

การพยากรณ์ราคาข้าวเปลือกเจ้าของไทยโดยใช้ วิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับใหม่ Using an Adjusted Reproducing Kernel Hilbert Space to Forecast the Price of Thai Rice

วสันต์ เดือนแจ้ง*

เสรี ชัดรัมย์**

พัชรี วงษ์เกษม***

*โรงเรียนสาธิตแห่งมหาวิทยาลัยเกษตรศาสตร์ วิทยาเขตกำแพงแสน ศูนย์วิจัยและพัฒนาการศึกษา
คณะศึกษาศาสตร์และพัฒนศาสตร์

**วิทยาลัยวิทยาการวิจัยและวิทยาการปัญญา มหาวิทยาลัยบูรพา

***ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยบูรพา

*Corresponding author. E-mail : kent2513@gmail.com

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อ 1) พัฒนารูปแบบการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซแบบปรับใหม่ กรณีความคลาดเคลื่อนสุ่มมีอัตราส่วนสัมพันธ์ 2) เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับใหม่กับวิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ ของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ โดยใช้แบบวัด 2 วิธี คือ วิธีที่ 1 Rules of Thumb และวิธีที่ 2 Silverman's Rules of Thumb ขนาดตัวอย่าง 6 ขนาด (5, 10, 15, 30, 50, 100) และกำหนดค่าอัตราส่วนสัมพันธ์ 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) ตัวประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซแบบปรับใหม่ คือ

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

2) ผลการเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีโปรดิวิส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับใหม่กับวิธีโปรดิวิส์ซิงเคอร์เนลฮิลเบิร์ตสเปซของ Ferraty (2007) แบนวิดจ์วิธี Rules of Thumb ปรากฏว่า มี 52 สถานการณ์ ที่วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีโปรดิวิส์ซิงเคอร์เนลฮิลเบิร์ตสเปซแบบปรับใหม่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง และค่าความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีโปรดิวิส์ซิงเคอร์เนลฮิลเบิร์ต สเปซของ Ferraty (2007) 3) ตัวแบบที่ใช้ในการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทย คือ ARIMA (1, 0, 0) ซึ่งผลการพยากรณ์ชี้ให้เห็นว่า สมประสิทธิ์การทำนายสามารถอธิบายความแปรผันได้ร้อยละ 60.30

คำสำคัญ: ฟังก์ชันเคอร์เนล วิธีโปรดิวิส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ การแจกแจงแบบโคชี การแจกแจงแบบปกติ ราคาข้าว แบนวิดจ์

Abstract

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 52 situations where the

adjusted Reproducing Kernel Hilbert Space had lower mean square error (MSE) and lower standard error (Std.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.

Keywords: Kernel Function, Producing Kernel Hilbert Space, Cauchy Distribution, Normal Distribution, Grain Price, bandwidth

บทนำ

การพยากรณ์ราคาผลผลิตสินค้าเกษตร เป็นการพยากรณ์เชิงปริมาณโดยการสร้างแบบจำลองทางเศรษฐมิติ (Econometric Model) ด้วยวิธีวิเคราะห์การถดถอย (Regression Analysis) ทั้งพืชและปศุสัตว์ ซึ่งจะแสดงความสัมพันธ์เชิง เศรษฐศาสตร์ระหว่างตัวแปรที่เกี่ยวข้องกัน แบบจำลองที่สร้างขึ้น มีหลายรูปแบบ มีทั้งแบบจำลองที่เป็นเชิงเส้นและไม่เป็นเชิงเส้น ในการพยากรณ์ใช้การวิเคราะห์ข้อมูลแบบอนุกรมเวลาในปัจจุบันตัวแบบที่นิยมใช้พยากรณ์มีหลายวิธี เช่น ตัวแบบ ARIMA และ วิธีอื่น ๆ ในทางปฏิบัติส่วนใหญ่ปรากฏว่าข้อมูล ที่นำมาใช้ในการวิเคราะห์ไม่เป็นไปตามข้อตกลงเบื้องต้นบางข้อทำให้ข้อมูลเกิดปัญหา เช่นการเกิดอัตสหสัมพันธ์ของความคลาดเคลื่อนมักเกิดในกรณีที่ค่าของตัวแปรตามและตัวแปรต้นมีการเก็บตามช่วงเวลา เรียกว่าอนุกรมเวลา (Box et al, 1994: 96) ข้อมูลอนุกรมเวลา (Time Series Data) ส่วนใหญ่ข้อมูลเป็นข้อมูลทางด้านเศรษฐศาสตร์และด้านธุรกิจ และลักษณะของอัตสหสัมพันธ์ของความคลาดเคลื่อนที่พบโดยทั่วไปจะเป็นอัตสหสัมพันธ์ที่หนึ่งช่วงเวลา หรือ อันดับที่ 1 (Autocorrelat ที่ lag 1) ซึ่งเกิดขึ้นกับข้อมูลอนุกรมเวลารายปี

