การเปรียบเทียบตัวแบบการพยากรณ์ราคาหุ้นโดยใช้แบบจำลองอารีมาและอารีแม็กซ์
A comparison of forecasting models of stock price using ARIMA and ARIMAX models

บุญกอง ทะกลโยธิน และ ยุพาภรณ์ อารีพงษ์*

ภาควิชาสถิติประยุกต์ คณะวิทยาศาสตร์ประยุกต์

มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ กรุงเทพมหานคร 10800

Boonkong Dhakonlayodhin and Yupaporn Areepong*

Applied Statistics Department, Faculty of Applied Science,

King Mongkut's University of Technology North Bangkok, Bangkok 10800

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์ราคาหุ้นโดยเปรียบเทียบจากแบบจำลองอารีมา (Autoregressive Integrated Moving Average: ARIMA) และแบบจำลองอารีแม็กซ์ (Autoregressive Integrated Moving Average with Exogenous Variable: ARIMAX) โดยใช้ข้อมูลราคาหุ้น BBL รายเดือน ตั้งแต่เดือนมกราคม พ.ศ. 2555 ถึงเดือนธันวาคม พ.ศ. 2559 และตัวแปรอิสระที่นำมาศึกษาคือ อัตราการแลกเปลี่ยน รวมทั้งสิ้น 60 ข้อมูล โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) เป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพของแบบจำลอง จากการพยากรณ์พบว่า แบบจำลอง ARIMAX เหมาะสมกับข้อมูลชุดนี้มากที่สุด เนื่องจากให้ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (MAPE) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำที่สุด

คำสำคัญ: แบบจำลองอารีมา แบบจำลองอารีแม็กซ์ ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ ความนิ่งของข้อมูล ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

Corresponding author: yupaporn.a@sci.kmutnb.ac.th

Abstract

This objective of this research is to forecast the stock price by using different forecasting models which are Autoregressive Integrated Moving Average (ARIMA) and Autoregressive Integrated Moving Average with Exogenous Variable (ARIMAX) models. This research is based on the BBL stock statement, being collected monthly from January 2012 to December 2016. The explanatory variable is the 60 records of the currency exchange rate which was collected monthly between 2012 and 2016. The forecasting accuracy and efficiency were considered by Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE). As the result, we found that the ARIMAX model is the most efficient one by providing the minimum Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE).

Keywords: ARIMA model, ARIMAX model, Mean Absolute Percentage Error (MAPE), Stationary,RootMean Squared Error (RMSE)

บทนำ

ในยุคปัจจุบันนี้การลงทุนในตลาดหลักทรัพย์ของประเทศไทยเป็นสิ่งที่น่าสนใจ เนื่องจากการลงทุนในตลาดหลักทรัพย์สามารถให้ผลตอบแทนที่สูงกว่าการฝากเงินไว้กับธนาคาร แต่เนื่องจากราคาหลักทรัพย์มีความผันผวนไม่แน่นอนอันเกิดมาจากสาเหตุหลายปัจจัย และนำไปสู่ความเสี่ยงในการลงทุน โดยเฉพาะอย่างยิ่งเมื่อราคาตกต่ำลงมากกว่าราคาเริ่มแรกที่นักลงทุนเข้าซื้อ แม้หุ้นจะเป็นหลักทรัพย์ที่ผู้ลงทุนหลายคนลังเลที่จะลงทุนเพราะไม่แน่ใจในความเสี่ยงที่อาจจะเกิดขึ้น แต่คงปฏิเสธไม่ได้ว่าหุ้นยังคงมีเสน่ห์สำหรับผู้ลงทุนอีกหลายคนที่ไม่ว่าจะเสี่ยงแค่ไหนก็ต้องมีหุ้นอยู่ในพอร์ตการลงทุนเสมอ นั่นเป็นเพราะการลงทุนในหุ้นมีโอกาสได้รับผลตอบแทนที่น่าสนใจ ทั้งในรูปของเงินปันผล (dividend) และกำไรจากการขายหุ้น (capital gain) อย่างไรก็ตามการลดความเสี่ยงจากการสูญเสียเงินทุนที่หายไปในตลาดหลักทรัพย์สามารถหลีกเลี่ยงได้ หากนำผลการวิเคราะห์ในเชิงเทคนิคเข้ามาช่วยในการตัดสินใจลงทุนเข้าซื้อขายหลักทรัพย์ โดยการเลือกเครื่องมือทางสถิติที่เหมาะสมกับทฤษฎีทางการเงิน เพื่อเสริมสร้างหรือหาจังหวะในการเข้าซื้อขายหุ้นในตลาดหลักทรัพย์ของประเทศไทย อันนำไปสู่การลดความเสี่ยงของการขาดทุนที่จะเกิดขึ้นแก่ผู้ลงทุน

