การศึกษาเปรียบเทียบการพยากรณ์ข้อมูลอนุกรมเวลาโดยใช้วิธีการทางพาราเมตริกและนอนพาราเมตริก

NONPARAMETRIC AND PARAMETRIC FORECASTING IN TIME SERIES DATA : A COMPARATIVE STUDY

*อัญชลี ทองกำเหนิด**

Unchalee Tonggumnead*

สาขาวิชาสถิติประยุกต์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี

Division of Applied Statistics, Faculty of Science and Technology,

Rajamangala University of Technology Thanyaburi.

***Corresponding author, E-mail:** unchalee_t@rmutt.ac.th*

บทคัดย่อ

การศึกษานี้เป็นการเปรียบเทียบประสิทธิภาพในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยวิธีการทางพาราเมตริกและนอนพาราเมตริก 3 วิธีด้วยกัน วิธีการทางพาราเมตริกคือวิธีบอกซ์–เจนกินส์และการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี ในขณะที่วิธีการทางนอนพาราเมตริก คือการประยุกต์ใช้ตัวประมาณ Nadaraya – Watson ในการพยากรณ์ค่าอนุกรมเวลา โดยศึกษาจากข้อมูลจริง จำนวน 2 ชุด ข้อมูลชุดที่ 1 คือราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2556 และข้อมูลชุดที่ 2 คือ ราคาทองคำในประเทศไทย รายเดือนตั้งแต่เดือน มกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2555 รูปแบบการศึกษาดำเนินการโดยแบ่งข้อมูลออกเป็น 2 ส่วน ส่วนแรกใช้สำหรับกำหนดรูปแบบสมการพยากรณ์ ข้อมูลส่วนที่ 2 จำนวน 12 ค่า ใช้สำหรับศึกษา ช่วงการพยากรณ์ล่วงหน้าที่ช่วงเวลา 3 เดือน 6 เดือน 9 เดือน และ 12 เดือน ทั้งนี้พิจารณาวิธีการพยากรณ์ที่เหมาะสมจากค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE%) มีค่าต่ำสุด ผลการศึกษาพบว่า วิธีที่มีประสิทธิภาพมากที่สุดในการพยากรณ์ข้อมูลอนุกรมเวลาล่วงหน้าที่ช่วงเวลา 3 เดือน 6 เดือน 9 เดือน และ 12 เดือน คือวิธีบอกซ์-เจนกินส์ ส่วนวิธีที่มีประสิทธิภาพรองลงมาคือ วิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson และวิธีที่มีประสิทธิภาพน้อยที่สุด คือ วิธีถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี

คำสำคัญ: การพยากรณ์ตัวประมาณ Nadaraya – Watson ตัวแปรดัมมีบอกซ์-เจนกินส์ฟัซซีอนุกรมเวลา

Abstract

This research aimed to compare the performance of forecasting techniques in time series data between parametric and nonparametric, There are tree forecasting methods that were used in this study : Box-jenkins, fuzzy regression with dummy variable, and nonparametric with Nadaraya – Watson estimator. There are two data sets were used in this study : the first group is the rubber price of central rubber market HATYAI monthly from January 2004

to December 2013, the second is the gold price monthly from January 2004 to December 2012. The time series in each group was devised into two groups. The first group for identification models.The second group for finding 4 periods of forecasting as 3, 6, 9 and 12 month. The suitable forecasting technique were chosen by considering the smallest value of MAPE (%). The best method is Box-jenkins, Nadaraya – Watson Estimator and fuzzy regression with dummy variable respectively.

Keywords: Forecasting, Nadaraya – Watson Estimator, Dummy Variable, Box-Jenkins, Fuzzy, Time Series

บทนำ

สิ่งที่สำคัญของการวิเคราะห์ข้อมูลอนุกรมเวลาคือ เป็นการพยากรณ์ข้อมูลในอนาคตโดยอาศัยข้อมูลในอดีตซึ่งมีเงื่อนไขว่ารูปแบบการเคลื่อนไหวของข้อมูลอนุกรมเวลาในอนาคตไม่แตกต่างไปจากข้อมูลในอดีต กล่าวคือให้ $Z_1, Z_2, ..., Z_n$ เป็นข้อมูลอนุกรมเวลา n ค่านวนจุดมุ่งหมายของการศึกษาคือต้องการพยากรณ์ค่าอนุกรมเวลาค่าที่ Z_{l+1} เมื่อ n เป็นจำนวนเต็มโดยที่ $l \geq 1$ หลักการทั่วไปของการของการสร้างสมการพยากรณ์จะพิจารณาจาก Autoregressive Process ซึ่งอยู่ในรูปแบบ

$$Z_t = g(Z_{t-1}, Z_{t-2}, ...) + \varepsilon_t \qquad (1)$$

เมื่อ ε_t คือ ความคลาดเคลื่อน $g(.)$ คือ ฟังก์ชันที่ต้องการประมาณค่า วิธีการทางพาราเมตริกซึ่งเป็นที่นิยมใช้คือ รูปแบบ ARMA (p,q) ของบอกซ์-เจนกินส์ (Box - Jenkins) ที่กำหนดรูปแบบ ARMA (p,q) ให้กับอนุกรมเวลาที่ค่าสังเกตมีสหสัมพันธ์กันแบบออโตและเป็นสเตชันนารี (Stationary) รูปแบบ ARMA (p,q) (Autoregressive and Moving Average) เป็นการรวมส่วนประกอบของ AR (p) ซึ่งเป็นรูปแบบที่แสดงว่าค่าสังเกต Z_t ขึ้นอยู่กับค่าสังเกตที่เวลาก่อนหน้าคือ $Z_{t-1}, ..., Z_{t-p}$ ในขณะที่รูปแบบ MA (q) เป็นรูปแบบที่แสดงให้เห็นว่าค่าสังเกต Z_t ขึ้นอยู่กับค่าความคลาดเคลื่อนที่เวลาก่อนหน้าคือ $\varepsilon_{t-1}, ..., \varepsilon_{t-p}$ การประมาณค่าพารามิเตอร์ในรูปแบบ ARMA (p,q) สามารถทำได้โดยวิธีการประมาณแบบง่าย วิธีกำลังสองน้อยที่สุด (Least Square) และวิธีภาวะน่าจะเป็นสูงสุด (Maximum Likelihood) ต่อมางานวิจัยเกี่ยวกับการพยากรณ์ข้อมูลอนุกรมเวลาแบบอิงพารามิเตอร์ได้ขยายไปสู่การประยุกต์ใช้ฟัซซีในการพยากรณ์ข้อมูลอนุกรมเวลา เนื่องจากมีความเชื่อว่าการพยากรณ์โดยใช้ข้อมูลในอดีตที่ขึ้นอยู่กับเวลานั้น รูปแบบของอนุกรมเวลาที่เก็บรวบรวมมาอาจมีความผิดพลาด คลาดเคลื่อนไม่ครบถ้วน หรืออาจมีความคลุมเครือ การประยุกต์ใช้ฟัซซีในการพยากรณ์ข้อมูลอนุกรมเวลาสามารถแก้ไขปัญหานี้ได้ [1] ได้ประยุกต์ใช้วิธีการถดถอยเชิงเส้นแบบฟัซซีในการพยากรณ์การใช้ไฟฟ้าระยะสั้นช่วงวันหยุด [2] ได้นำเสนอตัวแบบการพยากรณ์อนุกรมเวลาลำดับสูงแบบฟัซซีไทพ์ทูด้วยการถ่วงน้ำหนัก [3] ได้ศึกษาเกี่ยวกับการพยากรณ์ข้อมูลอนุกรมเวลาโดยใช้การถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีและการพยากรณ์โดยใช้ตัวแบบ FSARIMA ต่อมาวิธีในการพยากรณ์ข้อมูลอนุกรมเวลาได้ขยายไปสู่วิธีการทางนอนพาราเมตริกเนื่องจากมีความยืดหยุ่นต่อเงื่อนไขต่างๆ ในการประมาณค่าพารามิเตอร์ [4] ได้ศึกษาการพยากรณ์ข้อมูลอนุกรมเวลาโดยใช้ตัวประมาณประเภทเคอร์เนล (kernel) ได้แก่ Nadaraya – Watson และตัวประมาณ local linear [5] ได้ศึกษา

