคัญทางสถิติ .01 จากค่ารากที่สองของค่าเฉลี่ยค่าความคลาดเคลื่อนกำลังสองและค่าสัมประสิทธิ์ Thiel ที่มีค่าต่ำสุด จะได้ว่าตัวที่แบบมีความเหมาะสมที่สุด และพรหมภรณ์ แสงภัทรเนตร (2548) ได้ศึกษาเรื่องการพยากรณ์ราคาข้าวภายในประเทศซึ่งได้วิเคราะห์การเคลื่อนไหวของพื้นที่ปลูกผลผลิตและราคาข้าว ได้แก่ข้าวเปลือกเจ้านาปี 5%, 10%, 15%, 25% และหอมมะลิ 100% โดยใช้ข้อมูลราคาข้าวแต่ละชนิด เป็นรายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2527 จนถึงเดือนกุมภาพันธ์พ.ศ. 2548 และใช้เทคนิคการพยากรณ์เชิงปริมาณ ได้แก่วิธี Winters, วิธีBox–Jenkins, วิธีแยกส่วนประกอบอนุกรมเวลา และวิธีวิเคราะห์การถดถอย และใช้ค่า MAPE, MAD และMSD ในการเปรียบเทียบวิธีการพยากรณ์ ซึ่งจากผลการศึกษาชี้ให้เห็นว่า วิธีแยกส่วนประกอบเหมาะกับการพยากรณ์ราคาข้าวเปลือกเจ้านาปี 5%, 15%, 25% และหอมมะลิ 100% ส่วนวิธี Box–Jenkins เหมาะกับการพยากรณ์ราคาข้าวเปลือกเจ้านาปี 10%

ข้อเสนอแนะ

ในการวิจัยในครั้งนี้เป็นการพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ กรณีข้อมูลมีอัตตสหสัมพันธ์ เพื่อใช้ปรับข้อมูลให้เรียบก่อนที่จะนำเข้าสมการพยากรณ์ด้วยตัวแบบ ARIMA (1, 0, 0) ซึ่งผู้วิจัยมีข้อเสนอแนะ ดังนี้

1. ข้อเสนอแนะในการนำไปใช้

1.1 วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ เป็นวิธีการทางนอนพาราเมตริกภายใต้เงื่อนไขข้อมูลมีอัตตสหสัมพันธ์ ซึ่งเป็นการศึกษาที่การเลือกใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ การแจกแจงของตัวแปรสุ่มแบบโคชี และการใช้แบนวิดจ์แบบ Rule of Thumb เพื่อให้ได้วิธีการประมาณค่าพารามิเตอร์การถดถอยที่มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง MSE และค่าความคลาดเคลื่อนมาตรฐาน $Std.\,error$ ที่ต่ำผู้วิจัยควรคำนึงถึง ขนาดตัวอย่างที่มีขนาดตั้งแต่ 10 ตัวอย่างขึ้นไป

1.2 การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยมีปัจจัยด้านอื่นที่เป็นตัวแทรกแซงราคา จึงทำให้ผลการพยากรณ์ได้ค่าสัมประสิทธิ์การกำหนดสามารถอธิบายความแปรผันได้น้อย ควรศึกษาถึงปัจจัยที่ส่งผลกระทบต่อการกำหนดราคาพืชผลทางการเกษตรในปัจจุบัน และอนาคตด้วย

2. ข้อเสนอแนะในการศึกษาต่อไป

2.1 การศึกษาครั้งนี้ผู้วิจัยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ เป็นวิธีการทางนอนพาราเมตริก แล้วนำเข้าสมการพยากรณ์ด้วยตัวแบบ ARIMA โดยใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ การแจกแจงของตัวแปรสุ่มแบบโคชี และแบนวิดจ์แบบ Rule of Thumb ในการปรับข้อมูลให้เรียบ ดังนั้นในการศึกษาวิจัยครั้งต่อไปควรการเลือกใช้ฟังก์ชันเคอร์เนล การแจกแจงของตัวแปรสุ่ม และการเลือกใช้แบนวิดจ์ที่เหมาะสม กับชนิดหรือประเภทของข้อมูล

2.2 การพัฒนาฟังก์ชันเคอร์เนล การแจกแจงของตัวแปรสุ่ม และแบนวิดจ์ที่เหมาะสมในการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ แล้วจึงนำไปเข้าสมการพยากรณ์ จะเป็นประโยชน์กับการพยากรณ์ราคาของผลผลิตทางการเกษตรชนิดอื่น ๆ และโรงงานอุตสาหกรรมด้านการวางแผนการผลิตต่อไป

บรรณานุกรม

กิตติ ภักดีวัฒนะกุล และพนิดา พานิชกุล. (2554). *การวิเคราะห์เชิงปริมาณเพื่อการตัดสินใจ.* กรุงเทพฯ: เคทีพี.

กัลยา วานิชย์บัญชา. (2554). *การวิเคราะห์สถิติ: สถิติสำหรับการบริหารและวิจัย* (พิมพ์ครั้งที่ 13) กรุงเทพฯ: ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

จิตราภรณ์ ฟั่นศิริ (2547). *การพยากรณ์ราคาส่งออกข้าวโดยวิธีอารีมา*. วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาวิชาสถิติประยุกต์, บัณฑิตวิทยาลัย, มหาวิทยาลัยเชียงใหม่.

มุกดา แม้นมินทร์. (2549). *อนุกรมเวลาและการพยากรณ์.* กรุงเทพฯ: ประกายพรึก.

มนตรี พิริยะกุล. (2545). *เทคนิคการวิเคราะห์สมการการถดถอย.* กรุงเทพ: สำนักพิมพ์มหาวิทยาลัยรามคำแหง.

ปิยะฉัตร ลีลาศิลปศาสตร์. (2549). *การเลือกแบนวิดจ์สำหรับการประมาณความหนาแน่นเคอร์เนลของฟังก์ชันตัวแปรสุ่ม.* วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, สาขาวิชาสถิติประยุกต์, บัณฑิตวิทยาลัย, สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ.

ทรงศิริ แต้สมบัติ. (2549). *การพยากรณ์เชิงปริมาณ.* กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.

ศิริรัตน์ วงศ์ปกรณ์กุล. (2539). *การสร้างตัวเลขสุ่มเทียม.* กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

Ahmad, I.A., Mugdadi, A. R. (2004) Analysis of kernel density estimation of function of random variable. *Journal of Nonparametric Statistics*, *15(2003b),* 579-605.

Altman, N., Leger, C. (1995). Bandwidth selection for kernel distribution function estimation. *Journal of Statistical Planning and Inference*, 195-214.

Andrews et al. (2013) *Detecting an external influence on recent changes in oceanic oxygen using an optimal fingerprinting method*. *Biogeosciences, 10.*

Aronszajn, N. (1950). *Theory of reproducing kernels*. Trans. *Am. Math. Soc. 68*, 337-404.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. (2nd ed.). Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York: Columbia University.

Bergman, S. (1950). *The Kernel Function and Conformal Mapping*. Am. Math. Soc., New York. American Mathematical Society.

Bouboulis. (2011). Reproducing kernel Hilbert spaces and fractal interpolation. *Journal of Computational and Applied Mathematics,* 235(12).

Bowman, A. W. (1984). An alternative method of cross-validation for smoothing of density estimates*. Biometrika, 71*, 353-360.

Bower man, & O'Connell. (1993). A comparative study of linear and nonlinear Models for aggregate retail sales forecasting*. International Journal of Production Economics, 86*, 220–235.

Bower man & O'Connell. (1993). *Forecasting and Time Series*. An Applied Approach (3rd ed.). Duxbury Press, Belmont, CA.

Box, G. E. P., & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco: Wiley.

Breiman, Merisel & Purcell. (1977). On the bias of variable bandwidth curve estimators. *Oxford Journals Life Sciences & Mathematics biometrika*, *77*, 529-535.

Dong Zaheng Lu, at el, (2008). *A reproducing kernel hilbert space framework for pairwise time series distances*. International Conference on Machine *Learning.* 624-631.

Deheuvels. (1977). Nonparametric regression analysis of longitudinal data. springer verlay. *Journal of the American Statistical Association*, 93: 1403-1418.

Daniel Gianola & Johannes Kaam. (2008). *Reproducing Kernel Hilbert Spaces Regression Methods for Genomic Assisted Prediction of Quantitative Traits*. GENETICS *vol. 178 no. 4, 2289-2303.*

Fix, E., & Hodges, J. L. (1951*).* An Important contribution to nonparametric discriminant analysis and density estimation*. International Statistical Review, 57*(3) pp. 233–247.

Faraway, J. J., & Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *Journal of the American Statistical Association*, *85*, 1119-1122.

Ferraty F., & Vieu, P. (2003). Curves discrimination: A nonparametric approach. Comput. Statist. *Data Anal, 44*, 161–173.

Ferraty, F., & Vieu, P. (2004). *Nonparametric models for functional data with application in regression, time series prediction and curve discrimination*. Nonparametric Statist. 16 (1–2), 111–125.

Ferraty, F., & Vieu, P. (2007) *Nonparametric Functional Data Analysis*. Springer-Verlag.

Gauging Peter Zhang. (2003). A comparative Study of linear and nonlinear model for aggregate retail sales forecasting. International *Journal of Production Economic, 3*(11), 217-231.

Geng. (2011). Some error estimates for the reproducing kernel Hilbert Spaces method. *Journal of Computational and Applied Mathematics, 296*: 789-797.

Hansen, B. E. (2004). Bandwidth selection for nonparametric distribution estimation. *manuscript, University of Wisconsin*.

Higgins, J. J. (2004*). Introduction to modern non-parametric statistics*. Brooks/Cole-Thomson Learning, USA: Duxbury Press.

Hammersley, J. M., & Handscomb, D. C. (1965). *Monte Carlo methods*. New York: Wiley.

Holt, Charles C. (1957). Forecasting Trends and Seasonal by Exponentially Weighted Averages*. International Journal of Forecasting*. 20(1), 5–10.

Hristache, M., Juditsky, A., Polzehl, J., & Spokoiny, V. (2001). Structure adaptive approach for dimension reduction. *The Annals of Statistics, 29*(6), 1537-1566.

Janson, S. (1997). *Gaussian Hilbert Spaces*. Cambridge: Cambridge University Press.

Jones, M. C., Marron, J. S., & Sheather, S. J. (1996). A brief survey of bandwidth selection. *Journal of the American Statistical Association, 91*, 401–407.

Ker-Chau, Li. (1991). Sliced inverse regression for dimension reduction. (with discussion). *Journal of the American Statistical Association*, *86*, 316-342.

Lian, H. (2007). Nonlinear functional models for functional responses in reproducing kernel hilbert spaces. *Canadian Journal of Statistics*, *35*(4), 597–606.

Loftsgaarden & Queensberry. (1965). *A nonparametric estimate of a multivariate density function*. Ann Math Statist, 36(3): 1049-1051.

Marron, J. S., & Wand, M. P. (1992). Exact mean integrated squared error. *The Annals of Statistics, 20*, 712-736.

Mugdadi, A. R,. & Ahmad, I. A. (2004). A bandwidth selection for kernel density Estimation of functions of random variables*. Computation Statistics and Data Analysis, 47*, 49-62.

Murray Rosenblatt. (1956). Remarks on some Nonparametric Estimates of a Density Function. *Ann Math Statistics, 27*, 832-837.

Neter et al. (2004). *Applied Linear Statistical Models*. Springer-Verlag. Chicago: Irwin.

Parzen Emanuel. (1962). On Estimation a Probability Density Function and Mode. *Annals of Mathematical Statistics, 33*, 1065-1076.

Piera-Martinez, M., Vazquez, E., Walter, E., & Fleury, G. (2007). RKHS classification for multivariate extreme-value analysis. *In IASC 07-Statistics for Data Mining, Learning and Knowledge Extraction*. p. NA.

Preda, C. (2007*).* Regression models for functional data by reproducing kernel Hilbert spaces methods. *Journal of Statistical Planning and Inference, 137*, 829-840.

Ramsay, J. O,. & Silverman, B.W. (2005). *Functional Data Analysis.* 2nd ed. New York: Springer.

Roman. Rosipal**,.** & Leonard, J. Trejo. (2001). Regression in reproducing kernel hilbert space*. Kernel Machines, Section, 2*(12): 97-123.

Rosipal, R., & Trejo, L. J. (2001). Kernel partial least squares regression in reproducing kernel hilbert space. *Journal of machine learning research*, *2,* 97-123.