การประมาณความหนาแน่นแบบนอนพาราเมตริก ถือเป็นเครื่องมือสำคัญที่จะช่วยในการวิเคราะห์และนำเสนอโครงสร้างของข้อมูลได้อย่างมีประสิทธิภาพมีข้อตกลงเบื้องต้นน้อยกว่าการประมาณความหนาแน่นแบบอิงพารามิเตอร์ ไม่ต้องคำนึงถึงค่าพารามิเตอร์ใด ๆ กล่าวคือ เป็นการประมาณฟังก์ชันความหนาแน่นจากชุดข้อมูลโดยตรง สามารถทำได้หลายวิธีวิธีฮิสโตแกรมถือเป็นวิธีที่ง่ายและเก่าแก่ที่สุดนิยมใช้อย่างกว้างขวางเพื่อแสดงรูปร่างของฟังก์ชันความหนาแน่นของความน่าจะเป็น (Want & Jones, 1995: 5) ในเวลาต่อมาได้มีผู้คิดวิธีการต่าง ๆ หลายวิธีเพื่อประมาณความหนาแน่นของความน่าจะเป็น เช่น วิธีประมาณอย่างง่าย (Fix & Hodges, 1951: 233-247) วิธีประมาณความหนาแน่นแบบเคอร์เนล (Kernel

Density Estimation) ซึ่งถูกเสนอครั้งแรกโดย Fix and Hodge (1951, Quoted in Siliverman, 1986: 1-2) ตามมาด้วย Siliverman (1986) Rosenblatt (1956) Parzen (1962) วิธีฟังก์ชันถ่วงน้ำหนักทั่วไป (Whittle, 1958) วิธีเคอร์เนลแปรผัน (Breiman, Meisel & Purcell, 1977: 135-144) เป็นต้น

การประมาณความหนาแน่นแบบนอนพาราเมตริกด้วยวิธีเคอร์เนล (Kernel Density Estimation) เป็นวิธีหนึ่งที่สำคัญ การประมาณค่าความหนาแน่นแบบเคอร์เนล ถูกเสนอขึ้นอย่างเป็นทางการในปี ค.ศ.1956 โดย Murray Rosenblatt การประมาณค่าความหนาแน่นแบบเคอร์เนลเป็นวิธีนิยมอย่างแพร่หลายและเป็นวิธีที่มีประสิทธิภาพมากเพราะคุณสมบัติทางคณิตศาสตร์ที่เกี่ยวข้องสามารถทำความเข้าใจได้ไม่ยากปรับเปลี่ยน รูปแบบ ได้ง่าย เป็นพื้นฐานสำคัญที่นำไปสู่การศึกษาวิธีการประมาณความหนาแน่นแบบไมอิงพารามิเตอร์ วิธีอื่นการวิเคราะห์แบบรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซ (Reproducing Kernel Hilbert Space : RKHS) ช่วยให้สามารถสร้างแบบจำลองทางสถิติได้ดี การวิเคราะห์ข้อมูลที่มีความเจาะจงหลายรูปแบบในการสร้างตัวแบบทาง สถิติที่มีข้อมูลทั้งโดยตรงและโดยอ้อมสามารถจัดการกับปัญหาการกระจายของข้อมูล ที่หลากหลายโดยใช้ทฤษฎีต่าง ๆ เช่น ทฤษฎีของเกาส์ ใช้สกุลเลขยกกำลังทั่วไป (General Exponential Families) การประมาณที่มีค่าความแกร่ง (Robust Estimation) การสังเกตแบบช่วง (Interval Observations) ในแบบจำลองที่มีข้อจำกัดต่าง ๆ