เกี่ยวกับการวิเคราะห์การถดถอยแบบไม่อิงพารามิเตอร์สำหรับอนุกรมเวลาที่อยู่ในรูปแบบ Locally Stationary ในขณะเดียวกับ [6] ได้พยากรณ์ความต้องการใช้พลังงานไฟฟ้าโดยใช้เทคนิคของการถดถอยแบบไม่อิงพารามิเตอร์ (Nonparametric Regression) และได้เปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์กับวิธีของบอกซ์-เจนกินส์

สำหรับการศึกษานี้ต้องการเปรียบเทียบประสิทธิภาพในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยวิธีทางพาราเมตริกและนอนพาราเมตริก วิธีการทางพาราเมตริกคือวิธีบอกซ์-เจนกินส์และการถดถอยแบบฟัซซีที่ใช้ตัวแปร

ดัมมีในขณะที่วิธีการทางนอนพาราเมตริกที่ศึกษาในที่นี้คือประยุกต์ใช้ตัวประมาณ Nadaraya – Watson (NW) ในการพยากรณ์ข้อมูลอนุกรมเวลาเนื่องจากตัวประมาณ Nadaraya – Watson เป็นตัวประมาณที่มีความยืดหยุ่นสูงกว่าตัวประมาณด้วยวิธีการทางนอนพาราเมตริกวิธีอื่น และเป็นตัวประมาณที่มีความถูกต้องแม่นยำสูง โดยตัวประมาณ Nadaraya – Watson ขึ้นอยู่กับค่าพารามิเตอร์ที่เรียกว่า พารามิเตอร์ปรับให้เรียบหรือแบนวิดธ์ (Bandwidth) ซึ่งแทนด้วยสัญลักษณ์ h โดยที่แบนวิดธ์ของตัวประมาณ Nadaraya – Watson สามารถกำหนดเป็นค่าคงที่หรือตัวแปรก็ได้ โดยค่าแบนวิดธ์ที่เหมาะสมจะทำให้ความคลาดเคลื่อนกำลังสองรวมเฉลี่ย (Mean Integrated Square Error: MISE) มีค่าน้อยที่สุด [7] ในการศึกษานี้ได้ศึกษาจากข้อมูลจริง จำนวน 2 ชุด ข้อมูลชุดที่ 1 คือราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือนตั้งแต่เดือนมกราคม พ.ศ.2547 ถึงเดือนธันวาคม พ.ศ. 2556 จำนวน 120 ค่า และข้อมูลชุดที่ 2 คือ ราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2555 จำนวน 108 ค่า

วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบประสิทธิภาพในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยวิธีทางพาราเมตริกคือวิธีบอกซ์– เจนกินส์การถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี และวิธีการทางนอนพาราเมตริกคือประยุกต์ใช้ตัวประมาณ Nadaraya – Watson

วิธีดำเนินการวิจัย

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาวิธีการพยากรณ์ข้อมูลอนุกรมเวลา 3 วิธี ได้แก่ 1) วิธีบอกซ์– เจนกินส์ 2) การถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี และ 3) การประยุกต์ใช้ตัวประมาณ Nadaraya – Watson โดยข้อมูลที่ศึกษาเป็นข้อมูลอนุกรมเวลารายเดือน จำนวน 2 ชุด

ข้อมูลชุดที่ 1 ราคายางแผ่นดิบ รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2556 จำนวน 120 ค่า เป็นข้อมูลที่ได้จากตลาดกลางยางพารา อ.หาดใหญ่ จ.สงขลา [8] แบ่งข้อมูลออกเป็น 2 ส่วน ส่วนที่ 1 จำนวน 108 ค่า สำหรับกำหนดรูปแบบการพยากรณ์ ข้อมูลส่วนที่ 2 จำนวน 12 ค่า ใช้สำหรับพิจารณารูปแบบที่เหมาะสมของอนุกรมเวลา โดยนำมาใช้คำนวณค่าการพยากรณ์ล่วงหน้า 4 ช่วง คือ 3 เดือน 6 เดือน 9 เดือน และ 12 เดือน

ข้อมูลชุดที่ 2 ราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2555 จำนวน 108 ค่า เป็นข้อมูลที่ได้จากสมาคมค้าทอง [9] แบ่งข้อมูลออกเป็น 2 ส่วน ส่วนที่ 1 จำนวน 96 ค่า สำหรับกำหนดรูปแบบการพยากรณ์ ข้อมูลส่วนที่ 2 จำนวน 12 ค่า ใช้สำหรับพิจารณารูปแบบที่เหมาะสมของอนุกรมเวลา โดยนำมาใช้คำนวณค่าการพยากรณ์ล่วงหน้า 4 ช่วง คือ 3 เดือน 6 เดือน 9 เดือน และ 12 เดือน

เกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพของการพยากรณ์ ได้แก่ เปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE%)

ซึ่งเป็นค่าที่ใช้วัดความถูกต้องของการพยากรณ์ที่พิจารณาจากขนาดของผลต่างระหว่างค่าของข้อมูลจริง ณ เวลา t กับค่าพยากรณ์ของข้อมูล ณ เวลา t โดยเป็ค่าที่ไม่มีหน่วย ซึ่งมีความเหมาะสมสำหรับใช้เปรียบเทียบอนุกรมเวลาหลายๆ ชุดที่มีหน่วยในการวัดต่างกันเมื่อใช้วิธีการพยากรณ์เดียวกัน โดยวิธีการพยากรณ์ที่เหมาะสมคือวิธีที่ให้ค่าเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์เฉลี่ยต่ำ

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100 \qquad (2)$$

เมื่อ Y_t แทน ค่าของข้อมูลจริง ณ เวลา t
$\hat{Y}$ แทน ค่าของข้อมูลที่เกิดขึ้นจากการพยากรณ์ ณ เวลา t
n แทน จำนวนข้อมูลทั้งหมด

วิธีการพยากรณ์ข้อมูลอนุกรมเวลาทั้ง 3 วิธี มีรายละเอียดดังนี้

1. วิธีบอกซ์– เจนกินส์

วิธีบอกซ์– เจนกินส์ เป็นวิธีกำหนดรูปแบบให้กับอนุกรมเวลาโดยพิจารณาว่าค่าสังเกต Y_t ขึ้นอยู่กับค่าสังเกตก่อนหน้าที่เวลา $t-1,...,t-p$ คือ $Y_{t-1},...,Y_{t-p}$ และค่าสังเกตของ Y_t ขึ้นอยู่กับค่าความคลาดเคลื่อนก่อนหน้าที่เวลา $t-1,...,t-p$ คือ $\varepsilon_{t-1},...,\varepsilon_{t-p}$ หรือไม่ โดยการกำหนดตัวแบบของการพยากรณ์จะวัดจากค่าสหสัมพันธ์ระหว่างค่าสังเกตที่ช่วงเวลาห่าง k ช่วง (Y_t กับ Y_{t-k}) โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์แบบออโต (ACF: r_k) และค่าสัมประสิทธิ์สหสัมพันธ์บางส่วนแบบออโต (PACF : r_{kk}) ที่ช่วงเวลาห่าง k ขั้นตอนของวิธีบอกซ์–เจนกินส์ มีดังนี้ [10]

1) ตรวจสอบความเป็นสเตชันนารีของอนุกรมเวลา หากพบว่าไม่เป็นสเตชันนารีต้องดำเนินการแปลงข้อมูลด้วย การหาผลต่าง ผลต่างฤดูกาล หรือการแปลงค่าสังเกตในอนุกรมเวลาด้วยฟังก์ชันทางคณิตศาสตร์หากอนุกรมเวลา Y_t มีการเคลื่อนไหวเนื่องจากแนวโน้มให้หาผลต่างจนได้อนุกรมเวลาใหม่ Z_t ที่เป็นสเตชันนารี โดย $Z_t = \nabla^d Y_t$ เมื่อ d แทนจำนวนครั้งที่มีการหาผลต่าง หากอนุกรมเวลาเวลา Y_t มีการเคลื่อนไหวเนื่องจากฤดูกาลให้หาผลต่างจนได้อนุกรมเวลาใหม่ Z_t ที่เป็นสเตชันนารี โดย $Z_t = \nabla_L^d Y_t$ เมื่อ d แทนจำนวนครั้งที่มีการหาผลต่าง หากอนุกรมเวลาเวลา Y_t มีการเคลื่อนไหวเนื่องจากแนวโน้มฤดูกาลให้หาผลต่างจนได้อนุกรมเวลาใหม่ Z_t ที่เป็นสเตชันนารี โดย $Z_t = \nabla^d \nabla_L^d Y_t$, หากอนุกรมเวลา Y_t มีความแปรปรวนไม่คงที่ให้แปลงอนุกรมเวลา Y_t เป็นอนุกรมเวลาใหม่ Z_t ด้วยฟังก์ชันต่าง ๆ เช่น $Z_t = \ln Y_t$ $Z_t = Y_t^{1/a}$ เมื่อ a เป็นค่าคงที่