Robert, C., & Casella, G. (2005). Monte Carlo Statistical Methods. Journal of Applied Statistics, 32(6), 681-682.

Rosenblatt. (1956). Remarks on Some Nonparametric Estimates of a Density Function. *Ann Math. Statist*, *27*, *3*, 832-837.

Silvia Biancocin. (2008). The Henderson Smoother in reproducing kernel Hilbert. *Journal or Business & Economic Statistics, 26,* 536-545.

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society-B, 43*, 97–99.

Silverman, B. W. (1982). Kernel density estimation using the fast Fourier transform. *Applied Statistics, 31*, 93–97.

Silverman, B. W. (1986). *Density Estimation foe Statistics and Data Analysis*. New York: Chapman and Hall London.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel Density estimates*. Annals of Statistics, 12,* 1285–1297.

Taouali, O., Elaissi, E., & Messaoud, H. (2012). Design and comparative study of online kernel methods identification of nonlinear system in RKHS space. Artificial Intelligence Review, 37(4), 289-300.

Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika, 76*, 705-712.

Turlach, B. A. *Bandwidth Selection in Kernel Density Estimation: A Review*. www.citeceer.com.

Vogt, W. P., & Johnson, R. B. (2011). Dictionary of Statistics & Methodology: A Nontechnical Guide for the Social Sciences: A Nontechnical Guide for the Social Sciences. Sage.

Wahba, G. (1990). *Spline Models for Observational Data.* Series in Applied Mathematics, Vol.59, SIAM.

Winters, P. R. (1960). *Forecasting Sales by Exponentially Weighted Moving Averages. Management Science*, 6(3), 324–342.

Ying & Wei, L. J. (1994). The Kaplan Meier estimate for dependent failure time observations, *J. Multivariate Anal, 50*, 17-29.

Wolfgang Hardle. (1990). *Applied Nonparametric Regression.* Cambridge University Press-Business & Economics-333.

Wolfgang Hardle. (1990). *Smoothing Techniques with Implementation in S*. New York: Spinger-Verleg.

Zhang et al. (2001). A simulation study of artificial neural networks for nonlinear time series forecasting. *Computers & Operations Research, 28*, 381–396.

Zhang, S. (2013). Inverse Gaussian process-based corrosion growth model for energy pipelines considering the sizing error in inspection data. *Elsevier Corrosion Science, 73*(2), 309-320.

ภาคผนวก

ภาคผนวก ก

Code โปรแกรม R ในการจำลองสถานการณ์ด้วยวิธีมอนติคาร์โล ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน (Gaussian) การแจกแจงแบบปกติ

Code โปรแกรม R การแจกแจงปกติ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 5$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
```

```
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t--------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t--------------------------------------------------",'\n')
}
ypredict(5,0.1)
ypredict(5,0.2)
```

```
ypredict(5,0.3)
ypredict(5,0.4)
ypredict(5,0.5)
ypredict(5,0.6)
ypredict(5,0.7)
ypredict(5,0.8)
ypredict(5,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 10$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
```

```
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
```

```
std.err.y <- sqrt(mse.y)

cat("\t-----------------------------------------------------",'\n')

cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )

cat("\t-----------------------------------------------------",'\n')

}

ypredict(10,0.1)

ypredict(10,0.2)

ypredict(10,0.3)

ypredict(10,0.4)

ypredict(10,0.5)

ypredict(10,0.6)

ypredict(10,0.7)

ypredict(10,0.8)

ypredict(10,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 15$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)

{

M <- 1000
```

```
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)

A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
```

```
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(15,0.1)
ypredict(15,0.2)
ypredict(15,0.3)
ypredict(15,0.4)
ypredict(15,0.5)
ypredict(15,0.6)
```

```
ypredict(15,0.7)
ypredict(15,0.8)
ypredict(15,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 30$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
```

```
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
```

```
cat("\t-----------------------------------------------------",'\n')
}
ypredict(30,0.1)
ypredict(30,0.2)
ypredict(30,0.3)
ypredict(30,0.4)
ypredict(30,0.5)
ypredict(30,0.6)
ypredict(30,0.7)
ypredict(30,0.8)
ypredict(30,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 50$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
```

```
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
```

```
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t------------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE =",mse.y, '\n' )
cat("\t------------------------------------------------------",'\n')
}
ypredict(50,0.1)
ypredict(50,0.2)
ypredict(50,0.3)
ypredict(50,0.4)
ypredict(50,0.5)
ypredict(50,0.6)
ypredict(50,0.7)
ypredict(50,0.8)
ypredict(50,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 100$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
```

```
}

for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(100,0.1)
```

```
ypredict(100,0.2)

ypredict(100,0.3)

ypredict(100,0.4)

ypredict(100,0.5)

ypredict(100,0.6)

ypredict(100,0.7)

ypredict(100,0.8)

ypredict(100,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$ $h_2 = 1.06\,(\mathrm{n}^{\frac{1}{5}})$, $n = 5$, $\rho = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)

{

M <- 1000

y <- c(0,n)

x <- rep(0,n)

expected.y <- rep(0,M)

bias <- rep(0,M)

for(i in 1: M)

{

#create normal random variable
```

```
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
```

```
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(5,0.1)
ypredict(5,0.2)
ypredict(5,0.3)
ypredict(5,0.4)
ypredict(5,0.5)
ypredict(5,0.6)
ypredict(5,0.7)
ypredict(5,0.8)
ypredict(5,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$, $n = 10$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
```

```
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
```

```
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE =",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(10,0.1)
ypredict(10,0.2)
ypredict(10,0.3)
ypredict(10,0.4)
ypredict(10,0.5)
ypredict(10,0.6)
```

```
ypredict(10,0.7)
ypredict(10,0.8)
ypredict(10,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$, $n = 15$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
```

```
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-------------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
```

```
cat("\t-----------------------------------------------------","\n')
}
ypredict(15,0.1)
ypredict(15,0.2)
ypredict(15,0.3)
ypredict(15,0.4)
ypredict(15,0.5)
ypredict(15,0.6)
ypredict(15,0.7)
ypredict(15,0.8)
ypredict(15,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$, $n = 30$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
```

```
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
}


for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
```

```
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(30,0.1)
ypredict(30,0.2)
ypredict(30,0.3)
ypredict(30,0.4)
ypredict(30,0.5)
ypredict(30,0.6)
ypredict(30,0.7)
ypredict(30,0.8)
ypredict(30,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$, $n = 50$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)
bias.id <- mean(y) - y
```

```
}
for (j in 2:n){
y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-------------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-------------------------------------------------------",'\n')
}
ypredict(50,0.1)
ypredict(50,0.2)
```

```
ypredict(50,0.3)
ypredict(50,0.4)
ypredict(50,0.5)
ypredict(50,0.6)
ypredict(50,0.7)
ypredict(50,0.8)
ypredict(50,0.9)
```

....

Code โปรแกรม R การแจกแจงปกติ $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$, $n = 100$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rnorm(n,0,1)
```

```
xbar <- mean(x)

xsd <- sd(x)

A <- min(xsd,(IQR(x)/1.34))

h <- 1.06*xsd*n^(1/5)

sum1 <- 0

for (j in 1:1){

y[j] <- (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*xbar)

bias.id <- mean(y) - y

}

for (j in 2:n){

y[j] <-  (-1)/(2*n*(h^3)*sqrt(2*pi))*(sum1 + exp(xbar^2 - 2*xbar*x[j] + x[j]^2) + r*y[j-1])

bias.id <- mean(y) - y

}

expected.y[i] <- mean(y) #individual real y

bias[i] <- mean(bias.id)

#End loop M

}

y.n.real <- mean(expected.y)

aver.bias <- mean(bias)

var.y <- var(expected.y)

mse.y <- var.y + (aver.bias)^2
```

```
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(100,0.1)
ypredict(100,0.2)
ypredict(100,0.3)
ypredict(100,0.4)
ypredict(100,0.5)
ypredict(100,0.6)
ypredict(100,0.7)
ypredict(100,0.8)
ypredict(100,0.9)
```

....

ภาคผนวก ข

Code โปรแกรม R ในการจำลองสถานการณ์ด้วยวิธีมอนติ คาร์โล ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานีชนิคอฟ (Epanechnikov) การแจกแจงแบบโคชี

Code โปรแกรม R การแจกแจงโคชีมาตรฐาน $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 5$, $\rho = 0.1$, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
```

```
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-------------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-------------------------------------------------------",'\n')
}
ypredict(5,0.1)
ypredict(5,0.2)
ypredict(5,0.3)
```

```
ypredict(5,0.4)
ypredict(5,0.5)
ypredict(5,0.6)
ypredict(5,0.7)
ypredict(5,0.8)
ypredict(5,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 10$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
```

```
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
```

```
cat("\t----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE =",mse.y, '\n' )
cat("\t----------------------------------------------------",'\n')
}
ypredict(10,0.1)
ypredict(10,0.2)
ypredict(10,0.3)
ypredict(10,0.4)
ypredict(10,0.5)
ypredict(10,0.6)
ypredict(10,0.7)
ypredict(10,0.8)
ypredict(10,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 15$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
```

```
expected.y <- rep(0,M)

bias <- rep(0,M)

for(i in 1: M)

{

#create normal random variable

x <- rcauchy(n,0,1)

xbar <- mean(x)

xsd <- sd(x)

A <- min(xsd,(IQR(x)/1.34))

h <- 0.9*A*n^(1/5)

sum1 <- 0

for (j in 1:1){

y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)

bias.id <- mean(y) - y

}

for (j in 2:n){

y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )

bias.id <- mean(y) - y

}

expected.y[i] <- mean(y) #individual real y

bias[i] <- mean(bias.id)
```

```
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t----------------------------------------------------",'\n')
}
ypredict(15,0.1)
ypredict(15,0.2)
ypredict(15,0.3)
ypredict(15,0.4)
ypredict(15,0.5)
ypredict(15,0.6)
ypredict(15,0.7)
ypredict(15,0.8)
ypredict(15,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 30$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
```

```
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(30,0.1)
ypredict(30,0.2)
ypredict(30,0.3)
```

```
ypredict(30,0.4)
ypredict(30,0.5)
ypredict(30,0.6)
ypredict(30,0.7)
ypredict(30,0.8)
ypredict(30,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 50$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
```

```
bias.id <- mean(y) - y

}

expected.y[i] <- mean(y) #individual real y

bias[i] <- mean(bias.id)

#End loop M

}

y.n.real <- mean(expected.y)

aver.bias <- mean(bias)

var.y <- var(expected.y)

mse.y <- var.y + (aver.bias)^2

std.err.y <- sqrt(mse.y)

cat("\t----------------------------------------------------",'\n')

cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )

cat("\t----------------------------------------------------",'\n')

}

ypredict(50,0.1)

ypredict(50,0.2)

ypredict(50,0.3)

ypredict(50,0.4)

ypredict(50,0.5)

ypredict(50,0.6)
```

```
ypredict(50,0.7)
ypredict(50,0.8)
ypredict(50,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{SROT} = 0.9A(n^{\frac{1}{5}})$, $n = 100$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 0.9*A*n^(1/5)
```

```
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t------------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t------------------------------------------------------",'\n')
```

```
}
ypredict(100,0.1)
ypredict(100,0.2)
ypredict(100,0.3)
ypredict(100,0.4)
ypredict(100,0.5)
ypredict(100,0.6)
ypredict(100,0.7)
ypredict(100,0.8)
ypredict(100,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{ROT} = 1.06\hat{\sigma}(\mathrm{n}^{\frac{1}{5}})$, $n = 5$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
```

```
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
```

```
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE =",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(5,0.1)
ypredict(5,0.2)
ypredict(5,0.3)
ypredict(5,0.4)
ypredict(5,0.5)
ypredict(5,0.6)
ypredict(5,0.7)
ypredict(5,0.8)
ypredict(5,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{ROT} = 1.06\hat{\sigma}(\mathrm{n}^{\frac{1}{5}})$, $n = 10$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
```

```
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
```

```
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(10,0.1)
ypredict(10,0.2)
ypredict(10,0.3)
ypredict(10,0.4)
ypredict(10,0.5)
ypredict(10,0.6)
```