การพยากรณ์เป็นการคาดคะเนของเหตุการณ์ในอนาคต โดยอาศัยรูปแบบของการเกิดเหตุการณ์หรือการพยากรณ์ที่เก็บข้อมูลจากอดีต รวมถึงความรู้ความสามารถของผู้พยากรณ์ หากทราบเหตุการณ์ต่าง ๆ ในอนาคต มีความเป็นไปได้ที่จะเพิ่มความเชื่อมั่นให้กับการวางแผนดำเนินงานที่มีความถูกต้องและผิดพลาดน้อยที่สุด ซึ่งมีความสอดคล้องกับการคาดการณ์ราคาหุ้นและวิธีการพยากรณ์อนุกรมเวลาสำหรับการพยากรณ์อนุกรมเวลา (time series forecasting) มีอยู่ด้วยกันหลายวิธี วิธีที่นิยมใช้กันมาก คือ วิธีของ Box-Jenkins ด้วยตัวแบบ ARIMA (Autoregressive Integrated Moving Average model) ซึ่งเป็นวิธีหาตัวแบบพยากรณ์โดยอาศัยความสัมพันธ์ของข้อมูลในอดีตเพื่อหาตัวแบบที่แสดงพฤติกรรมของข้อมูล และใช้เป็นแนวทางในการพยากรณ์ค่าในอนาคต [1-3] ในบางครั้งจะพบว่าข้อมูลอนุกรมเวลาอาจจะได้รับผลกระทบมาจากปัจจัยอื่น ๆ ได้ (exogenous variable) ซึ่งอาจจะส่งผลต่อความแม่นยำของค่าพยากรณ์ได้ ดังนั้นจึงมีการพัฒนาแบบจำลองที่เป็นการพยากรณ์ข้อมูลอนุกรมเวลาโดยอาศัยปัจจัยอื่น ซึ่งเรียกว่าแบบจำลองอารีแม็กซ์ (Autoregressive Integrated Moving Average with Exogenous model: ARIMAX) นั่นเอง [4, 5] โดยในงานวิจัยนี้ได้สนใจการพยากรณ์ราคาหุ้น

โดยพบว่าการเปลี่ยนแปลงของราคาหุ้นขึ้นอยู่กับอัตราแลกเปลี่ยนกล่าวคือ การเปลี่ยนแปลงของอัตราแลกเปลี่ยนมีทิศทางเดียวกันกับการเคลื่อนไหวของตลาดหุ้น โดยเฉพาะในตลาดเกิดใหม่อย่างไทยที่ต้องพึ่งพาเงินลงทุนจากต่างชาติ คือ เมื่อค่าเงินหรืออัตราแลกเปลี่ยนอ่อนค่าลง ตลาดหุ้นก็จะปรับตัวลดลงด้วย ดังนั้นผู้วิจัยจึงมีแนวคิดว่าหากนำปัจจัยที่ส่งผลต่อข้อมูลที่นำมาศึกษามาร่วมในตัวแบบการพยากรณ์น่าจะส่งผลให้ค่าพยากรณ์มีความแม่นยำสูงขึ้น จากที่กล่าวมาข้างต้น คณะผู้วิจัยจึงมีความสนใจในการเปรียบเทียบประสิทธิภาพระหว่างแบบจำลอง ARIMA และแบบจำลอง ARIMAX เพื่อหาตัวแบบที่เหมาะสมในการพยากรณ์ราคาหุ้นให้มีความถูกต้องและผิดพลาดน้อยที่สุด

วิธีดำเนินการวิจัย

1. การคัดเลือกข้อมูล

นำข้อมูลราคาปิดหุ้น BBL รายเดือน ตั้งแต่เดือนมกราคม พ.ศ. 2555 ถึงเดือนธันวาคม พ.ศ. 2559 จากตลาดหลักทรัพย์แห่งประเทศไทย โดยแบ่งข้อมูลออกเป็น 2 ชุด คือข้อมูลตั้งแต่เดือนมกราคม พ.ศ. 2555 ถึงเดือนธันวาคม พ.ศ. 2558 เป็นชุดข้อมูลสำหรับนำไปสร้างตัวแบบ และข้อมูลปี พ.ศ. 2559 เป็นข้อมูลสำรองสำหรับการทดสอบความแม่นยำของตัวแบบในการพยากรณ์

2. การทดสอบความนิ่งของข้อมูลอนุกรมเวลา (unit root test)

การทดสอบความนิ่งของข้อมูลอนุกรมเวลาเป็นการทดสอบว่าข้อมูลที่นำมาศึกษามีลักษณะนิ่ง (stationary) หรือไม่ โดยจำเป็นที่จะต้องมีการทดสอบก่อนที่จะนำข้อมูลไปประมาณค่า ซึ่งโดยทั่วไปข้อมูลแบบอนุกรมเวลาจะมีลักษณะ “ไม่นิ่ง” (nonstationary) และข้อมูลมาจากกระบวนการเชิงสุ่ม (random process) ดังนั้นจึงต้องนำข้อมูลมาทดสอบว่ามีความนิ่งหรือไม่ โดยการทดสอบยูนิทรูท (unit root) ดังสมการต่อไปนี้