จากการศึกษาวิธีการวิเคราะห์ข้อมูลของข้อมูลที่มีความคลาดเคลื่อนมีความสัมพันธ์กัน การใช้วิธีประมาณค่าพารามิเตอร์ด้วยวิธีกำลังสองน้อยที่สุดไม่มีประสิทธิภาพพอ และส่งผลให้ตัวประมาณค่าไม่มีคุณสมบัติเป็น BLUE โดยเฉพาะการพยากรณ์ข้อมูลทางเศรษฐศาสตร์ เพราะข้อมูลทางเศรษฐศาสตร์เป็นข้อมูลที่มีการจัดเก็บเป็นช่วงเวลาที่เราเรียกว่า ข้อมูลอนุกรมเวลาค่าสังเกตและค่าความคลาดเคลื่อนมีความสัมพันธ์กัน ดังนั้นในการพยากรณ์ข้อมูลทางเศรษฐศาสตร์จึงจำเป็นต้องตระหนักถึงวิธีการประมาณค่าที่เหมาะสมกับข้อมูลเป็นอย่างมาก เพราะถ้าข้อมูลที่ต้องการใช้ไม่เป็นไปตามข้อตกลงเบื้องต้น อาจทำให้เกิดความคลาดเคลื่อนในการพยากรณ์ได้ การมีวิธีการสร้างตัวแบบพยากรณ์สำหรับอนุกรมเวลาที่น่าเอาสหสัมพันธ์ที่ปรากฏไปวิเคราะห์ใช้ในปัจจุบันวิธีที่รู้จักและใช้กันมากคือวิธีกำลังสองน้อยที่สุดทั่วไป แต่วิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซ ของ Ferraty (2007) เป็นวิธีการแบบนอนพาราเมตริก (Nonparametric) ที่มีกรอบแนวคิดสำหรับการประมาณค่าฟังก์ชันที่มีความยืดหยุ่น ในการสร้างตัวแบบทางสถิติที่มีข้อมูลทั้งโดยตรงและโดยอ้อมที่สามารถจัดการกับปัญหาและการกระจายของข้อมูลที่หลากหลายการประมาณค่าที่มี

ค่าความแกร่ง (Robust Estimation) ผู้วิจัยจึงสนใจที่จะพัฒนาเพื่อนำไปใช้ในการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซแบบปรับใหม่ ให้สามารถควบคุมค่าความคลาดเคลื่อนที่เกิดขึ้นจากการพยากรณ์ให้มีความ ถูกต้องมากยิ่งขึ้นในกรณีที่ข้อมูลเป็นอนุกรมเวลาที่ความคลาดเคลื่อนมีอัตตสหสัมพันธ์ และนำวิธีประมาณค่าพารามิเตอร์ที่ปรับใหม่ ไปใช้ในการพยากรณ์ราคาข้าวเปลือกเจ้าล่วงหน้าของไทย โดยใช้ตัวแบบ ARIMA (p,d,q) ที่มีสมบัติทั่วไปคือ ARIMA (1,0,0) ในการพยากรณ์ราคาข้าวเปลือกเจ้าของไทยล่วงหน้า ซึ่งสอดคล้องกับงานวิจัยของ จิตราภรณ์ พันศิริ (2547) ที่ศึกษาการพยากรณ์ราคาส่งออกข้าวด้วยวิธี ARIMA และเพื่อให้ผู้ที่เกี่ยวข้องนำผลการศึกษาที่ได้ไปใช้ในเรื่องของการวางแผนการผลิตการเพิ่มมูลค่าสินค้าทางการเกษตรในด้านของราคาของผลผลิตที่มีประสิทธิภาพให้ดีขึ้นทั้งในด้านปริมาณและคุณภาพอันจะนำไปสู่การพัฒนาการผลิตที่ยั่งยืนต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซแบบปรับใหม่ กรณีความคลาดเคลื่อนสุ่ม มีอัตตสหสัมพันธ์
2. เพื่อเปรียบเทียบประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับใหม่ กับวิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ (2x6x9) ได้แก่ แบบวัดจ 2 วิธีคือวิธีที่ 1 Rules of thumb และวิธีที่ 2 Siliverman's rules of thumb ขนาดตัวอย่าง 6 ขนาด คือ 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
3. เพื่อพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยล่วงหน้าโดยใช้ตัวแบบ ARIMA