```
ypredict(10,0.7)
ypredict(10,0.8)
ypredict(10,0.9)
```

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{ROT} = 1.06\hat{\sigma}(\mathrm{n}^{\frac{1}{5}})$, $n = 15$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
```

```
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
```

```
}
ypredict(15,0.1)
ypredict(15,0.2)
ypredict(15,0.3)
ypredict(15,0.4)
ypredict(15,0.5)
ypredict(15,0.6)
ypredict(15,0.7)
ypredict(15,0.8)
ypredict(15,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{ROT} = 1.06\hat{\sigma}(\mathrm{n}^{\frac{1}{5}})$, $n = 30$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
```

```
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
```

```
aver.bias <- mean(bias)

var.y <- var(expected.y)

mse.y <- var.y + (aver.bias)^2

std.err.y <- sqrt(mse.y)

cat("\t-----------------------------------------------------",'\n')

cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )

cat("\t-----------------------------------------------------",'\n')

}

ypredict(30,0.1)

ypredict(30,0.2)

ypredict(30,0.3)

ypredict(30,0.4)

ypredict(30,0.5)

ypredict(30,0.6)

ypredict(30,0.7)

ypredict(30,0.8)

ypredict(30,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{ROT} = 1.06\hat{\sigma}(\mathrm{n}^{\frac{1}{5}})$, $n = 50$, $\rho =$ 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9

```
ypredict <- function(n,r)
```

```
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
sum1 <- 0
for (j in 1:1){
y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)
bias.id <- mean(y) - y
}
for (j in 2:n){
y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )
```

```
bias.id <- mean(y) - y
}
expected.y[i] <- mean(y) #individual real y
bias[i] <- mean(bias.id)
#End loop M
}
y.n.real <- mean(expected.y)
aver.bias <- mean(bias)
var.y <- var(expected.y)
mse.y <- var.y + (aver.bias)^2
std.err.y <- sqrt(mse.y)
cat("\t-----------------------------------------------------",'\n')
cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )
cat("\t-----------------------------------------------------",'\n')
}
ypredict(50,0.1)
ypredict(50,0.2)
ypredict(50,0.3)
ypredict(50,0.4)
ypredict(50,0.5)
ypredict(50,0.6)
```

```
ypredict(50,0.7)
ypredict(50,0.8)
ypredict(50,0.9)
```

....

Code โปรแกรม R การแจกแจงโคชี่มาตรฐาน $h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$, $n = 100$, $\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$

```
ypredict <- function(n,r)
{
M <- 1000
y <- c(0,n)
x <- rep(0,n)
expected.y <- rep(0,M)
bias <- rep(0,M)
for(i in 1: M)
{
#create normal random variable
x <- rcauchy(n,0,1)
xbar <- mean(x)
xsd <- sd(x)
A <- min(xsd,(IQR(x)/1.34))
h <- 1.06*xsd*n^(1/5)
```

```
sum1 <- 0

for (j in 1:1){

y[j] <- (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*xbar)

bias.id <- mean(y) - y

}

for (j in 2:n){

y[j] <-  (3)/(4*n*(h^3))*(sum1 + ((h^2)- (xbar - x[j])^2) + r*y[j-1] )

bias.id <- mean(y) - y

}

expected.y[i] <- mean(y) #individual real y

bias[i] <- mean(bias.id)

#End loop M

}

y.n.real <- mean(expected.y)

aver.bias <- mean(bias)

var.y <- var(expected.y)

mse.y <- var.y + (aver.bias)^2

std.err.y <- sqrt(mse.y)

cat("\t-----------------------------------------------------",'\n')

cat("\tFor sample sizes(n) =",n,"Rho =",r,"y =",y.n.real,"Std. Error =",std.err.y, "MSE
=",mse.y, '\n' )