2) สร้างรูปแบบอนุกรมเวลาที่เป็นสเตชันนารี Z_t ด้วยรูปแบบ $ARMA(p,q)\times SARMA(P,Q)_L$ และจะได้รูปแบบของอนุกรมเวลา Y_t คือ $ARIMA(p,d,q)\times SARIMA(P,D,Q)_L$ การเลือกรูปแบบที่เหมาะสมพิจารณาจากค่า r_k และ r_{kk} ของค่าสังเกตที่ช่วงเวลาห่าง k

3) ประมาณค่าพารามิเตอร์ของตัวแบบพยากรณ์โดยใช้วิธีกำลังสองน้อยที่สุด (Least Square Method)

4) ตรวจสอบความเหมาะสมของตัวแบบโดยพิจารณาจากการทดสอบค่าพารามิเตอร์ในตัวแบบมีค่าเท่ากับ 0 หรือไม่ ด้วยการทดสอบแบบ t หรือ t บางส่วน และการพิจารณาความเหมาะสมของตัวแบบจากค่า $r_k(e_t)$ ของความความคลาดเคลื่อนและจากการทดสอบของ Box-Ljung

5) นำตัวแบบที่ได้ในขั้นตอนที่ 4) ไปใช้ในการพยากรณ์

2. วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี

ตัวแบบการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี มีตัวแบบของสมการพยากรณ์ดังต่อไปนี้

$$\tilde{Y}_t = \tilde{\beta}_0 + \tilde{\beta}_1 t + \sum_{i=2}^{12} \tilde{\beta}_i X_{it} \qquad (3)$$

เมื่อ $\tilde{Y}_t$ แทน ค่าพยากรณ์ของตัวแปรตาม

t แทน ตัวแปรอิสระในการศึกษานี้คือเวลา

$\tilde{\beta}_i$ แทน ฟัซซีพารามิเตอร์ประกอบด้วยค่ากลางและการกระจาย (α_i, c_i)

X_{it} แทน ตัวแปรดัมมีโดย X_{it} = 1 เมื่อค่าสังเกตอยู่ในฤดูกาลที่ (i – 1)

X_{it} = 0 เมื่อค่าสังเกตไม่เป็นค่าในฤดูกาลที่ (i – 1)

สำหรับการศึกษานี้กำหนดให้ฟัซซีพารามิเตอร์ $\tilde{\beta}_i$, $i = 1,2,3,...,12$ มีฟังก์ชันความเป็นสมาชิกแบบสามเหลี่ยมสมมาตรคือ

$$\mu_{\hat{\beta}_i}(\beta_i) = \begin{cases} 1 - \dfrac{|\alpha_i - \beta_i|}{c_i} & ; \ \alpha_i - c_i \le \beta_i \le \alpha_i + c_i \\ 0 & ; \ \text{otherwise} \end{cases} \qquad (4)$$

เมื่อ $\mu_{\hat{\beta}_i}(\beta_i)$ แทน ฟังก์ชันความเป็นสมาชิกของแบบฟัซซีของพารามิเตอร์ β_i

α_i แทน ค่ากลางของตัวเลขฟัซซี

c_i แทน ค่าการกระจายของตัวเลขฟัซซีรอบค่ากลาง

วัตถุประสงค์ของการประมาณค่าพารามิเตอร์โดยใช้การถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีคือต้องการให้สมการพยากรณ์ที่ได้มีความคลุมเครือน้อยที่คือ minimize สมการเป้าหมาย

สมการเป้าหมายคือ $S = \sum_{t=1}^{k} \left(c_0 + c_1 t + \sum_{i=2}^{12} c_i X_{it} \right)$ (5)

โดยมีสมการเงื่อนไขคือ $1 - \dfrac{\left| Y_t - [\alpha_0 + \alpha_1 t + \sum_{i=2}^{12} \alpha_i X_{it}] \right|}{c_0 + c_1 t + \sum_{i=2}^{12} c_i X_{it}} \ge h$ (6)

เมื่อ h แทน ระดับความเป็นสมาชิกน้อยที่สุดของ Y_t ที่อยู่ในช่วงของการพยากรณ์ด้วยการถดถอยแบบฟัซซีซึ่ง $h \in [0,1]$ โดยค่า h จะมีผลต่อการกระจายของค่าพารามิเตอร์เมื่อ h มีค่ามากค่าพารามิเตอร์จะมีการกระจายมากขึ้น ในการศึกษานี้กำหนดให้ h = 0.5 เมื่อประมาณค่าพารามิเตอร์โดยเทคนิคกำหนดการเชิงเส้นจะได้ $\tilde{\beta}_i = (\hat{\alpha}_i, \hat{c}_i)$ ดังนั้นสมการพยากรณ์คือ $\hat{Y}_t = (\hat{\alpha}_0, \hat{c}_0) + (\hat{\alpha}_1, \hat{c}_1)t + \sum_{i=2}^{12} (\hat{\alpha}_i, \hat{c}_i) X_{it}$ [3]

3. วิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson

พิจารณารูปแบบอนุกรมเวลาที่เป็นสเตชันนารี Z_t โดยที่ $1 \le t \le n$ กำหนดให้ตัวแบบอยู่ในรูป

$$Z_t = g(Z_{t-i_1}, Z_{t-i_2}, ..., Z_{t-i_p}) + \varepsilon_t \quad (7)$$

ε_t คือ ความคลาดเคลื่อนซึ่งเป็นอิสระกันและมีค่าเฉลี่ยเท่ากับ 0 เนื่องจากต้องมีการพยากรณ์ข้อมูลตัวที่ Z_{t+1} จึงพิจารณาข้อมูล $(\underline{X}_j, Y_j)$ ซึ่งมีค่าอยู่ใน $R^p \times R = R^{p+1}$ เมื่อ $\underline{X}_j = (Z_{j+i_p-i_1}, Z_{j+i_p-i_2}, ..., Z_i)$ และ $Y_j = Z_{j+i_p}$ ดังนั้นจากสมการที่ (7) จะได้ว่า

$$\hat{g}(\underline{X}_{n-i_p+1}) = \hat{g}(Z_{n+i_p-i_1}, Z_{n+i_p-i_2}, ..., Z_n) \quad (8)$$

ดังนั้นในการพยากรณ์ค่า Z_{n+1} ก็คือการหาตัวประมาณ $\hat{g}(\underline{u})$ เมื่อ $\underline{u} = (\underline{X}_{n-i_p+1})$ ซึ่งโดยปกติตัวประมาณด้วยวิธีการทางนอนพาราเมตริกจะอยู่ในรูปของ

$$\hat{g}(\underline{u}) = \sum_{j=1}^{n} W(\underline{u}, \underline{X}_j) Y_j \quad (9)$$

เมื่อ $W(\underline{u}, \underline{X}_j)$ แทน น้ำหนักที่ให้กับข้อมูล ซึ่งจะมีค่าสูงเมื่อ $\underline{u}$ มีค่าอยู่ใกล้ $\underline{X}_j$ และจะมีค่าน้อยเมื่อ $\underline{u}$ มีค่าอยู่ห่างจาก $\underline{X}_j$

n แทน จำนวนข้อมูลทั้งหมด

Y_j แทน ค่าสังเกตค่าที่ j [4]

สำหรับตัวประมาณ Nadaraya – Watson จะมีน้ำหนักแทนด้วย

$$W_{H,NW}(\underline{u}, \underline{X}_j) = \frac{K_H(\underline{X}_j - \underline{u})}{\sum_{i=1}^{n} K_H(\underline{X}_i - \underline{u})} \quad (10)$$