วิธีการดำเนินงาน

1. พัฒนารูปแบบการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตตสหสัมพันธ์
2. เปรียบเทียบประสิทธิภาพของวิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซ (RKHS) แบบปรับใหม่ กับวิธีรีโพรดิวซ์ซิงเคอร์เนลฮิลเบิร์ตสเปซ ของ Ferraty(2007) ภายใต้สถานการณ์ 108 สถานการณ์โดยการจำลองข้อมูลด้วยเทคนิคมอนติ คาร์โล (Monte Carlo Technique)
3. การพยากรณ์ราคาของข้าวเปลือกเจ้าของไทยด้วยตัวแบบ ARIMA

ผลการวิจัย

1. การพัฒนาวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีโปรเจกต์ซิงเคอร์เนลฮิลเบิร์ตสเปซ กรณีความคลาดเคลื่อนสุ่มมีอัตราสัมพัทธ์

กำหนดสมการ

$$y = F(x) + \mathcal{E}_i \dots\dots\dots(1)$$

เมื่อข้อมูลเป็น AR(1) ค่าความคลาดเคลื่อน (GAF SEBER and OJ. WILD, 1988, p. 18)

$$y = F(x) + p\mathcal{E}_{i-1} + a_i \dots\dots\dots(2)$$

$$F(x) = \frac{1}{h} \sum K(u) \dots\dots\dots(3)$$

การพัฒนาวิธีการประมาณค่าแบบปรับใหม่โดยใช้ฟังก์ชันเคอร์เนลของอีพานิชนิคอฟ (Epanechnikov Kernel Function) เมื่อ

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

เมื่อ $u = \frac{x - x_i}{h} \dots\dots\dots(5)$

แทนค่า $u = \frac{x - x_i}{h}$ ลงในสมการที่.....(4)

$$K(u) = \frac{3}{4} (1 - (\frac{x - x_i}{h})^2) I \dots\dots\dots(6)$$

$$= \frac{3}{4} (1 - (\frac{x - x_i}{h})^2) I \dots\dots\dots(7)$$

$$K(u) = \frac{3}{4} (\frac{h^2 - (x - x_i)^2}{h^2}) I \dots\dots\dots(8)$$

แทนค่า $K(u)$ ในสมการที่ 8 ลงในสมการที่ 3

$$F(x) = \frac{1}{nh} \sum \frac{3}{4} (\frac{h^2 - (x - x_i)^2}{h^2}) I \dots\dots\dots(9)$$

$$= \frac{3}{4nh} \sum (\frac{h^2 - (x - x_i)^2}{h^2}) I \dots\dots\dots(10)$$

$$= \frac{3}{4nh} \sum \frac{1}{h^2} (h^2 - (x - x_i)^2) I \dots\dots\dots(11)$$

$$F(x) = \frac{3}{4nh^3} \sum (h^2 - (x - x_i)^2) I \dots\dots\dots(12)$$

นำค่า $F(X)$ ในสมการที่ 12 แทนลงในสมการที่ 2 ได้สมการที่ปรับเปลี่ยน

$$\hat{y} = \frac{3}{4nh^3} \sum (h^2 - (x - x_i)^2) I + \rho \varepsilon_{i-1} + a_i \dots\dots\dots(13)$$

2. ผลการจำลองสถานการณ์ด้วยวิธีมอนติ คาร์โล (Monte Carlo Technique) ของการหาประสิทธิภาพวิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับเปลี่ยน กับวิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซของ Ferraty (2007) ภายใต้สถานการณ์ 108 สถานการณ์ 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 ($\rho = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$) พบว่า

วิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับเปลี่ยน ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ การแจกแจงแบบโคชี มีแบนวิดจ์แบบ Rules of Thumb (h_{rot}) ใน 54 สถานการณ์ ปรากฏว่ามี 52 สถานการณ์ที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Std. error) ต่ำกว่าวิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ ยกเว้นกรณี 5 ที่ 0.1, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 ปรากฏว่า วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับเปลี่ยน มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Std. error) ต่ำกว่าวิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซของ Ferraty (2007) และที่ $\rho = 0.2, 0.3$ วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซของ Ferraty (2007) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Std. error) ต่ำกว่าวิธีโรดิวิงส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับเปลี่ยน ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี (แสดงกราฟได้ดังรูปหน้าถัดไป)

3. ผลของพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยโดยใช้วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีโปรติวส์ซิงเคอร์เนลลิลเบอร์ทสเปซแบบปรับใหม่ จากข้อมูลราคาข้าวย้อนหลัง 5 ปีตั้งแต่เดือนมกราคม พ.ศ. 2552 ถึงเดือนธันวาคม พ.ศ. 2556 นำไปสร้างตัวแบบ ARIMA(p,d,q) มีสมบัติทั่วไปคือ ARIMA(1,0,0) ดังนี้
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

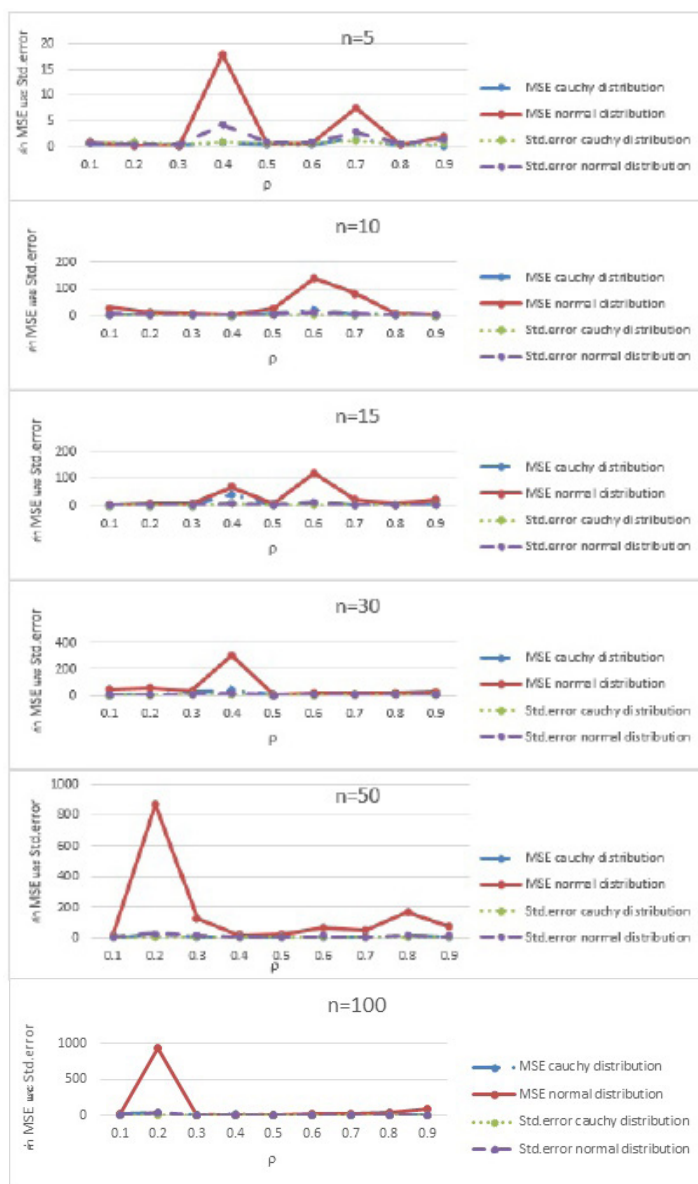
Nonseasonal อันดับที่ 1 (nonseasonal auto-regressive process of order 1)

เมื่อแทนค่า $\theta_0 = 9465.825$, $\phi_1 = 0.848$

$$\hat{y}_t = 9465.825 - 9465.825 \times 0.848 + 0.848 y_{t-1} + \varepsilon_t \quad \text{ได้สมการ}$$