cat("\t-----------------------------------------------------",'\n')
```

```
}
ypredict(100,0.1)
ypredict(100,0.2)
ypredict(100,0.3)
ypredict(100,0.4)
ypredict(100,0.5)
ypredict(100,0.6)
ypredict(100,0.7)
ypredict(100,0.8)
ypredict(100,0.9)
```

....

ภาคผนวก ค

ผลการจำลองสถานการณ์ด้วยวิธีมอนติ คาร์โล ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

(MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ด้วยการประมาณ

ค่าพารามิเตอร์การถดถอยโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่

การแจกแจงแบบโคชี และวิธีรีโปรดิวส์ซิงเคอร์เนล ฮิลเบอร์ทสเปซของ Ferraty (2007)

การแจกแจงแบบปกติ

ตารางที่ ค.1 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) เมื่อ $n=5$ การแจกแจงแบบปกติ

Normal		$h_{SORT}=0.9A(n^{\frac{1}{5}})$		$h_{ROT}=1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.1	69668250	8346.751	0.7539949	0.5685084
	0.2	3.98E+13	6312315	0.1272428	0.3567111
	0.3	8.67E+22	2.94401E+11	0.06947197	0.2635754
	0.4	2.97E+33	1.72E+12	17.87866	4.228316
5	0.5	9.82E+23	9.90748E+11	0.6145895	0.7839576
	0.6	2.60E+24	1.61E+12	0.4982718	0.7058837
	0.7	1.93E+19	4390183018	7.411393	2.722387
	0.8	5.83E+18	2.41E+09	0.3094456	0.5562783
	0.9	1.08E+16	103780714	1.89925	1.378133

ตารางที่ ค.2 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) เมื่อ $n=10$ การแจกแจงแบบปกติ

Normal		$h_{SORT}=0.9A(n^{\frac{1}{5}})$		$h_{ROT}=1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.1	211.8464	1.46E+01	29.1842	5.402241
	0.2	2.70E+03	51.97183	12.76805	3.573241
	0.3	256.7443	16.02324	7.116949	2.667761
	0.4	3.16E+01	5.617159	4.071698	2.017845
10	0.5	7.32E+01	8.554825	25.96916	5.095994
	0.6	4.37E+02	20.91463	138.1992	11.75582

ตารางที่ ค.2 (ต่อ)

Normal		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.error$	MSE	$Std.error$
	0.7	3.78E+05	614.4602	82.48132	9.081923
	0.8	2.78E+03	52.7158	7.594851	2.755876
	0.9	1.48E+03	38.40806	1.644086	1.282219

ตารางที่ ค.3 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) เมื่อ $n = 15$ การแจกแจงแบบปกติ

Normal		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.error$	MSE	$Std.error$
	0.1	151.5478	12.31048	1.566023	1.251408
	0.2	4.82E+01	6.945996	6.762281	2.600439
	0.3	1.96E+03	44.2269	5.918349	2.432766
	0.4	1.27E+03	35.65924	67.64166	8.224455
15	0.5	2.04E+02	14.267	9.736672	3.120364
	0.6	1.03E+01	3.216892	120.4375	10.9744
	0.7	3.33E+02	18.23463	20.63742	4.542843
	0.8	6.28E+03	79.21766	5.384255	2.3204
	0.9	5.98E+01	7.733792	22.45317	4.738477

ตารางที่ ค.4 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) เมื่อ $n = 30$ การแจกแจงแบบปกติ

Normal		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\ error$	MSE	$Std.\ error$
	0.1	17.51259	4.184805	42.85186	6.546133
	0.2	7.37E+01	8.587233	59.48774	7.71283
	0.3	6.37E+02	25.23162	35.73703	5.978046
	0.4	2.64E+02	16.24477	305.08	17.46654
30	0.5	4.64E+02	21.53365	3.165722	1.779248
	0.6	3.01E+03	54.89506	18.92993	4.350854
	0.7	1.55E+02	12.4386	10.72347	3.274672
	0.8	7.93E+02	28.1517	18.21116	4.267454
	0.9	8.06E+01	8.979027	24.21221	4.920591

ตารางที่ ค.5 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) เมื่อ $n = 50$ การแจกแจงแบบปกติ

Normal		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\ error$	MSE	$Std.\ error$
	0.1	150.5159	12.26849	8.066216	2.840108
	0.2	1.85E+02	13.58648	872.6413	29.5405
	0.3	1.64E+02	12.79659	123.8391	11.1283
	0.4	8.26E+01	9.087965	15.87558	3.984417
50	0.5	1.03E+01	3.21155	18.39301	4.288707
	0.6	2.25E+01	4.743227	61.65818	7.852272

ตารางที่ ค.5 (ต่อ)

Normal		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.7	1.29E+01	3.585561	48.66674	6.976155
	0.8	1.52E+01	3.904113	171.6599	13.1019
	0.9	1.56E+02	12.4841	75.58655	8.694053

ตารางที่ ค.6 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) เมื่อ $n = 100$ การแจกแจงแบบปกติ

Normal		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.1	9.59208	3.097108	16.40854	4.050745
	0.2	2.32E+02	15.22515	936.0035	30.59418
	0.3	1.46E+03	38.18212	1.181845	1.087127
	0.4	4.37E+01	6.610801	10.30073	3.209476
100	0.5	4.96E+02	22.27686	7.282934	2.698691
	0.6	4.36E+00	2.087637	14.00966	3.742949
	0.7	1.13E+01	3.357755	21.08249	4.591568
	0.8	7.72E+02	27.78739	42.72618	6.536527
	0.9	1.88E+00	1.372823	81.41796	9.02319

ตารางที่ ค.7 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ เมื่อ $n=5$ การแจกแจงแบบโคชี

Cauchy		$h_{SORT}=0.9A(n^{\frac{1}{5}})$		$h_{ROT}=1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.1	1.59E+12	1262486	0.3947371	0.5913857
	0.2	3.18E+24	1.78E+12	0.6122232	0.7824469
	0.3	5.11E+16	226077393	0.09307431	0.3050808
	0.4	5.68E+13	7.54E+06	0.8607768	0.9277806
5	0.5	1.36E+27	3.69E+13	0.2228366	0.4720557
	0.6	3.50E+17	5.92E+08	0.3367094	0.5802667
	0.7	2.17E+21	46563731448	1.482699	1.217661
	0.8	6.04E+25	7.77E+12	0.1808493	0.4252638
	0.9	7.20E+22	2.6841E+11	0.1644902	0.4055739

ตารางที่ ค.8 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ เมื่อ $n=10$ การแจกแจงแบบโคชี

Cauchy		$h_{SORT}=0.9A(n^{\frac{1}{5}})$		$h_{ROT}=1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.1	3415319	1.85E+03	1.664089	1.289996
	0.2	3.31E+05	574.9508	4.699224	2.167769
	0.3	2724269	1650.536	3.791083	1.94707
	0.4	1.66E+06	1289.882	0.1735095	0.4165447
10	0.5	1.45E+06	1202.444	9.565195	3.092765
	0.6	1.33E+07	3648.124	22.02206	4.692766

ตารางที่ ค.8 (ต่อ)

Cauchy		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.error$	MSE	$Std.error$
	0.7	1.03E+08	10132.9	2.416365	1.554466
	0.8	2.76E+04	166.075	4.975977	2.23069
	0.9	4.19E+06	2047.716	1.071894	1.035323

ตารางที่ ค.9 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ เมื่อ $n = 15$ การแจกแจงแบบโคชี

Cauchy		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.error$	MSE	$Std.error$
	0.1	16790.5	129.5782	0.6582165	0.8113054
	0.2	9.05E+03	95.13253	0.8156667	0.9031427
	0.3	3.26E+05	571.2711	0.5590166	0.7476742
	0.4	8.98E+04	299.5972	42.49066	6.518486
15	0.5	3.83E+05	618.9955	4.391799	2.095662
	0.6	8.86E+02	29.76327	10.10806	3.179317
	0.7	3.67E+05	605.7881	4.685716	2.164651
	0.8	7.47E+04	273.2994	2.956788	1.719531
	0.9	1.47E+05	383.9889	9.062405	3.010383

ตารางที่ ค.10 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ เมื่อ $n=30$ การแจกแจงแบบโคชี

Cauchy		$h_{SORT}=0.9A(n^{\frac{1}{5}})$		$h_{ROT}=1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\ error$	MSE	$Std.\ error$
	0.1	42558.58	206.2973	1.318778	1.14838
	0.2	2.05E+05	452.7375	5.027516	2.242212
	0.3	5.01E+04	223.7958	25.20197	5.020156
	0.4	3.43E+04	185.2277	38.69058	6.220175
30	0.5	4.96E+04	222.6261	2.02698	1.423721
	0.6	7.74E+04	278.2094	3.326179	1.823781
	0.7	1.99E+05	446.2979	10.89838	3.301269
	0.8	1.83E+04	135.1675	17.84559	4.224404
	0.9	1.93E+04	139.0552	11.23695	3.352156

ตารางที่ ค.11 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\ error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ เมื่อ $n=50$ การแจกแจงแบบโคชี

Cauchy		$h_{SORT}=0.9A(n^{\frac{1}{5}})$		$h_{ROT}=1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\ error$	MSE	$Std.\ error$
	0.1	318241.9	564.1293	0.4523936	0.6726021
	0.2	3.14E+03	56.0247	24.98548	4.998548
	0.3	1.38E+03	37.11695	2.0848	1.443884
	0.4	8.55E+03	92.48325	2.522758	1.588319
50	0.5	3.12E+04	176.5048	1.808578	1.344834