เมื่อ K แทน p – variate เคอร์เนลฟังก์ชัน

H แทน แบนวิดธ์เมตริกซ์ขนาด $p \times p$

จากสมการที่ (10) แบนวิดธ์เมตริกมีสมาชิกในแนวเส้นทแยงมุมหลักอยู่ในรูปของ $H = diag\{h, ..., h\}$ เมื่อพิจารณาในเชิง 1 – dimension สามารถเขียนเคอร์เนลฟังก์ชันให้อยู่ในรูป $K(\overline{v}) = \prod_{i=1}^{p} K(v_t)$ ดังนั้นจะได้ว่าจากสมการที่ (10) สามารถเขียนให้อยู่ในรูปทั่วไปได้ดังนี้ [11]

$$W_{H,NW}(\underline{u}, \underline{X}_j) = \frac{\Pi_{t=1}^{p} K_h(X_{jt} - u_t)}{\sum_{i=1}^{n} \Pi_{t=1}^{p} K_h(X_{it} - u_t)} \quad (11)$$

ดังนั้นจะได้ว่าตัวประมาณ Nadaraya – Watson คือ

$$\hat{g}_{h,NW}(\underline{u}) = \frac{\sum_{j=1}^{n} (\Pi_{t=1}^{p} K_h(X_{jt} - u_t)) Y_j}{\sum_{i=1}^{n} \Pi_{t=1}^{p} K_h(X_{it} - u_t)} \quad (12)$$

เนื่องจากการหาตัวประมาณด้วยวิธีการทางนอนพาราเมตริก Nadaraya – Watson มีความเกี่ยวข้องกับเวกเตอร์ $(Z_{t-i_1}, Z_{t-i_1}, ..., Z_{t-i_p})$ ดังนั้นจึงต้องมีการเลือกค่า autoregressor (p) ซึ่งก็คือการเลือกจำนวนของระยะห่าง (lag) $i_1, i_2, ..., i_p$ ในการศึกษานี้ได้ใช้หลักการที่ถูกพัฒนามาจาก martin [12] โดยใช้หลักเกณฑ์ที่เรียกว่า final prediction error (FPE) โดยให้ Ω_s^1 เป็นสับเซตของ $\{Z_{t-s}\}$ โดยจะหาเซตของ Ω_s^1 ที่ minimizes สมการที่ (13) คือ

$$FPE(\Omega_s^p) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{g}_{h,NW}(\underline{u}))^2 \frac{1+(nh^p)^{-1}J^pB_p}{1-(nh^p)^{-1}(2K^p(0)-J^P)}B_p \quad (13)$$

เมื่อ J เป็นการหาปริพันธ์ของเคอร์เนลฟังก์ชันโดยที่ $J = \int K^2(v)dv$, ซึ่งในขั้นตอนที่ 1 จะกำหนดให้ p เท่ากับ 1 หลังจากนั้นจะเพิ่มค่า p ขึ้นเรื่อย ๆ และจะหยุดเมื่อไม่สามารถลดค่า final prediction error (FPE) ได้อีกแล้ว

$$B_p = \sum_{i=1}^{n}\sum_{j=1}^{n}(\Pi_{t=1}^{p}K_h(X_{jt} - X_{it})) \quad (14)$$

ในขณะเดียวกันในการศึกษานี้คำนวณหาแบนวิดธ์ด้วยวิธี cross-validation [4] คือ

$$CV(h) = \frac{1}{n}\sum_{j=1}^{n}(Y_i - \hat{g}_{h,\ \mathrm{NW}}(\underline{X}_j))^2 w(\underline{X}_j) \quad (15)$$

เมื่อ $\hat{g}_{h,\ \mathrm{NW}}(\underline{X}_j)$ คือตัวประมาณแบบนอนพาราเมตริกของ $g_{h,\ \mathrm{NW}}(\underline{X}_j)$ h คือ bandwidth และฟังก์ชันถ่วงน้ำหนัก ($w(\underline{X}_j)$)อยู่ในรูปของ

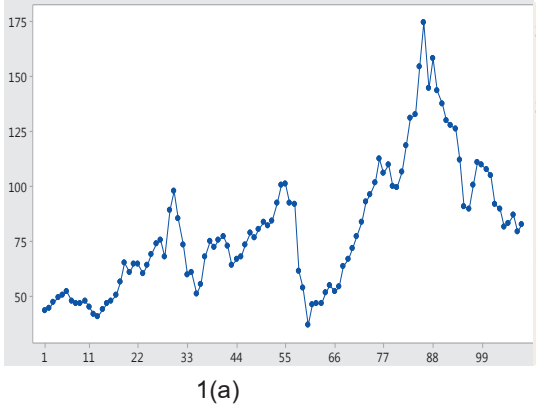
$$w(\underline{X}_j) = \Pi_{t=1}^{p}\Phi(\frac{X_{jt} - X_t}{0.2\sigma_Z}) \quad (16)$$

เมื่อ $\Phi(u)$ แทน ฟังก์ชันความหนาแน่นน่าจะเป็นที่มีการแจกแจงแบบปกติ

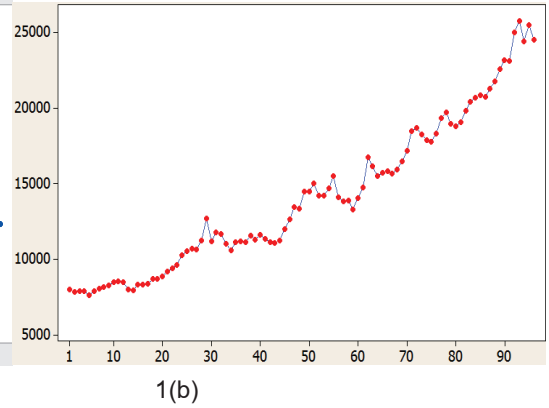
σ_Z แทน ส่วนเบี่ยงเบนมาตรฐานของอนุกรมเวลา

ผลการวิจัย

ผลการวิเคราะห์ลักษณะการเคลื่อนไหวของข้อมูลอนุกรมเวลาทั้ง 2 ชุด พบว่ามีลักษณะดังภาพที่ 1(a) และภาพที่ 1(b)



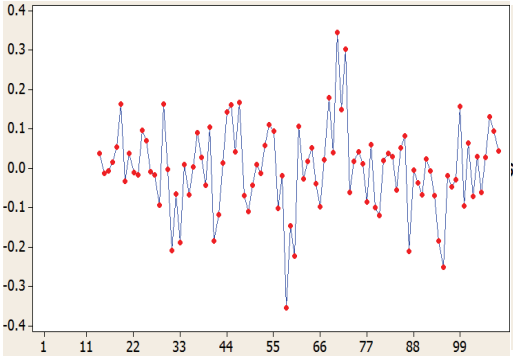
1(a)



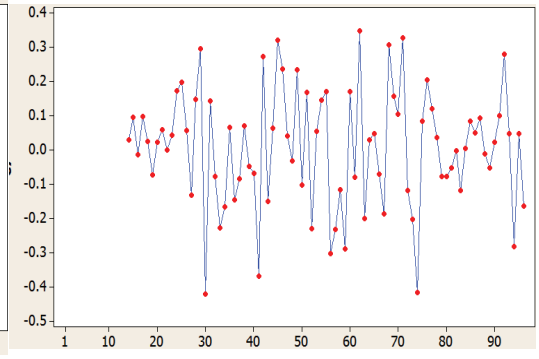
1(b)

ภาพที่ 1 (a) การเคลื่อนไหวของอนุกรมเวลาราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2556 **ภาพที่ 1 (b)** การเคลื่อนไหวของอนุกรมเวลาราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือนตั้งแต่เดือน มกราคม พ.ศ. 2547 ถึงเดือน ธันวาคม พ.ศ. 2555

จากภาพการเคลื่อนไหวของอนุกรมเวลาทั้ง 2 ชุด พบว่าไม่เป็นสเตชันนารีเนื่องจากมีส่วนประกอบของแนวโน้ม ฤดูกาล และความแปรปรวนไม่คงที่ ดังนั้นจึงสร้างอนุกรมเวลาใหม่ $Z_t = Y_t^{1/4}$ และกำจัดอิทธิพลของแนวโน้มและฤดูกาลโดยการสร้างอนุกรมเวลาใหม่ จะได้อนุกรมเวลาที่เป็นสเตชันนารีที่มีการเคลื่อนไหวดังภาพที่ 2(a) และภาพที่ 2(b)