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

ซึ่งผลการพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยที่ได้จากสมการตัวแบบ ARIMA(1,0,0) เมื่อนำไปเทียบกับราคาข้าวเปลือกเจ้าของไทยปี พ.ศ. 2557 ผลการตรวจสอบความเหมาะสมของตัวแบบ ARIMA (1,0,0) ซึ่งเห็นว่าสัมประสิทธิ์การกำหนดสามารถอธิบายความแปรผันได้ 60.30 % และเมื่อใช้สมการนี้ในการพยากรณ์ราคาข้าวมีคลาดเคลื่อนกำลังสองเฉลี่ย 764.718 บาทหรือโดยประมาณ 765 บาท



ภาพที่ 1 กราฟแสดงค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อน มาตรฐาน (Std. Error)

สรุปผลการวิจัย

วิธีวิธีโปรติวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟ ที่มีการแจกแจงแบบโคชี มีแบนวิดจ์แบบ Rules of Thumb (h_{ROT}) ใน 54 สถานการณ์ ปรากฏว่ามี 52 สถานการณ์ ที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Std. error) ต่ำกว่าวิธีวิธีโปรติวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ ของ Ferraty (2007) ที่ใช้ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ โดยเมื่อ 5 ที่ $\rho = 0.1, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9$ ปรากฏว่า วิธีการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีวิธีโปรติวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Std. error) ต่ำกว่าวิธีวิธีโปรติวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) และที่ $\rho = 0.2, 0.3$ วิธีวิธีโปรติวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซของ Ferraty (2007) ฟังก์ชันเคอร์เนลแบบเกาส์เซียน ที่มีการแจกแจงแบบปกติ มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Std. error) ต่ำกว่าวิธีวิธีโปรติวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซ แบบปรับใหม่ ที่ใช้ฟังก์ชันเคอร์เนลแบบอีพานิชนิคอฟที่มีการแจกแจงแบบโคชี

ในการประมาณค่าพารามิเตอร์การถดถอยโดยวิธีวิธีโปรติวส์ซิงเคอร์เนลฮิลเบอร์ทสเปซแบบปรับใหม่โดยใช้ฟังก์ชันเคอร์เนล แบบอีพานิชนิคอฟ (Epanechnikov) ซึ่งเป็นฟังก์ชันเคอร์เนลที่ประสิทธิภาพสัมพัทธ์สูงที่สุด (1.00) ที่ใช้แบนวิดจ์แบบ $h_{ROT} = 1.06\sigma(n^{\frac{1}{5}})$ การแจกแจงแบบโคชี จะมีค่าเฉลี่ย ความคลาดเคลื่อนกำลังสอง (MSE) และค่าความคลาดเคลื่อนมาตรฐาน (Std. Error) ลดลงเมื่อมีขนาดเพิ่มขึ้น ซึ่งสอดคล้องกับงานวิจัยของ Mugdadi และ Ahman (2004) โดยเมื่อมีขนาดกลุ่มตัวอย่างเพิ่มขึ้น ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (MSE) จะลดลงในทุกกรณี

การพยากรณ์ราคาข้าวเปลือกเจ้า 15% ของไทยด้วยตัวแบบ ARIMA ผลการตรวจสอบความเหมาะสมของตัวแบบ ARIMA (1,0,0) ชี้ให้เห็นว่าสัมประสิทธิ์การกำหนดสามารถอธิบายความแปรผันได้ 60.30 % และเมื่อใช้สมการในการพยากรณ์ราคาข้าวมีความคลาดเคลื่อนโดยเฉลี่ย 764.718 บาทหรือโดยประมาณ 765 บาท และจากสถิติ Ljung-Box ที่มีค่า p-value เท่ากับ .40 สามารถสรุปได้ว่า ตัวแบบ ARIMA (1,0,0) มีความเหมาะสมในการพยากรณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ข้อเสนอแนะในการนำไปใช้

1. วิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับใหม่ เป็นวิธีการทางนอนพาราเมตริก ภายใต้เงื่อนไขข้อมูลมีอัตราสัมพันธ์ที่ศึกษาสามารถใช้ในการประมาณค่าพารามิเตอร์การถดถอยกรณีข้อมูล มีอัตราสัมพันธ์ และข้อมูลมีขนาดตั้งแต่ 30 ขึ้นไป
2. การพยากรณ์ราคาข้าวเปลือกเจ้าของไทย มีปัจจัยด้านอื่นที่เป็นตัวแทรกแซงราคา จึงทำให้ผลการพยากรณ์ได้ค่าสัมประสิทธิ์การกำหนดสามารถอธิบายความแปรผันได้น้อย ผู้ศึกษาต่อไปควรศึกษาถึงปัจจัยที่ส่งผลกระทบต่อข้อกำหนด ราคาพืชผลทางการเกษตรในปัจจุบันและอนาคตด้วย

ข้อเสนอแนะในการศึกษาต่อไป

1. วิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับใหม่ เป็นวิธีการทางนอนพาราเมตริก สามารถใช้ประมาณค่าพารามิเตอร์การถดถอยกับพืชผลทางการเกษตรอื่นๆ ได้ในกรณีข้อมูลมีอัตราสัมพันธ์
2. การพัฒนาฟังก์ชันเคอร์เนล การแจกแจงของตัวแปรสุ่ม และแบบนิวตจ์ที่เหมาะสม ในการประมาณค่าพารามิเตอร์การถดถอยโดยใช้วิธีรีโพรดิวส์ซิงเคอร์เนลฮิลเบิร์ตสเปซ แบบปรับใหม่ แล้วจึงนำไปเข้าสมการพยากรณ์ จะเป็นประโยชน์กับการพยากรณ์ราคาของผลผลิตทางการเกษตร และโรงงานอุตสาหกรรมด้านการวางแผนการผลิตต่อไป

เอกสารอ้างอิง

- กิตติ ภัคดีวัฒนกุล และพินดา พานิชกุล. (2554). การวิเคราะห์เชิงปริมาณเพื่อ
การตัดสินใจ. กรุงเทพฯ: เคทีพี.
- กัลยา วานิชย์บัญชา. (2549). การวิเคราะห์สถิติ: สถิติสำหรับการบริหารและวิจัย
พิมพ์ครั้งที่ 13. กรุงเทพฯ: ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี
จุฬาลงกรณ์มหาวิทยาลัย.
- จิตราภรณ์ พันศิริ (2547). การพยากรณ์ราคาส่งออกข้าว โดยวิธีอาร์มา. วิทยานิพนธ์
วิทยาศาสตร์มหาบัณฑิต, สาขาวิชาสถิติประยุกต์ บัณฑิตวิทยาลัย มหาวิทยาลัย
เชียงใหม่.
- มุกดา แม่นมินทร์. (2549). อนุกรมเวลาและการพยากรณ์. กรุงเทพฯ: ประกายพริก.
- มนตรี พิริยะกุล. (2545). เทคนิคการวิเคราะห์สมการการถดถอย. กรุงเทพฯ: สำนักพิมพ์
มหาวิทยาลัยรามคำแหง.

- ปิยะนัทร สีสาคิลปะศาสตร์. (2549). การเลือกแบนวิดจสำหรับการประมาณความหนาแน่น
เคอร์เนลของฟังก์ชันตัวแปรสุ่ม. วิทยานิพนธ์ วิทยาศาสตร์มหาบัณฑิต
สาขาวิชาสถิติประยุกต์ บัณฑิตวิทยาลัย สถาบันเทคโนโลยีพระจอมเกล้า
พระนครเหนือ.
- ทรงศิริ แท้สมบัติ. (2549). การพยากรณ์เชิงปริมาณ. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัย
เกษตรศาสตร์.
- ศิริรัตน์ วงศ์ปกรณกุล. (2539). การสร้างตัวเลขสุ่มเทียม. กรุงเทพฯ: สำนักพิมพ์
จุฬาลงกรณ์มหาวิทยาลัย.
- 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**. Bio geosciences, 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.

- Bowerman & 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, et al. (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 verlag. **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 for 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 Hurdle. (1990). **Applied Nonparametric Regression**. Cambridge University Press-Business & Economics-333.
- Wolfgang Hurdle. (1990). **Smoothing Techniques with Implementation in S**. New York: Springer-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.