	0.6	3.91E+03	62.56735	7.499222	2.738471

ตารางที่ ค.11 (ต่อ)

Cauchy		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.7	2.49E+04	157.8609	1.309465	1.144319
	0.8	1.38E+05	371.2015	13.13541	3.62428
	0.9	1.21E+04	109.8761	1.126489	1.061362

ตารางที่ ค.12 ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน ($Std.\,error$) ของวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่ เมื่อ $n = 100$ การแจกแจงแบบโคชี

Cauchy		$h_{SORT} = 0.9A(n^{\frac{1}{5}})$		$h_{ROT} = 1.06\hat{\sigma}(n^{\frac{1}{5}})$	
n	ρ	MSE	$Std.\,error$	MSE	$Std.\,error$
	0.1	5674.845	75.33157	11.76728	3.430347
	0.2	7.35E+02	27.11286	41.94124	6.476206
	0.3	2.77E+03	52.62469	1.174519	1.083752
	0.4	3.41E+02	18.46832	11.54988	3.398511
100	0.5	8.63E+04	293.692	1.362921	1.167442
	0.6	2.54E+03	50.41717	5.59004	2.364326
	0.7	1.18E+04	108.7378	0.7780196	0.8820542
	0.8	8.11E+04	284.8054	1.322509	1.150004
	0.9	3.97E+04	199.3697	0.7656246	0.8749998

ภาคผนวก ง

ข้อมูลราคาข้าวที่ปรับโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ และOUTPUT
จากโปรแกรมสำเร็จรูป

ตารางที่ ง.1 ข้อมูลราคาข้าวเปลือกเจ้า 15% ที่ปรับโดยวิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่

ไตรมาสที่	ราคาข้าว	x-xi	(x-xi)^2	คำนวณ	res1	a_t
1	6955	2868.46667	8228101.018	140935.9413	-3381.71093	6954.999785
2	7439	2384.46667	5685681.284	2683355.675	-2880.31282	7438.999785
3	9453	370.466667	137245.5511	8231791.408	-848.91471	9452.999785
4	13259	-3435.5333	11802889.28	-3433852.325	2974.4834	13258.99978
5	13042	-3218.5333	10358956.82	-1989919.859	2774.88151	13041.99978
6	12162	-2338.5333	5468738.151	2900298.808	1912.27962	12161.99978
7	12203	-2379.5333	5662178.884	2706858.075	1970.67773	12202.99978
8	11592	-1768.5333	3127710.151	5241326.808	1377.07585	11591.99978
9	11857	-2033.5333	4135257.818	4233779.141	1659.47396	11856.99978
10	10692	-868.53333	754350.1511	7614686.808	511.87207	10691.99978
11	9843	-19.533333	381.5511111	8368655.408	-319.72982	9842.999785
12	9618	205.466667	42216.55111	8326820.408	-527.33171	9617.999785
13	10005	-181.53333	32954.35111	8336082.608	-122.9336	10004.99978
14	10029	-205.53333	42243.95111	8326793.008	-81.53549	10028.99978
15	10392	-568.53333	323230.1511	8045806.808	298.86262	10391.99978
16	10047	-223.53333	49967.15111	8319069.808	-28.73927	10046.99978
17	10003	-179.53333	32232.21778	8336804.741	-55.34116	10002.99978
18	10192	-368.53333	135816.8178	8233220.141	151.05695	10191.99978
19	10182	-358.53333	128546.1511	8240490.808	158.45506	10181.99978
20	9200	623.466667	388710.6844	7980326.275	-806.14683	9199.999785
21	9318	505.466667	255496.5511	8113540.408	-670.74872	9317.999785
22	8460	1363.46667	1859041.351	6509995.608	-1511.35061	8459.999785
23	8948	875.466667	766441.8844	7602595.075	-1005.9525	8947.999785
24	9825	-1.5333333	2.351111111	8369034.608	-111.55439	9824.999785
25	9326	497.466667	247473.0844	8121563.875	-593.15627	9325.999785
26	9400	423.466667	179324.0178	8189712.941	-501.75816	9399.999785
27	8636	1187.46667	1410077.084	6958959.875	-1248.36005	8635.999785
28	8257	1566.46667	2453817.818	5915219.141	-1609.96194	8256.999785
29	7916	1907.46667	3638429.084	4730607.875	-1933.56383	7915.999785
30	7985	1838.46667	3379959.684	4989077.275	-1847.16572	7984.999785
31	8993	830.466667	689674.8844	7679362.075	-821.76761	8992.999785
32	9523	300.466667	90280.21778	8278756.741	-274.3695	9522.999785
33	9949	-125.53333	15758.61778	8353278.341	169.02861	9948.999785

ตารางที่ ง.1 (ต่อ)

ไตรมาสที่	ราคาข้าว	x-xi	(x-xi)^2	คำนวณ	res1	a_t
34	10217	-393.53333	154868.4844	8214168.475	454.42672	10216.99978
35	10305	-481.53333	231874.3511	8137162.608	559.82483	10304.99978
36	10322	-498.53333	248535.4844	8120501.475	594.22294	10321.99978
37	9641	182.466667	33294.08444	8335742.875	-69.37895	9640.999785
38	9645	178.466667	31850.35111	8337186.608	-47.98084	9644.999785
39	9921	-97.533333	9512.751111	8359524.208	245.41727	9920.999785
40	10000	-176.53333	31164.01778	8337872.941	341.81538	9999.999785
41	10242	-418.53333	175170.1511	8193866.808	601.21349	10241.99978
42	10467	-643.53333	414135.1511	7954901.808	843.61161	10466.99978
3	10326	-502.53333	252539.7511	8116497.208	720.00972	10325.99978
44	9914	-90.533333	8196.284444	8360840.675	325.40783	9913.999785
45	10584	-760.53333	578410.9511	7790626.008	1012.80594	10583.99978
46	10123	-299.53333	89720.21778	8279316.741	569.20405	10122.99978
47	10160	-336.53333	113254.6844	8255782.275	623.60216	10159.99978
48	10221	-397.53333	158032.7511	8211004.208	702.00027	10220.99978
49	10526	-702.53333	493553.0844	7875483.875	1024.39838	10525.99978
50	10559	-735.53333	541009.2844	7828027.675	1074.79649	10558.99978
51	10158	-334.53333	111912.5511	8257124.408	691.1946	10157.99978
52	9753	70.4666667	4965.551111	8364071.408	303.59271	9752.999785
53	9799	24.4666667	598.6177778	8368438.341	366.99082	9798.999785
54	9897	-73.533333	5407.151111	8363629.808	482.38893	9896.999785
55	9578	245.466667	60253.88444	8308783.075	180.78704	9577.999785
56	9329	494.466667	244497.2844	8124539.675	-50.81485	9328.999785
57	8480	1343.46667	1804902.684	6564134.275	-882.41674	8479.999785
58	8280	1543.46667	2382289.351	5986747.608	-1065.01863	8279.999785
59	8126	1697.46667	2881393.084	5487643.875	-1201.62051	8125.999785
60	8134	1689.46667	2854297.618	5514739.341	-1176.2224	8133.999785
คำนวณ						
std	1203.37678		สูตร	416703392.6	ε_i	$\rho =$-0.252
mean	9823.46667		3	1250110178	1.89478E-14	
ค่า	1.06	1.1236	1.1910			-4.77485E-15
std^2	1448115.68		4	5.81065E+12		0.000215141
std^3	1742628782		ผลลัพธ์	0.000215141		

ตารางที่ ง.1 (ต่อ)

ไตรมาสที่	ราคาข้าว	x-xi	(x-xi)^2	คำนวณ	res1	a_t
n^8/5	60	699.909681		0.000215141		
n^2/5	60	5.1435208				

ตารางที่ ง.2 ข้อมูลราคาข้าวเปลือกเจ้า 15% ของไทย

เดือนที่	ราคาข้าว(บาท)	เดือนที่	ราคาข้าว(บาท)	เดือนที่	ราคาข้าว(บาท)
1	6955	21	9318	41	10242
2	7439	22	8460	42	10467
3	9453	23	8948	43	10326
4	13259	24	9825	44	9914
5	13042	25	9326	45	10584
6	12162	26	9400	46	10123
7	12203	27	8636	47	10160
8	11592	28	8257	48	10221
9	11857	29	7916	49	10526
10	10692	30	7985	50	10559
11	9843	31	8993	51	10158
12	9618	32	9523	52	9753
13	10005	33	9949	53	9799
14	10029	34	10217	54	9897
15	10392	35	10305	55	9578
16	10047	36	10322	56	9329
17	10003	37	9641	57	8480
18	10192	38	9645	58	8280
19	10182	39	9921	59	8126
20	9200	40	10000	60	8134

ที่มา: ส่วนปฏิบัติการข้อมูลการเกษตร ศูนย์สารสนเทศการเกษตร สำนักงานเศรษฐกิจการเกษตร กระทรวงเกษตรและสหกรณ์

ตารางที่ ง.3 Model Description

Model Description			
			Model Type
Model ID	at	Model_1	ARIMA(1,0,0)

ตารางที่ ง.4 Model Fit

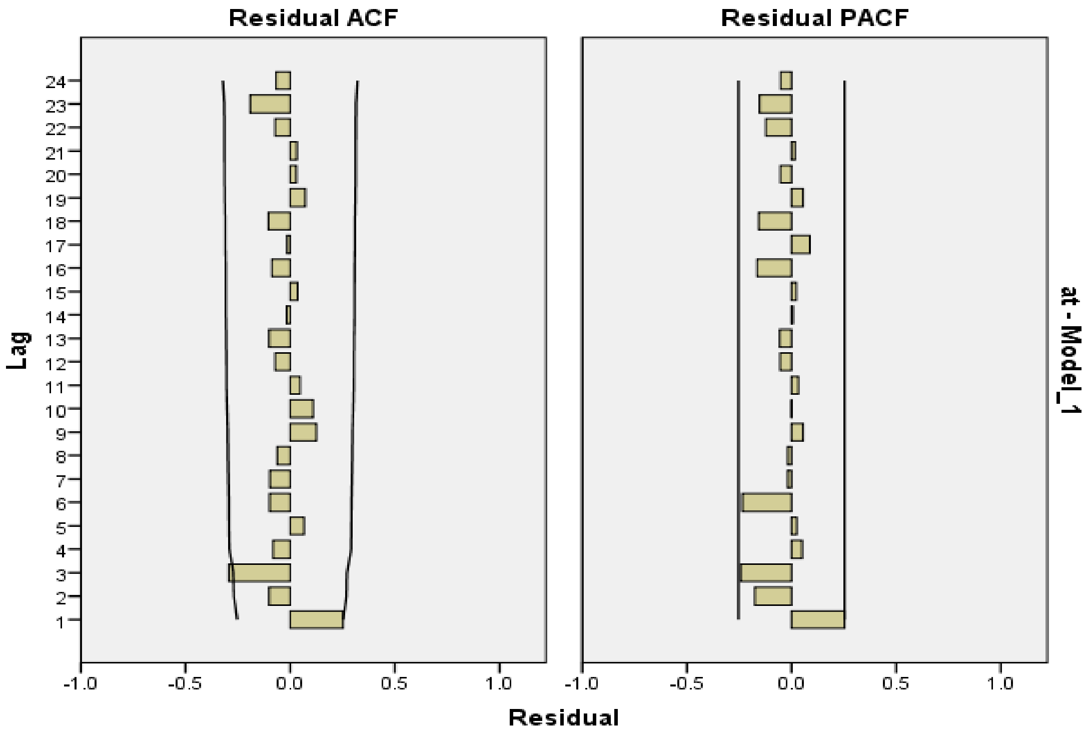
Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.603	.	.603	.603	.603	.603	.603	.603	.603	.603	.603
R-squared	.603	.	.603	.603	.603	.603	.603	.603	.603	.603	.603
RMSE	764.71	.	764.71	764.71	764.71	764.71	764.71	764.71	764.71	764.71	764.71
MAPE	4.841	.	4.841	4.841	4.841	4.841	4.841	4.841	4.841	4.841	4.841
MaxAPE	36.101	.	36.101	36.101	36.101	36.101	36.101	36.101	36.101	36.101	36.101
MAE	466.83	.	466.83	466.83	466.83	466.83	466.83	466.83	466.83	466.83	466.83
MaxAE	3804.0	.	3804.055	3804.055	3804.055	3804.055	3804.055	3804.055	3804.055	3804.055	3804.055