2(a)

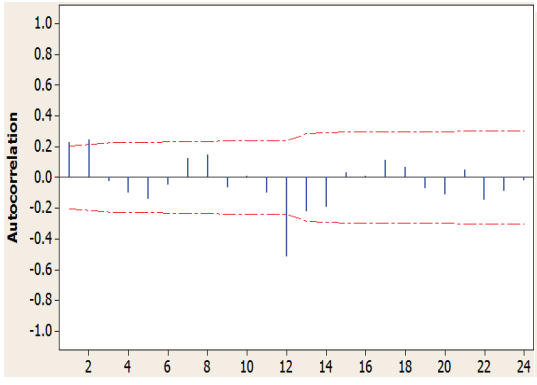


2(b)

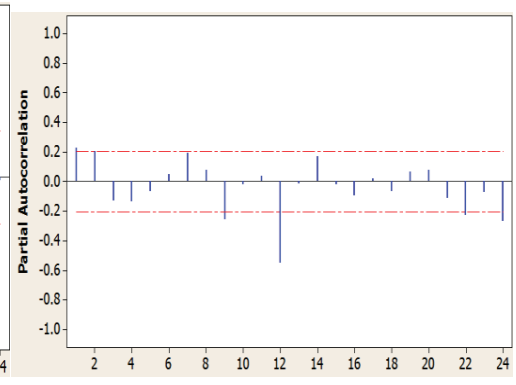
ภาพที่ 2(a) : การเคลื่อนไหวของอนุกรมเวลาราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือนตั้งแต่เดือนมกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2556 ที่เป็นสเตชันนารี **ภาพที่ 2(b)** การเคลื่อนไหวของอนุกรมเวลาราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือนตั้งแต่เดือน มกราคม พ.ศ. 2547 ถึงเดือนธันวาคม พ.ศ. 2555 ที่เป็นสเตชันนารี

1. ผลการวิเคราะห์ข้อมูลอนุกรมเวลาด้วยวิธีบอกซ์– เจนกินส์

1.1 ข้อมูลชุดที่ 1 : ราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือน จากอนุกรมเวลาที่เป็นสเตชันนารีของราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือน กำหนดรูปแบบสมการพยากรณ์โดยพิจารณาจากกราฟ AFC และ PACF ได้ดังภาพที่ 3(a) และภาพที่ 3(b)



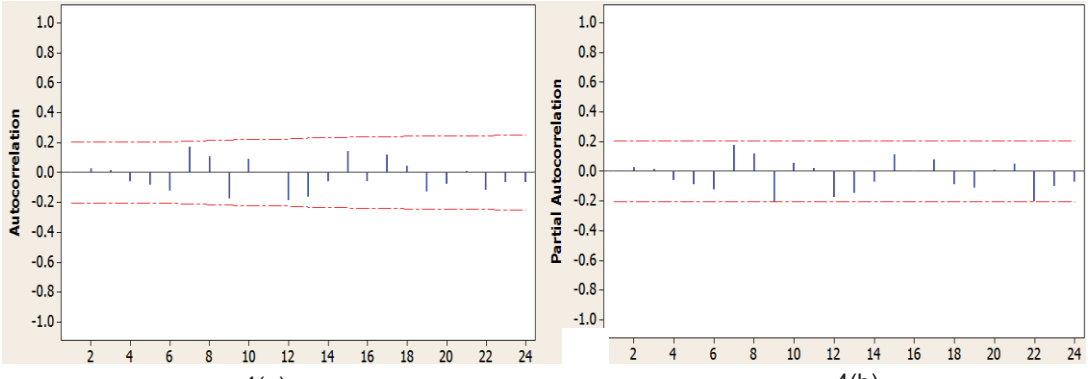
3(a)



3(b)

ภาพที่ 3(a) และ**ภาพที่ 3(b)** กราฟ ACF และ PACF ของราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่

จากภาพที่ 3(a) และ 3(b) พิจารณารูปแบบของสมการพยากรณ์จากอนุกรมเวลา W_t ที่เป็นสเตชันนารี จากกราฟ ACF และ PACF พบว่ารูปแบบที่เหมาะสมคือรูปแบบ $ARMA(0,2)\times SARMA(0,1)_{12}$ โดยมีตัวแบบสมการพยากรณ์คือ $W_t=\theta_0+\varepsilon_t+\theta_1\varepsilon_{t-1}+\theta_2\varepsilon_{t-2}+\theta_{12}\varepsilon_{t-12}$ และสามารถประมาณค่าพารามิเตอร์ได้ดังสมการต่อไปนี้ $W_t=-0.0093+\varepsilon_t-0.1143\varepsilon_{t-1}-0.2776\varepsilon_{t-2}+0.8464\varepsilon_{t-12}$ ซึ่งจากการตรวจสอบความเหมาะสมของรูปแบบ จากการประมาณ θ_1 ได้ค่า $t=-1.14$ (p=0.049) จากการประมาณ θ_2 ได้ค่า $t=-2.77$ (p=0.007) จากการประมาณค่า θ_{12} ได้ค่า $t=8.77$ (p=0.000) แสดงว่ารูปแบบของสมการพยากรณ์มีความเหมาะสมอย่างมีนัยสำคัญที่ระดับ 0.05 เนื่องจากรูปแบบสมการพยากรณ์ W_t คือ W_t~ $ARMA(0,2)\times SARMA(0,1)_{12}$ ในทำนองเดียวกันสามารถเขียนสมการพยากรณ์ในรูปของอนุกรมเวลา Y_t คือ Y_t~$ARIMA(0,1,2)\times SARIMA(0,1,1)_{12}$ โดยหาค่าสมการพยากรณ์ Y_t จากความสัมพันธ์ของสมการ $Z_t=W_t+Z_{t-1}+Z_{t-12}-Z_{t-13}$ เนื่องจาก $W_t=\nabla^1\nabla^1_{12}=Z_t-Z_{t-1}-Z_{t-12}+Z_{t-13}$ เมื่อ $Z_t=Y_t^{1/4}$ และจากการพิจารณากราฟ ACF และ PACF ของความคลาดเคลื่อน ดังภาพที่ 4(a) และภาพที่ 4(b) พบว่าที่ทุก ๆ ค่า k ตกอยู่ในขอบเขตช่วงความเชื่อมั่น 95 เปอร์เซ็นต์ แสดงว่าความคลาดเคลื่อนไม่มีสหสัมพันธ์ต่อกัน ดังนั้นรูปแบบสมการพยากรณ์ที่กำหนดขึ้นจึงมีความเหมาะสม

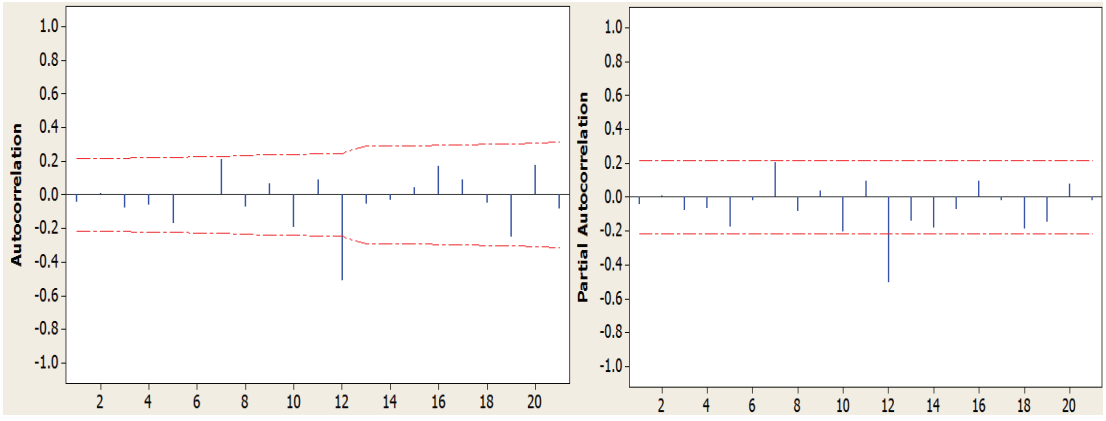


4(a) 4(b)