Normalized BIC	13.415	.	13.415	13.415	13.415	13.415	13.415	13.415	13.415	13.415	13.415

ตารางที่ ง.5 ARIMA Model Parameters

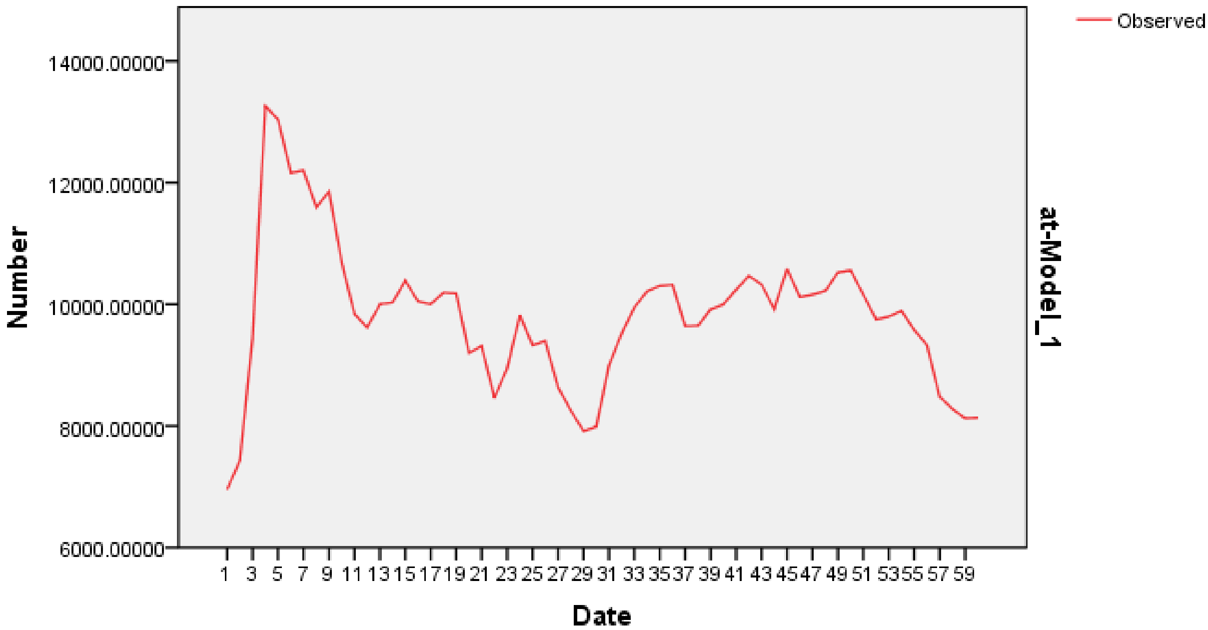
ARIMA Model Parameters				Estimate	SE	t	Sig.
at-Model_1	No Transformation	Constant		9465.825	559.074	16.931	.000
		AR	Lag 1	.848	.069	12.293	.000

ตารางที่ ง.6 การตรวจสอบความเหมาะสมของตัวแบบ

Model Statistics								
Model	Number of Predictors	Model Fit statistics			Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	R-squared	RMSE	Statistics	DF	Sig.	
at-Model_1	0	.603	.603	764.718	17.831	17	.400	0



ภาพที่ ง-1 Residual ACF และResidual PACF



ภาพที่ ง-2 กราฟของค่าพยากรณ์

ตารางที่ ง.7 One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Noise residual from at-Model_1	60	35.4033592	757.36866217	97.77587385

ตารางที่ ง.8 One-Sample Test Test Value=0

	Test Value = 0				
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Noise residual from at-Model_1	.362	59	.719	35.40335921	-160.2457125	231.0524309

ตารางที่ ง.9 ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	557288.018	1	557288.018	.971	.329[b]
	Residual	33285542.118	58	573888.657		
	Total	33842830.136	59			

a. Dependent Variable: Noise residual from at-Model_1

b. Predictors: (Constant), time

ตารางที่ ง.10 ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	730937.674	1	730937.674	.933	.340[b]
	Residual	32909598.566	42	783561.871		
	Total	33640536.240	43			

a. Dependent Variable: Noise residual from at-Model_1

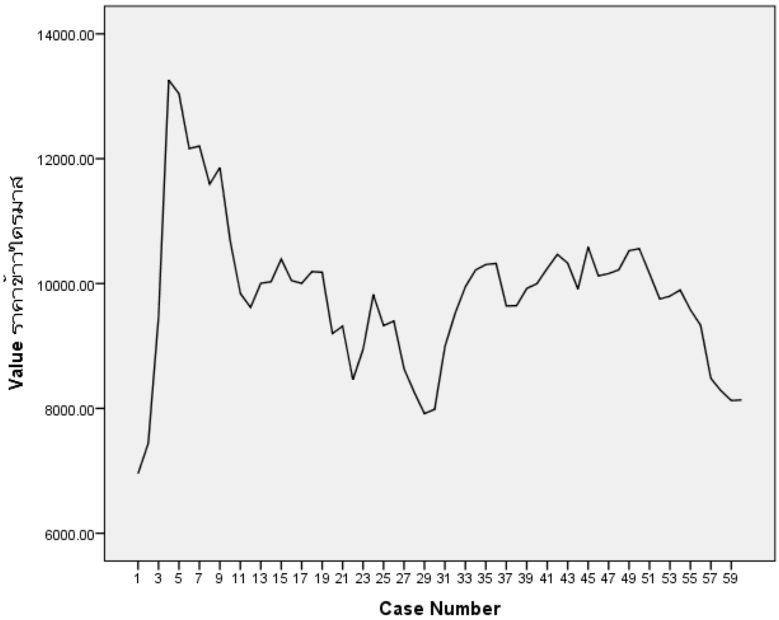
b. Predictors: (Constant), time

ตารางที่ ง.11 ตารางค่า F

		F (คำนวณ)	F (เปิดตาราง)
MS(Residual1)	783561.871		F(0.05,20,20)
MS(Residual2)	573888.657	1.365355216	2.124155213

ตารางที่ ง.12 print out

Graph ราคาข้าว



Model Description

Model Name		MOD_1
Series Name	1	ราคาข้าวไตรมาส
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		No periodicity
Maximum Number of Lags		16
Process Assumed for Calculating the Standard Errors of the Autocorrelations		Independence (white noise) [a]
Display and Plot		All lags

Applying the model specifications from MOD_1

a. Not applicable for calculating the standard errors of the partial autocorrelations.

Case Processing Summary

		ราคาข้าวไตรมาส
Series Length		60
Number of Missing Values	User-Missing	0
	System-Missing	0
Number of Valid Values		60
Number of Computable First Lags		59

ราคาข้าวไตรมาส

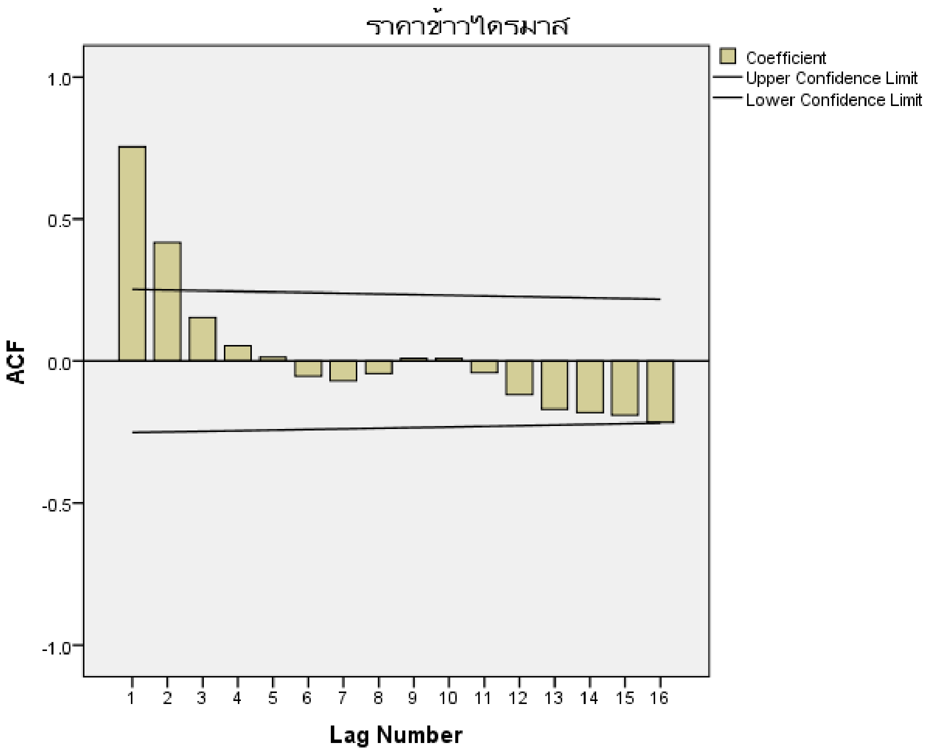
Autocorrelations

Series: ราคาข้าวไดรมาส

Lag	Autocorrelation	Std. Error[a]	Box-Ljung Statistic Value	df	Sig.[b]
1	.754	.126	35.884	1	.000
2	.417	.125	47.030	2	.000
3	.153	.124	48.552	3	.000
4	.054	.123	48.744	4	.000
5	.014	.122	48.756	5	.000
6	-.054	.120	48.956	6	.000
7	-.069	.119	49.290	7	.000
8	-.044	.118	49.429	8	.000
9	.010	.117	49.436	9	.000
10	.009	.116	49.442	10	.000
11	-.041	.115	49.570	11	.000
12	-.117	.114	50.636	12	.000
13	-.169	.112	52.904	13	.000
14	-.181	.111	55.559	14	.000
15	-.190	.110	58.555	15	.000
16	-.216	.109	62.487	16	.000

a. The underlying process assumed is independence (white noise).

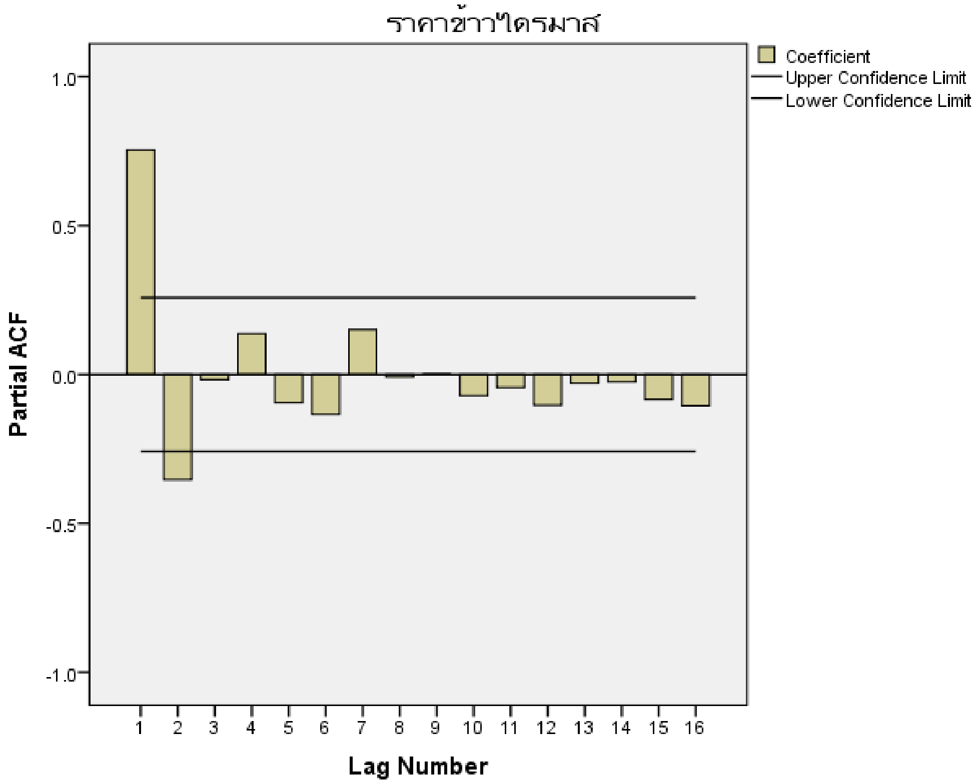
b. Based on the asymptotic chi-square approximation.



Partial Autocorrelations

Series: ราคาข้าวไตรมาส

Lag	Partial Autocorrelation	Std. Error
1	.754	.129
2	-.353	.129
3	-.017	.129
4	.137	.129
5	-.095	.129
6	-.134	.129
7	.151	.129
8	-.009	.129
9	.001	.129
10	-.071	.129
11	-.044	.129
12	-.103	.129
13	-.029	.129
14	-.024	.129
15	-.084	.129
16	-.105	.129



Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	time[b]	.	Enter

a. Dependent Variable: ราคาข้าวไตรมาส

b. All requested variables entered.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.252[a]	.064	.048	1174.381

a. Predictors: (Constant), time

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	5446983	1	5446983	3.949	.052[b]
	Residual	8.00E+7	58	1379170		
	Total	8.54E+7	59			

a. Dependent Variable: ราคาข้าวไตรมาส

b. Predictors: (Constant), time

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10354.11	307.054		33.721	.000
	time	-17.398	8.755	-.252	-1.987	.052

a. Dependent Variable: ราคาข้าวไตรมาส

Time Series Modeler

Model Description

			Model Type
Model ID	ราคาข้าวไตรมาส	Model_1	ARIMA(2,0,0)

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.652	.	.652	.652	.652	.652	.652	.652	.652	.652	.652
R-squared	.652	.	.652	.652	.652	.652	.652	.652	.652	.652	.652
RMSE	721.976	.	721.976	721.976	721.976	721.976	721.976	721.976	721.976	721.976	721.976
MAPE	4.765	.	4.765	4.765	4.765	4.765	4.765	4.765	4.765	4.765	4.765
MaxAPE	39.848	.	39.848	39.848	39.848	39.848	39.848	39.848	39.848	39.848	39.848
MAE	457.389	.	457.389	457.389	457.389	457.389	457.389	457.389	457.389	457.389	457.389
MaxAE	2839.759	.	2839.759	2839.759	2839.759	2839.759	2839.759	2839.759	2839.759	2839.759	2839.759
Normalized BIC	13.369	.	13.369	13.369	13.369	13.369	13.369	13.369	13.369	13.369	13.369

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
ราคาข้าวไตรมาส-Model_1	0	.652	13.308	16	.650	0

ARIMA Model Parameters

					Estimate	SE	t	Sig.
ราคาข้าวไตรมาส-Model_1	ราคาข้าวไตรมาส	No Transformation	Constant		9726.422	319.381	30.454	.000
			AR	Lag 1	1.186	.120	9.909	.000
				Lag 2	-.445	.120	-3.701	.000

Residual ACF

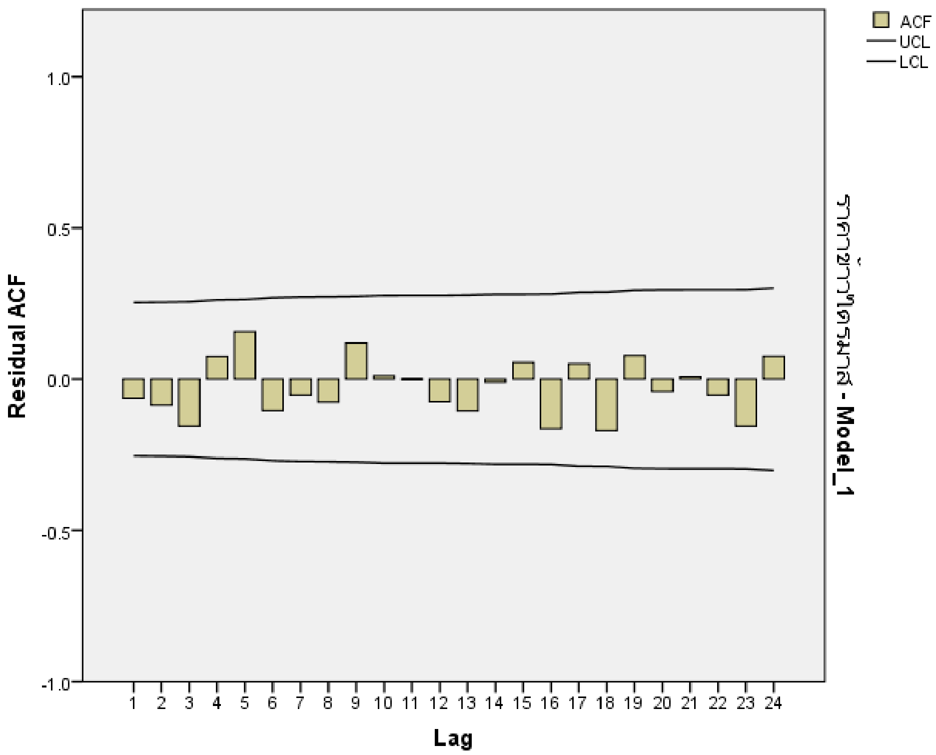
Model		1	2	3	4	5	6	7	8	9	10	11	12
ราคาข้าวไตรมาส-Model_1	ACF	-.063	-.085	-.154	.075	.158	-.104	-.052	-.075	.120	.011	.000	-.073
	SE	.129	.130	.131	.134	.134	.137	.139	.139	.140	.141	.141	.141

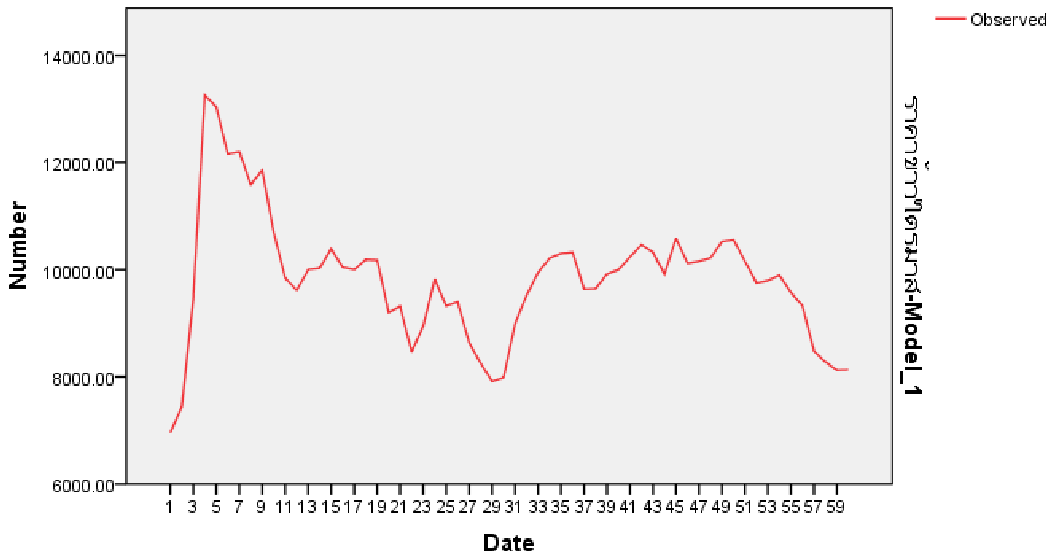
13	14	15	16	17	18	19	20	21	22	23	24
-.105	-.009	.056	-.164	.051	-.170	.078	-.041	.008	-.053	-.155	.075
.142	.143	.143	.144	.147	.147	.150	.151	.151	.151	.151	.154

Residual PACF

Model		1	2	3	4	5	6	7	8	9	10	11	12
ราคาข้าวไตรมาส-Model_1	PACF	-.063	-.089	-.168	.045	.143	-.100	-.026	-.056	.060	-.006	.028	-.040
	SE	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129

13	14	15	16	17	18	19	20	21	22	23	24
-.119	-.073	.033	-.208	.069	-.187	-.004	-.083	.001	-.095	-.151	-.035
.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129


ACF
UCL
LCL
Residual ACF
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Lag
ราคาข้าวไตรมาส - Model_1


Observed
Number
14000.00
12000.00
10000.00
8000.00
6000.00
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Date
ราคาข้าวไตรมาส-Model_1

PPlot

Model Description

Model Name		MOD_2
Series or Sequence	1	Noise residual from at-Model_1
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		No periodicity
Standardization		Not applied
Distribution	Type	Normal
	Location	estimated
	Scale	estimated
Fractional Rank Estimation Method		Blom's
Rank Assigned to Ties		Mean rank of tied values

Applying the model specifications from MOD_2

Case Processing Summary

		Noise residual from at-Model_1
Series or Sequence Length		60
Number of Missing Values in the Plot	User-Missing	0
	System-Missing	0

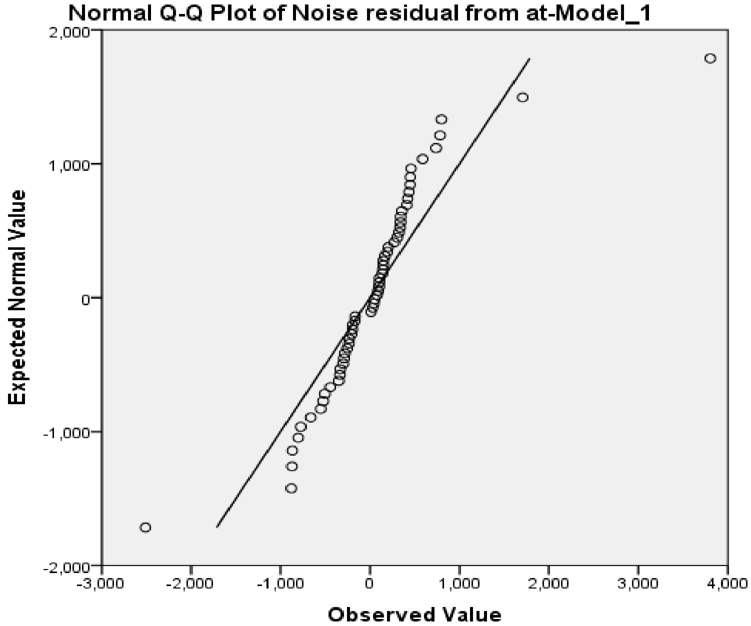
The cases are unweighted.

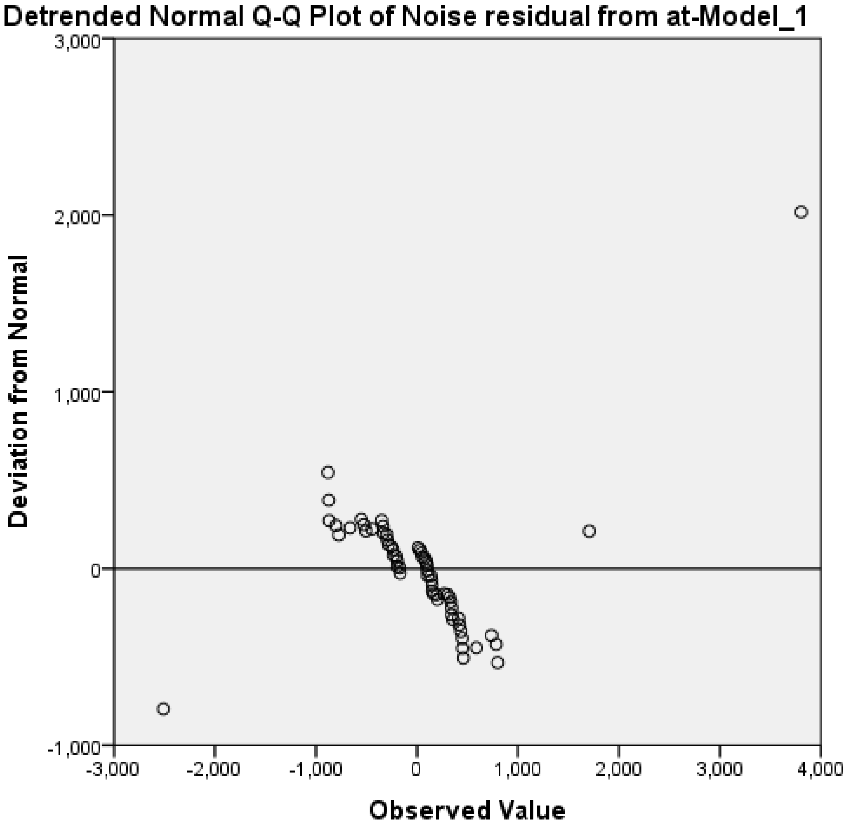
Estimated Distribution Parameters

		Noise residual from at-Model_1
Normal Distribution	Location	35.40336
	Scale	757.3687

The cases are unweighted.

Noise residual from at-Model_1





```
NPAR TESTS
  /K-S(NORMAL)=NResidual_at_Model_1_A
  /MISSING ANALYSIS.
```

NPar Tests

One-Sample Kolmogorov-Smirnov Test

		Noise residual from at-Model_1
N		60
Normal Parameters[a,b]	Mean	35.40336
	Std. Deviation	757.3687
Most Extreme Differences	Absolute	.188
	Positive	.188
	Negative	-.122
Test Statistic		.188
Asymp. Sig. (2-tailed)		.000[c]

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Model Summary

```
COMPUTE a=2 / sqrt (60).
EXECUTE.
NPAR TESTS
  /K-S(NORMAL)=NResidual_ราคาข้าวไตรมาส_Model_1_A
  /MISSING ANALYSIS.
```

NPar Tests

Time Series Modeler

Model Description

			Model Type
Model ID	ราคาข้าวไตรมาส	Model_1	ARIMA(0,0,2)

Model Summary

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.600	.	.600	.600	.600	.600	.600	.600	.600	.600	.600
R-squared	.600	.	.600	.600	.600	.600	.600	.600	.600	.600	.600
RMSE	773.973	.	773.973	773.973	773.973	773.973	773.973	773.973	773.973	773.973	773.973
MAPE	5.307	.	5.307	5.307	5.307	5.307	5.307	5.307	5.307	5.307	5.307
MaxAPE	40.520	.	40.520	40.520	40.520	40.520	40.520	40.520	40.520	40.520	40.520
MAE	503.608	.	503.608	503.608	503.608	503.608	503.608	503.608	503.608	503.608	503.608
MaxAE	2973.061	.	2973.061	2973.061	2973.061	2973.061	2973.061	2973.061	2973.061	2973.061	2973.061
Normalized BIC	13.508	.	13.508	13.508	13.508	13.508	13.508	13.508	13.508	13.508	13.508

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18) Statistics	DF	Sig.	Number of Outliers
ราคาข้าวไตรมาส-Model_1	0	.600	12.621	16	.700	0

ARIMA Model Parameters

					Estimate	SE	t	Sig.
ราคาข้าวไตรมาส-Model_1	ราคาข้าวไตรมาส	No Transformation	Constant		9773.134	225.938	43.256	.000
			MA	Lag 1	-.944	.116	-8.112	.000
				Lag 2	-.531	.119	-4.445	.000

Residual ACF

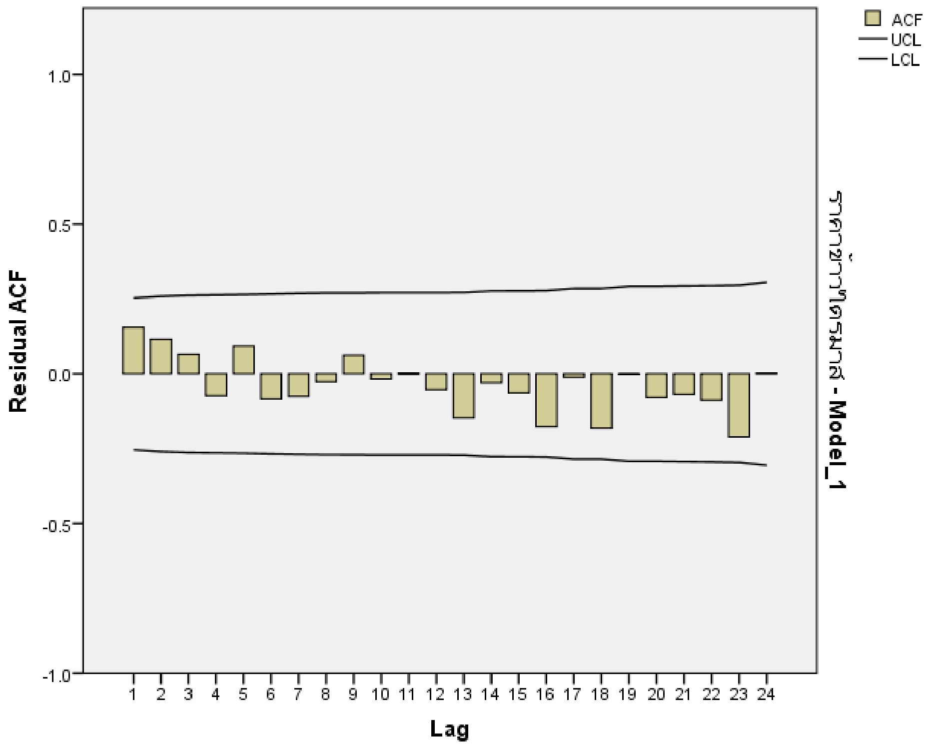
Model		1	2	3	4	5	6	7	8	9	10	11	12
ราคาข้าวไตรมาส-Model_1	ACF	.155	.116	.064	-.072	.092	-.083	-.075	-.027	.062	-.018	.001	-.052
	SE	.129	.132	.134	.134	.135	.136	.137	.138	.138	.138	.138	.138

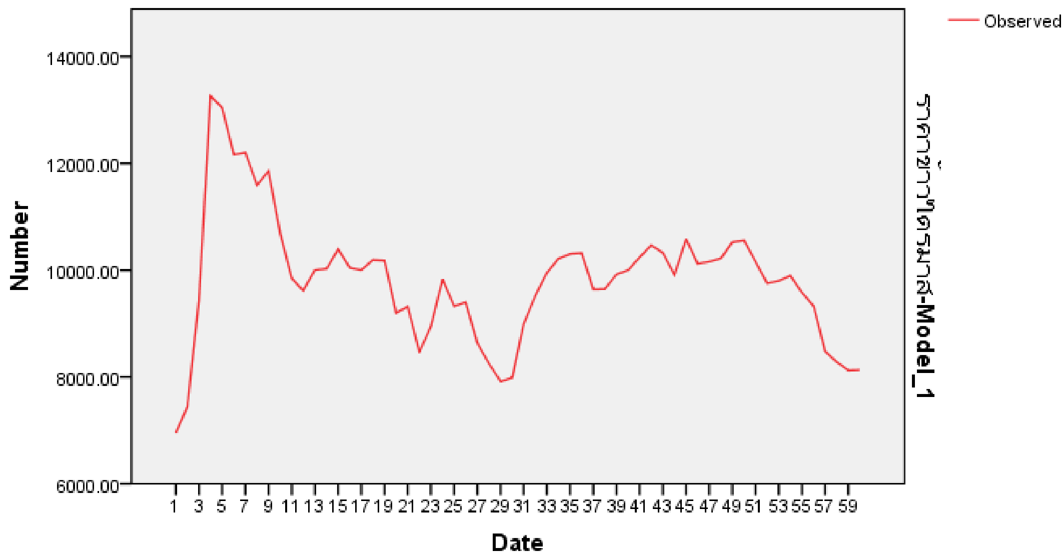
13	14	15	16	17	18	19	20	21	22	23	24
-.146	-.030	-.063	-.176	-.011	-.181	-.002	-.078	-.068	-.088	-.211	.002
.138	.141	.141	.142	.145	.145	.149	.149	.150	.150	.151	.156

Residual PACF

Model		1	2	3	4	5	6	7	8	9	10	11	12
ราคาข้าวไตรมาส-Model_1	PACF	.155	.094	.035	-.100	.111	-.102	-.062	-.007	.117	-.068	.004	-.055
	SE	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129

13	14	15	16	17	18	19	20	21	22	23	24
-.126	-.017	.001	-.170	.038	-.157	.041	-.131	.020	-.154	-.170	.012
.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129


ACF
UCL
LCL
Residual ACF
1.0
0.5
0.0
-0.5
-1.0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Lag
ราคาข้าวไตรมาส - Model_1


Observed
Number
14000.00
12000.00
10000.00
8000.00
6000.00
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59
Date
ราคาข้าวไตรมาส-Model_1

NPar Tests

One-Sample Kolmogorov-Smirnov Test

		Noise residual from ราคาข้าวไตรมาส-Model_1
N		60
Normal Parameters[a,b]	Mean	1.1336
	Std. Deviation	760.7404
Most Extreme Differences	Absolute	.162
	Positive	.162
	Negative	-.100
Test Statistic		.162
Asymp. Sig. (2-tailed)		.000[c]

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

Time Series Modeler

Model Description

			Model Type
Model ID	ราคาข้าวไตรมาส	Model_1	ARIMA(0,0,1)

Model Summary

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	.506	.	.506	.506	.506	.506	.506	.506	.506	.506	.506
R-squared	.506	.	.506	.506	.506	.506	.506	.506	.506	.506	.506
RMSE	852.862	.	852.862	852.862	852.862	852.862	852.862	852.862	852.862	852.862	852.862
MAPE	6.078	.	6.078	6.078	6.078	6.078	6.078	6.078	6.078	6.078	6.078
MaxAPE	40.737	.	40.737	40.737	40.737	40.737	40.737	40.737	40.737	40.737	40.737
MAE	579.197	.	579.197	579.197	579.197	579.197	579.197	579.197	579.197	579.197	579.197
MaxAE	3265.538	.	3265.538	3265.538	3265.538	3265.538	3265.538	3265.538	3265.538	3265.538	3265.538
Normalized BIC	13.634	.	13.634	13.634	13.634	13.634	13.634	13.634	13.634	13.634	13.634

Model Statistics

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Ljung-Box Q(18): Statistics	DF	Sig.	Number of Outliers
ราคาข้าวไตรมาส-Model_1	0	.506	29.912	17	.027	0

ARIMA Model Parameters

					Estimate	SE	t	Sig.
ราคาข้าวไตรมาส-Model_1	ราคาข้าวไตรมาส	No Transformation	Constant		9788.283	188.771	51.853	.000
			MA	Lag 1	-.829	.080	-10.388	.000

Residual ACF

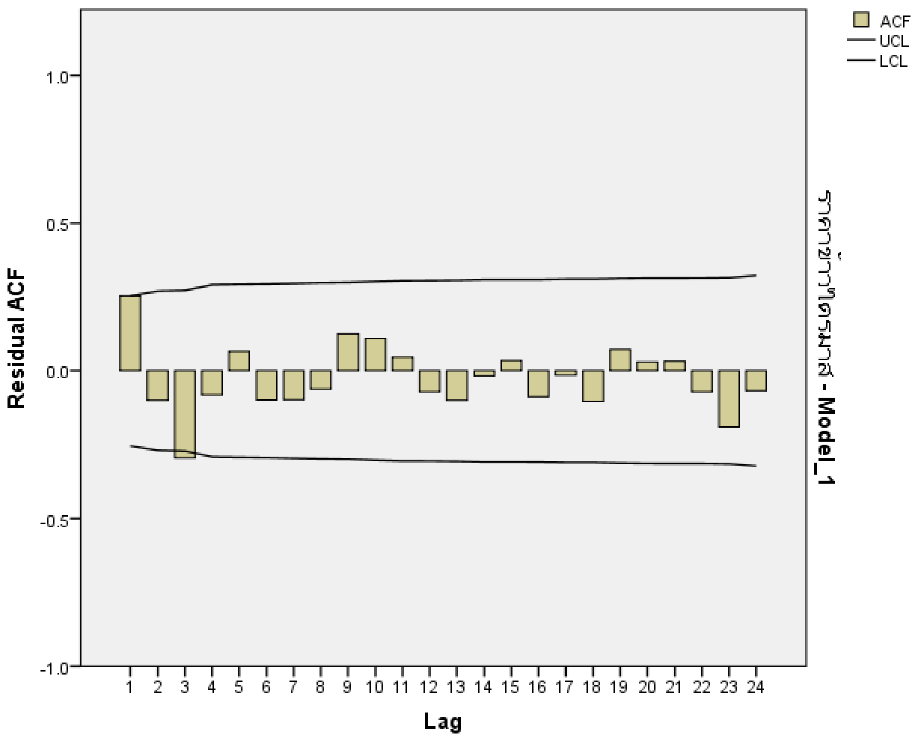
Model		1	2	3	4	5	6	7	8	9	10	11	12
ราคาข้าวไตรมาส-Model_1	ACF	.322	.364	-.076	.101	-.047	-.032	-.074	-.048	.042	-.011	.002	-.113
	SE	.129	.142	.157	.157	.158	.159	.159	.159	.159	.160	.160	.160

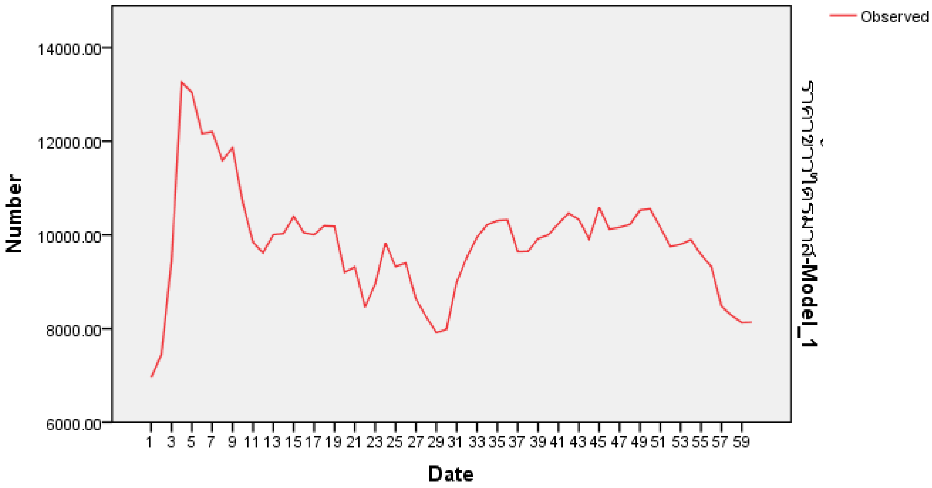
Model		13	14	15	16	17	18	19	20	21	22	23	24
ราคาข้าวไตรมาส-Model_1	ACF	-.100	-.017	.035	-.088	-.014	-.104	.072	.029	.032	-.072	-.191	-.067
	SE	.156	.157	.157	.157	.158	.158	.159	.160	.160	.160	.160	.164

Residual PACF

Model		1	2	3	4	5	6	7	8	9	10	11	12
ราคาข้าวไตรมาส-Model_1	PACF	.253	-.176	-.242	.051	.024	-.235	-.020	-.019	.054	.000	.034	-.054
	SE	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129

Model		13	14	15	16	17	18	19	20	21	22	23	24
ราคาข้าวไตรมาส-Model_1	PACF	-.056	.006	.022	-.164	.088	-.156	.055	-.053	.017	-.121	-.152	-.052
	SE	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129	.129





T-Test

One-Sample Statistics

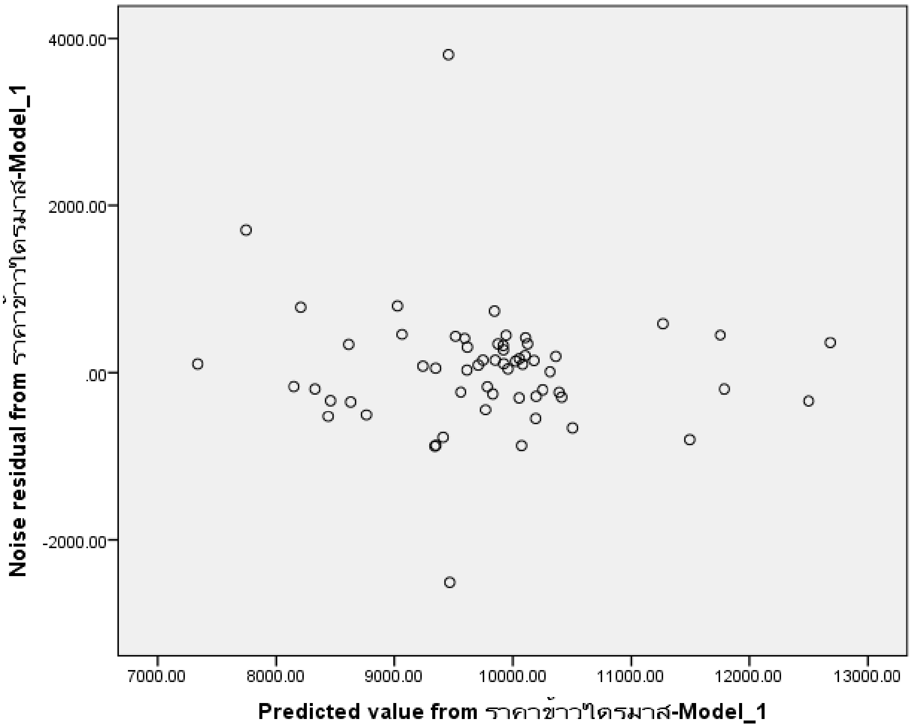
	N	Mean	Std. Deviation	Std. Error Mean
Noise residual from ราคาข้าวไตรมาส-Model_1	60	35.4034	757.3687	97.77587

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
Noise residual from ราคาข้าวไตรมาส-Model_1	.362	59	.719	35.40336	-160.246	231.0524