ภาพที่ 4(a) และ**ภาพที่ 4(b)** กราฟ ACF และ PACF ของความคลาดเคลื่อนของราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่

1.2 ข้อมูลชุดที่ 2 : ราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือน

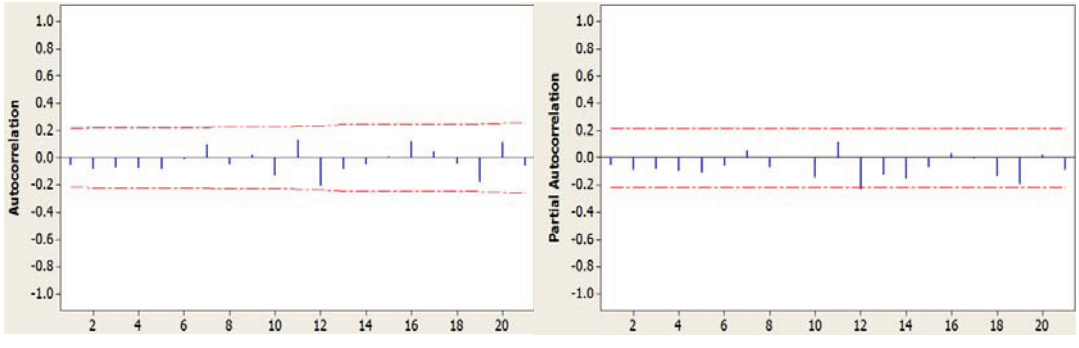
จากอนุกรมเวลาที่เป็นสเตชันนารีของราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือน กำหนดรูปแบบสมการพยากรณ์โดยพิจารณาจากกราฟ ACF และ PACF ได้ดังภาพที่ 5(a) และภาพที่ 5(b)



5(a) 5(b)

ภาพที่ 5(a) และ**ภาพที่ 5(b)** กราฟ ACF และ PACF ของราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือน

จากภาพที่ 5(a) และ 5(b) พิจารณารูปแบบของสมการพยากรณ์จากอนุกรมเวลา ที่เป็นสเตชันนารีจากกราฟ ACF และ PACF พบว่ารูปแบบที่เหมาะสมคือรูปแบบ $ARMA(0,0)\times SARMA(0,1)_{12}$ โดยมีตัวแบบสมการพยากรณ์อยู่ในรูป $W_t = \theta_0 + \varepsilon_t + \theta_{12}\varepsilon_{t-12}$ และสามารถประมาณค่าพารามิเตอร์ได้ดังสมการต่อไปนี้ $W_t = 0.002634 + \varepsilon_t + 0.8849\varepsilon_{t-12}$ ซึ่งจากการตรวจสอบความเหมาะสมของรูปแบบ จากการประมาณ θ_{12} ได้ค่า $t = 9.06$ (p=0.000) แสดงว่ารูปแบบของสมการพยากรณ์มีความเหมาะสมอย่างมีนัยสำคัญที่ระดับ 0.05 เนื่องจากรูปแบบสมการพยากรณ์ W_t คือ W_t~$ARMA(0,0)\times SARMA(0,1)_{12}$ ในทำนองเดียวกันสามารถเขียนสมการพยากรณ์ในรูปของอนุกรมเวลา Y_t คือ Y_t~$ARIMA(0,1,0)\times SARIMA(0,1)_{12}$ โดยหาค่าสมการพยากรณ์ Y_t จากความสัมพันธ์ของสมการ $Z_t = W_t + Z_{t-1} + Z_{t-12} - Z_{t-13}$ เนื่องจาก $W_t = \nabla^1\nabla^1_{12} = Z_t - Z_{t-1} - Z_{t-12} + Z_{t-13}$ เมื่อ $Z_t = Y_t^{1/4}$ และจากการพิจารณากราฟ ACF และ PACF ของความคลาดเคลื่อน ดังภาพที่ 6(a) และภาพที่ 6(b) พบว่าที่ทุก ๆ ค่า k ตกอยู่ในขอบเขตช่วงความเชื่อมั่น 95 เปอร์เซ็นต์ แสดงว่าความคลาดเคลื่อนไม่มีสหสัมพันธ์ต่อกัน ดังนั้นรูปแบบสมการพยากรณ์ที่กำหนดขึ้นจึงมีความเหมาะสม



6(a) 6(b)

ภาพที่ 6(a) และ**ภาพที่ 6(b)** กราฟ ACF และ PACF ของราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือน

2. ผลการวิเคราะห์ข้อมูลอนุกรมเวลาวิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี

ผลการวิเคราะห์ข้อมูลอนุกรมเวลาด้วยวิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีโดยกำหนดให้เดือนแทนตัวแปรดัมมีวิเคราะห์ข้อมูลโดยใช้โปรแกรม SAS 9.1 ได้รูปแบบสมการพยากรณ์ของข้อมูลอนุกรมเวลาดังต่อไปนี้ ข้อมูลชุดที่ 1 ราคายางพารารายเดือนเมื่อพิจารณาที่ค่ากลางของตัวเลขฟัซซี (α_i) มีรูปแบบของสมการพยากรณ์คือ

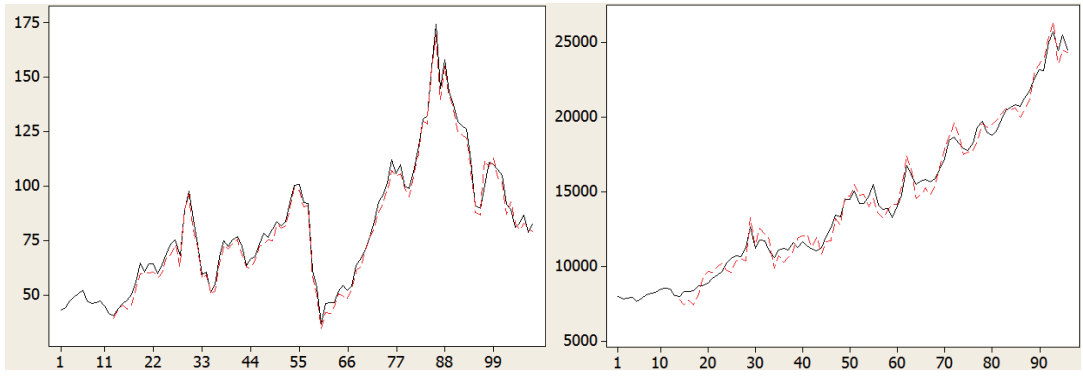
$$\hat{Y}_t = 32.13 + 0.89t + 1.12X_2 + 3.19X_3 + 1.84X_4 + 1.44X_5 + 1.57X_6 - 4.70X_7 + 1.12X_8 + 1.09X_9 + 1.95X_{10} - 2.43X_{11} - 3.39X_{12}$$

ข้อมูลชุดที่ 2 ราคาทองคำรายเดือนเมื่อพิจารณาที่ค่ากลางของตัวเลขฟัซซี (α_i) มีรูปแบบของสมการพยากรณ์คือ

$$\hat{Y}_t = 5,074 + 181.73t + 1123.34X_2 + 1135.24X_3 + 1,653.34X_4 + 1046.54X_5 + 1,146.68X_6 + 904.53X_7 + 1011.11X_8 + 886.76X_9 + 695.34X_{10} + 678.3X_{11} + 589.65X_{12}$$

3. ผลการวิเคราะห์ข้อมูลอนุกรมเวลาด้วย วิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson

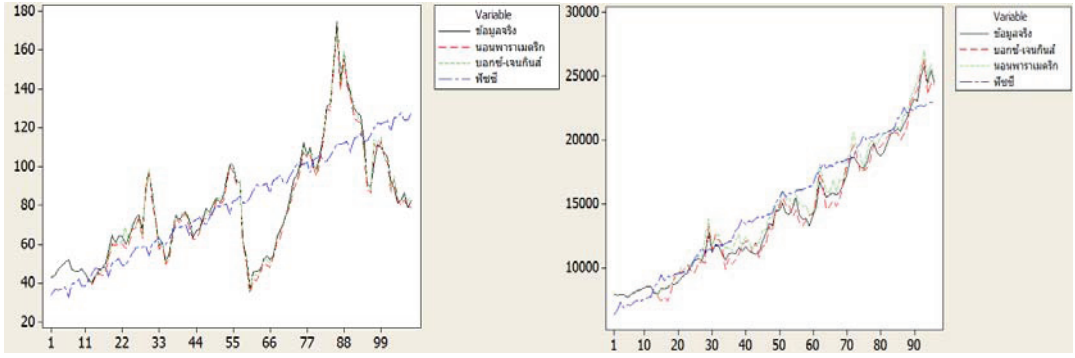
จากอนุกรมเวลาที่เป็นสเตชันนารีดำเนินการตามขั้นตอนของวิธีการทางนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson โดยใช้โปรแกรม SAS 9.1 และหาค่าพยากรณ์โดยใช้สมการที่ (7) ในการศึกษานี้กำหนด kernel function ให้อยู่ในรูปของการแจกแจงปกติมาตรฐาน การคำนวณหาค่า bandwidth ใช้วิธี cross-validation function ดังสมการที่ (10) และเลือกค่า autoregressive (p) หรือการเลือกจำนวนของระยะห่าง (lag) ตามสมการที่ (8) ซึ่งในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยวิธีการทางนอนพาราเมตริกจะไม่ให้ตัวแบบในการพยากรณ์ในรูปแบบของการประมาณค่าพารามิเตอร์แต่จะให้ค่าพยากรณ์ที่แต่ละจุดเวลาผลการวิเคราะห์ข้อมูลทั้ง 2 ชุดแสดงดังภาพที่ 7(a) และภาพที่ 7(b)



7(a) 7(b)

ภาพที่ 7(a) ราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือนเทียบกับค่าพยากรณ์ด้วยวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson (เส้นทึบแทนค่าจริง เส้นประแทนค่าพยากรณ์)
ภาพที่ 7(b) ราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือนเทียบกับค่าพยากรณ์ด้วยวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson (เส้นทึบแทนค่าจริง เส้นประแทนค่าพยากรณ์)

ผลการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ข้อมูลอนุกรมเวลาด้วยวิธีบอกซ์-เจนกินส์ วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีและวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson แสดงดังภาพที่ 8(a) และภาพที่ 8(b)



8(a) 8(b)

ภาพที่ 8(a) ราคายางแผ่นดิบราคาประมูลของตลาดกลางยางพารา อ.หาดใหญ่ รายเดือนเทียบกับค่าพยากรณ์ด้วยวิธีบอกซ์-เจนกินส์ วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีและวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson **ภาพที่ 8(b)** ราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือนเทียบกับค่าพยากรณ์ด้วยวิธีบอกซ์-เจนกินส์ วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีและวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson

ผลการเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ข้อมูลอนุกรมเวลาระหว่างวิธีบอกซ์-เจนกินส์ วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีและวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson ของข้อมูลอนุกรมเวลาทั้ง 2 ชุด เมื่อพิจารณาจากค่า MAPE (%) พบว่าวิธีบอกซ์-เจนกินส์มีประสิทธิภาพดีที่สุดสำหรับข้อมูลทั้ง 2 ชุดโดยมีค่า MAPE (%) 2.42 สำหรับข้อมูลชุดที่ 1 ราคายางพารารายเดือน และมีค่าเท่ากับ 4.08 สำหรับข้อมูลชุดที่ 2 ราคาทองคำรายเดือน ในขณะที่วิธีถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีมีประสิทธิภาพน้อยที่สุดโดยมีค่า MAPE (%) 3.68 สำหรับข้อมูลชุดที่ 1 ราคายางพารารายเดือน และมีค่าเท่ากับ 5.94 สำหรับข้อมูลชุดที่ 2 ราคาทองคำรายเดือน แต่อย่างไรก็ตามการพยากรณ์ข้อมูลด้วยวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya– Watson ให้ค่า MAPE (%) ที่ใกล้เคียงกับวิธีบอกซ์-เจนกินส์ โดยมีค่าเท่ากับ 2.75 สำหรับข้อมูลชุดที่ 1 ราคายางพารารายเดือน และมีค่าเท่ากับ 4.28 สำหรับข้อมูลชุดที่ 2 ราคาทองคำรายเดือน ผลการวิเคราะห์ข้อมูลแสดงดังตารางที่ 1

ตารางที่ 1 การเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ข้อมูลอนุกรมเวลาระหว่างวิธีบอกซ์-เจนกินส์ วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี และวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson พิจารณาจากค่า MAPE(%)

วิธีการพยากรณ์ ยางพารารายเดือน	ข้อมูลชุดที่ 1 ราคาทองคำรายเดือน	ข้อมูลชุดที่ 2 ราคาทองคำรายเดือน
1. วิธีบอกซ์-เจนกินส์	2.42	4.08
2. วิธีถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี	3.68	5.94
3. วิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson	2.75	4.28

ผลการเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์ข้อมูลอนุกรมเวลาระหว่างวิธีบอกซ์-เจนกินส์ วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมีและวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson ของข้อมูลอนุกรมเวลาทั้ง 2 ชุด พิจารณาจากค่า MAPE (%) ที่ช่วงเวลาล่วงหน้า 4 ช่วงเวลา พบว่าวิธีบอกซ์-เจนกินส์มีประสิทธิภาพดีที่สุด โดยให้ค่า MAPE (%) ที่ต่ำกว่าในข้อมูลทั้ง 2 ชุด โดยข้อมูลชุดที่ 1 ราคายางพารารายเดือน พบว่าการพยากรณ์ล่วงหน้า 3 เดือน ให้ค่า MAPE (%) เท่ากับ 3.33 รองลงมาคือการพยากรณ์ล่วงหน้า 6 เดือน ให้ค่า MAPE (%) เท่ากับ 4.71 การพยากรณ์ล่วงหน้า 9 เดือนให้ค่า MAPE (%) เท่ากับ 5.98 และการพยากรณ์ล่วงหน้า 12 เดือนให้ค่า MAPE (%) เท่ากับ 6.22 ตามลำดับ สำหรับข้อมูลชุดที่ 2 ราคาทองคำรายเดือนพบว่าการพยากรณ์ล่วงหน้า 3 เดือนให้ค่า MAPE (%) เท่ากับ 7.58 รองลงมาคือการพยากรณ์ล่วงหน้า 6 เดือนให้ค่า MAPE (%) เท่ากับ 7.64 การพยากรณ์ล่วงหน้า 9 เดือนให้ค่า MAPE (%) เท่ากับ 7.86 และการพยากรณ์ล่วงหน้า 12 เดือนให้ค่า MAPE (%) เท่ากับ 8.04 ส่วนวิธีที่มีประสิทธิภาพรองลงมาคือวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson และวิธีที่มีประสิทธิภาพน้อยที่สุดคือวิธีถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี ผลการวิเคราะห์ข้อมูลแสดงดังตารางที่ 2

ตารางที่ 2 ผลการเปรียบเทียบช่วงการพยากรณ์ล่วงหน้า ระหว่างวิธีบอกซ์-เจนกินส์ วิธีการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี และวิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson พิจารณาจากค่า MAPE(%)

วิธีการพยากรณ์	ข้อมูลชุดที่ 1 ราคายางพารารายเดือน				ข้อมูลชุดที่ 2 ราคาทองคำรายเดือน			
	3 เดือน	6 เดือน	9 เดือน	12 เดือน	3 เดือน	6 เดือน	9 เดือน	12 เดือน
1. วิธีบอกซ์-เจนกินส์	3.33	4.71	5.98	6.22	7.58	7.64	7.86	8.04
2. วิธีถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี	4.35	5.12	6.45	7.38	8.54	9.38	10.12	11.78
3. วิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson	3.58	4.89	5.12	6.39	8.19	8.64	9.11	9.89