```
GRAPH
  /SCATTERPLOT(BIVAR)=Predicted_ราคาข้าวไตรมาส_Model_1_D WITH NResidual_ราคาข้าวไตรมาส_Model_1_D
  /MISSING=LISTWISE.
```

Graph




```
NPAR TESTS
  /CHISQUARE=NResidual_ราคาข้าวไตรมาส_Model_1_D
  /EXPECTED=EQUAL
  /MISSING ANALYSIS.
```

NPar Tests

Chi-Square Test

Test Statistics

	Noise residual from ราคาข้าวไตรมาส-Model_1
Chi-Square	.000[a]
df	59
Asymp. Sig.	1.000

a. 60 cells (100.0%) have expected frequencies less than 5. The minimum expected cell frequency is 1.0.

PPlot

[DataSet1] G:\ \ดุษฎีส่ง 270759\ราคาข้าวแก้ไข110659.sav

Model Description

Model Name		MOD_1
Series or Sequence	1	Noise residual from ราคาข้าวไตรมาส-Model_1
Transformation		None
Non-Seasonal Differencing		0
Seasonal Differencing		0
Length of Seasonal Period		No periodicity
Standardization		Not applied
Distribution	Type	Normal
	Location	estimated
	Scale	estimated
Fractional Rank Estimation Method		Blom's
Rank Assigned to Ties		Mean rank of tied values

Applying the model specifications from MOD_1

Case Processing Summary

		Noise residual from ราคาข้าวไตรมาส-Model_1
Series or Sequence Length		60
Number of Missing Values in the Plot	User-Missing	0
	System-Missing	0

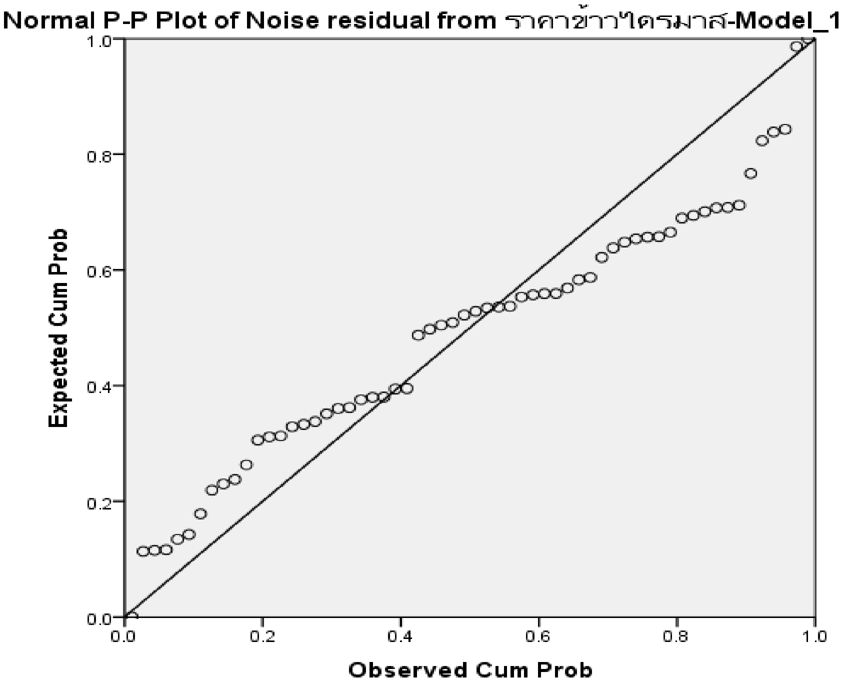
The cases are unweighted.

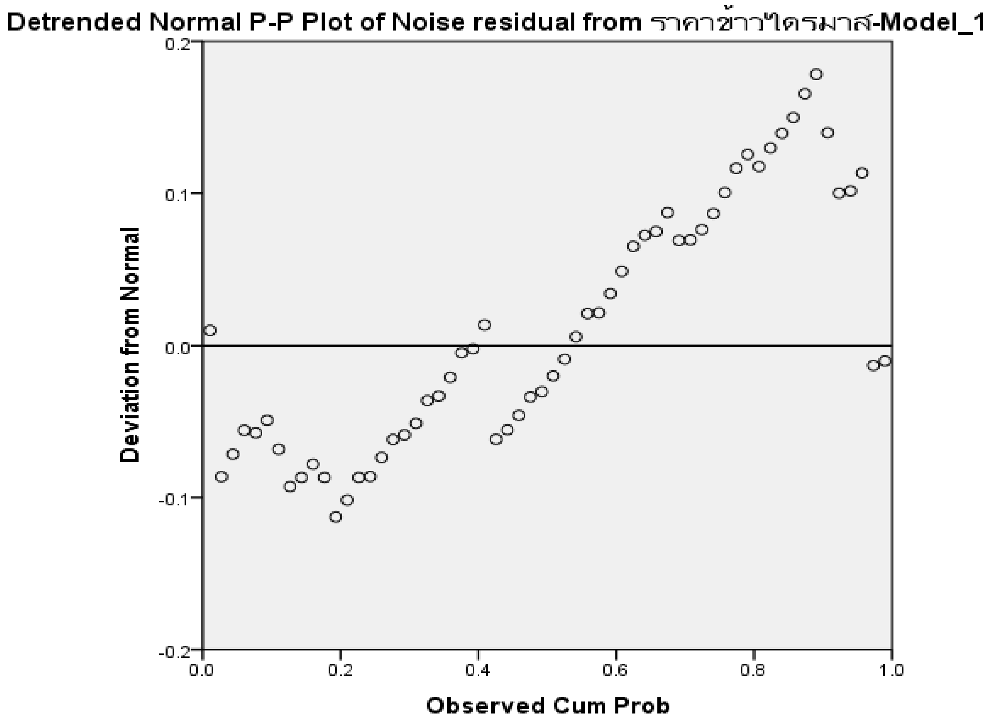
Estimated Distribution Parameters

		Noise residual from ราคาข้าวไตรมาส-Model_1
Normal Distribution	Location	35.4034
	Scale	757.3687

The cases are unweighted.

Noise residual from ราคาข้าวไดรมาส-Model_1





ผลงานวิจัยตีพิมพ์

สุภาพร ศรีหามี, วสันต์ เดือนแจ้ง และกรองทิพย์ เนียมถนอม. (2553). การศึกษาระดับพฤติกรรมทางคุณธรรมจริยธรรมของนักเรียนโรงเรียนมัธยมสาธิต มหาวิทยาลัยราชภัฏบ้านสมเด็จเจ้าพระยา. *ครุศาสตร์สาร*, *4*(1), 148-165.

ชัยยา น้อยนารถ, กิตติศักดิ์ จังพานิช, อัญชลี ชุ่มบัวทอง, พัชรา เกรียงไกร, ณัฏฐ์พัชร์ วณิชย์กุล, นุชนาฎ เลี้ยงอำนวย, วสันต์ เดือนแจ้ง, กรองทิพย์ เนียมถนอม และพูลพงศ์ สุขสว่าง. โมเดลความสันพันธ์เชิงสาเหตุการรับรู้โครงสร้างมหาวิทยาลัยและความผูกพันต่อองค์กรที่ส่งผลต่อความพึงพอใจในการปฏิบัติงานของอาจารย์ใน 3 จังหวัดชายแดนภาคใต้. *วารสารนานาชาติ มหาวิทยาลัยขอนแก่น, 6*(3). (in press).

วสันต์ เดือนแจ้ง, เสรี ชัดแช้ม และพัชรี วงษ์เกษม. การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยโดยใช้วิธีรีโปรดิวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่. *ครุศาสตร์สาร*, 11(2). (in press).