สรุปและอภิปรายผล

ผลจากการเปรียบเทียบประสิทธิภาพของวิธีการพยากรณ์ข้อมูลอนุกรมเวลาทั้ง 3 วิธี ได้แก่วิธีการทางพาราเมตริก 2 วิธี คือ วิธีบอกซ์- เจนกินส์ และการถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี ในขณะที่วิธีการทางนอนพาราเมตริกคือการใช้ตัวประมาณ Nadaraya – Watson ในการพยากรณ์ค่าอนุกรมเวลา โดยศึกษาจากข้อมูลจริงจำนวน 2 ชุด ข้อมูลชุดที่ 1 คือ ราคายางแผ่นดิบรายเดือน และข้อมูลชุดที่ 2 คือ ราคาทองคำในประเทศไทย (น้ำหนักทอง 1 บาท) รายเดือนพบว่าวิธีที่มีประสิทธิภาพมากที่สุดในการพยากรณ์ข้อมูลอนุกรมเวลาการพยากรณ์ข้อมูลอนุกรมเวลาล่วงหน้าที่ช่วงเวลา 3 เดือน 6 เดือน 9 เดือน และ 12 เดือนคือ วิธีบอกซ์-เจนกินส์ เกณฑ์ที่ใช้การพิจารณาคือค่า MAPE (%) ส่วนวิธีที่มีประสิทธิภาพรองลงมาคือ วิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson และวิธีที่มีประสิทธิภาพน้อยที่สุดคือ วิธีถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี แต่อย่างไรก็ตามวิธีการพยากรณ์ข้อมูลโดยวิธีบอกซ์-เจนกินส์ ก็มีข้อจำกัดในบางกรณี เช่น ต้องมีการพิจารณาเลือกรูปแบบของแนวโน้มให้สอดคล้องกับลักษณะแนวโน้มของข้อมูลเพื่อนำไปสู่ตัวแบบการพยากรณ์ที่มีประสิทธิภาพ แต่ในบางกรณีข้อมูลไม่ได้มีรูปแบบแนวโน้มหรือฤดูกาลที่เด่นชัด จนอาจต้องใช้ข้อมูลในอดีตเป็นจำนวนมากจึงจะเห็นส่วนประกอบการเปลี่ยนแปลงของวัฏจักรอย่างเด่นชัด ในขณะเดียวกัน การเลือกรูปแบบโดยการพิจารณาจากค่า ACF และ PACF นั้นต้องพิจารณาอย่างละเอียด เนื่องจากข้อมูลในสถานการณ์จริงที่เกิดขึ้นนั้นไม่ตรงตามทฤษฎี อาจส่งผลต่อการเลือกรูปแบบของสมการพยากรณ์ จากการศึกษานี้ วิธีนอนพาราเมตริกโดยใช้ตัวประมาณ Nadaraya – Watson เป็นทางเลือกอีกหนึ่งทางเลือกสำหรับการพยากรณ์ข้อมูลอนุกรมเวลา เนื่องจากมีความยืดหยุ่นต่อเงื่อนไขต่างๆ ในการประมาณค่าพารามิเตอร์ แต่ในขณะเดียวกันมีประสิทธิภาพในการพยากรณ์ข้อมูลอนุกรมเวลาที่เทียบเคียงได้กับวิธีบอกซ์-เจนกินส์ซึ่งในการพยากรณ์ข้อมูลอนุกรมเวลาด้วยวิธีการทางนอนพาราเมตริกจะไม่ให้ตัวแบบในการพยากรณ์ในรูปแบบของการประมาณค่าพารามิเตอร์แต่จะให้ค่าพยากรณ์ที่แต่ละจุดเวลาโดยใช้หลักการของ Smoothing Technique ซึ่งการเลือกพารามิเตอร์ปรับให้เรียบ (Bandwidth) มีความสำคัญอย่างยิ่งเนื่องจาก เมื่อ Bandwidth มีค่าสูงจะทำให้รูปแบบของค่าพยากรณ์ที่ได้ในแต่ละจุดเวลาลักษณะเรียบ (Smooth) แต่ไม่เหมาะสมกับข้อมูล (Well Fitting) ในขณะเดียวกันเมื่อ bandwidth มีค่าต่ำรูปแบบของค่าพยากรณ์ที่ได้ในแต่ละจุดเวลาจะเหมาะสมกับข้อมูลแต่ไม่เรียบ หากพิจารณาในแง่ของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (MSE) และความเอนเอียง (Bias) พบว่า ถ้าให้ความสำคัญกับความเรียบความเอนเอียงจะสูงแต่ถ้าให้ความสำคัญกับความเหมาะสมของข้อมูลความแปรปรวนจะสูงดังนั้นในการนำวิธีการพยากรณ์ข้อมูลอนุกรมเวลาไปใช้ จึงควรพิจารณาตามลักษณะของข้อมูลอนุกรมเวลาที่มีอยู่และข้อจำกัดในเรื่องต่างๆ ของการกำหนดตัวแบบในการพยากรณ์

ข้อเสนอแนะ

จากการศึกษานี้ วิธีการทางนอนพาราเมตริกในการพยากรณ์ข้อมูลอนุกรมเวลาเป็นอีก 1 ทางเลือกในการพยากรณ์ข้อมูลอนุกรมเวลา เนื่องจากมีประสิทธิภาพที่ใกล้เคียงกับวิธีทางพาราเมตริก นอกจากนั้นยังมีความยืดหยุ่นต่อเงื่อนไขต่างๆ ในการหาค่าพยากรณ์ ดังนั้นควรขยายไปสู่วิธีการทางนอนพาราเมตริกในรูปแบบอื่น เช่น ตัวประมาณ Local Linear, Spline, Loess, Lowess เพื่อเปรียบเทียบประสิทธิภาพกับตัวประมาณ Nadaraya – Watson วิธีบอกซ์- เจนกินส์ การประยุกต์ใช้ฟัซซีในการพยากรณ์ข้อมูลอนุกรมเวลาหรือวิธีการอื่นๆ เป็นต้น

เอกสารอ้างอิง

[1] Song, K. B., Baek, Y. S., Hong, D. H.; and Jang, G. (2005). Short-term load forecasting for the holidays using fuzzy linear regression method. *Power Systems, IEEE Transactions on*. 20(1): 96-101.

[2] ยุทธชัย เลิศวรปรัชญ์. (2558). การพยากรณ์อนุกรมเวลาแบบฟัซซีไทพ์ทูอันดับสูงด้วยการถ่วงน้ำหนัก. ใน *ACTIS & NCOBA 2015*. หน้า 171-178.

[3] กนกวรรณ วิไลศรี. (2547). *การพยากรณ์อนุกรมเวลาที่มีฤดูกาลโดยใช้การถดถอยแบบฟัซซีที่ใช้ตัวแปรดัมมี*. ปริญญานิพนธ์ วท.ม. (สถิติประยุกต์). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยธรรมศาสตร์.

[4] Vilar-Fernández, J. M. and Cao, R. (2007). Nonparametric Forecasting in Time Series–AComparative Study. *Communications in Statistics-Simulation and Computation®*. 36(2): 311-334.

[5] Vogt, M. (2012). Nonparametric regression for locally stationary time series. *The Annals of Statistics*. 40(5): 2601-2633.

[6] Vilar, J. M., Cao, R. and Aneiros, G. (2012). Forecasting next-day electricity demand andprice using nonparametric functional methods. *International Journal of Electrical Power & Energy Systems*. 39(1): 48-55.

[7] Aljuhani, K.H. (2014). Modification of the Adaptive Nadaraya-Watson Kernel Regression Estimator. *Scientific Research and Essays*. 9(22): 966-971.

[8] ตลาดกลางยางพารา อ. หาดใหญ่ จ. สงขลา. (2556). *ราคายางพารารายเดือน*. สืบค้นเมื่อ 20 เมษายน 2558, จาก http://www.live-rubber.com/index.php/para-rubber-articles/40-para-rubber-statistic/104-rubber-price-in-thailand

[9] สมาคมค้าทอง. (2555). *ราคาทองคำรายเดือน*. สืบค้นเมื่อ 15 เมษายน 2558, จาก http://www.goldtraders.or.th

[10] ทรงศิริ แต้สมบัติ. (2549). *การพยากรณ์เชิงปริมาณ*. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์.

[11] Hárdle, W. and Linton, O. (1994). Applied nonparametric methods. *Handbook of Econometrics*. 4: 2295-2339.

[12] Tjostheim, D. and Auested, B.H. (1994). NonparametricIdentification of Nonlinear Time Series: Selecting Significant Lags. *Journal of American Statistics Association*. 89: 1410-1419.