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad (1)$$

โดยที่ ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของประชากรและมีสมมติฐานในการทดสอบ คือ

$$\mathrm{H}_0 : \rho = 1$$

$$\mathrm{H}_1 : |\rho| < 1$$

ถ้ายอมรับสมมติฐาน H_0 แสดงว่าข้อมูลมีลักษณะไม่นิ่ง แต่ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าข้อมูลนั้นมีลักษณะนิ่ง

การทดสอบแบบ Dickey-Fuller (DF) ถูกนำเสนอโดย Dickey และ Fuller ในปี ค.ศ. 1979 [6] โดยมีแนวคิดในการทดสอบเพื่อให้ง่ายต่อการทดสอบยูนิทรูทของข้อมูลอนุกรมเวลาว่ามีความนิ่งหรือไม่ ซึ่งได้เสนอให้ θ มีค่าเท่ากับศูนย์ โดยนำ X_{t-1} ลบทั้งสองข้างของสมการที่ (1) และสามารถจัดรูปสมการใหม่ได้ดังนี้

$$\Delta X_t = \theta X_{t-1} + \varepsilon_t \qquad (2)$$

โดยมีสมมติฐานการทดสอบ คือ

$$\mathrm{H}_0 : \theta = 1$$

$$\mathrm{H}_1 : \theta \neq 1$$

ถ้ายอมรับสมมติฐาน H_0 แสดงว่าข้อมูลมีลักษณะไม่นิ่ง แต่ถ้าปฏิเสธสมมติฐาน H_0 แสดงว่าข้อมูลนั้นมีลักษณะนิ่ง และทำการทดสอบรูปแบบสมการเป็น 3 กรณี ดังนี้

กรณีไม่มีค่าคงที่และแนวโน้มเวลา $\Delta X_t = \theta X_{t-1} + \varepsilon_t$

กรณีมีเฉพาะค่าคงที่ $\Delta X_t = \alpha + \theta X_{t-1} + \varepsilon_t$

กรณีมีทั้งค่าคงที่และแนวโน้มเวลา $\Delta X_t = \alpha + \beta_t + \theta X_{t-1} + \varepsilon_t$

โดยที่ θ คือ ค่าพารามิเตอร์ของการทดสอบยูนิทรูท

β_t คือ ค่าคงที่แนวโน้ม

α คือ ค่าคงที่ระดับ

ต่อมาในปี ค.ศ. 1984 Said และ Dickey [7] ได้นำเสนอวิธีการทดสอบ Augmented Dickey-Fuller Test (ADF) โดยเพิ่มจำนวนพจน์ของ lagged difference เข้าไปในสมการเพื่อแก้ปัญหาความสัมพันธ์ในตัว (autocorrelation) ดังสมการต่อไปนี้

กรณีไม่มีค่าคงที่และแนวโน้มเวลา $\Delta X_t = \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_1 \Delta X_{t-1} + \varepsilon_t$

กรณีมีเฉพาะค่าคงที่ $\Delta X_t = \alpha + \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_1 \Delta X_{t-1} + \varepsilon_t$

กรณีมีทั้งค่าคงที่และแนวโน้มเวลา $\Delta X_t = \alpha + \beta_t \theta X_{t-1} + \sum_{i=1}^{\rho} \phi_1 \Delta X_{t-1} + \varepsilon_t$

ในการตรวจสอบว่าข้อมูลมีลักษณะนิ่งหรือไม่นิ่งนั้น สามารถเปรียบเทียบค่าสถิติ t ที่คำนวณได้กับค่าวิกฤต (critical value) ในตาราง ADF โดยใช้สถิติ t (t-statistic) ซึ่งมีสูตรดังต่อไปนี้

$$t = \frac{\hat{\theta}}{S.E._{\hat{\theta}}} \tag{3}$$

3. การวิเคราะห์ตัวแบบ นำชุดข้อมูลที่ใช้สำหรับสร้างตัวแบบมาทำการวิเคราะห์เพื่อหาสมการของทั้ง 2 ตัวแบบ ดังนี้

3.1 แบบจำลองอารีมา (ARIMA (p,d,q))

ตัวแบบ ARIMA (p,d,q) มีตัวแบบดังนี้ [8]

$$(1-\mathrm{B})^{\mathrm{d}}(1-\phi_1\mathrm{B}-\phi_2\mathrm{B}^2-\ldots-\phi_{\mathrm{P}}\mathrm{B}^{\mathrm{P}})\mathrm{Y}_{\mathrm{t}} = \theta_0 + (1-\theta_1\mathrm{b}-\theta_2\mathrm{B}^2-\ldots-\theta_{\mathrm{q}}\mathrm{B}^{\mathrm{q}})\varepsilon_{\mathrm{t}} \tag{4}$$

เมื่อ θ_0 คือ ค่าคงที่

d คือ จำนวนครั้งของการหาผลต่าง

ϕ_{i} คือ พารามิเตอร์ของการถดถอยในตัว (autoregressive parameter)

θ_{i} คือ พารามิเตอร์ของค่าเฉลี่ยเคลื่อนที่ (moving average parameter)

ε_{t} คือ ค่าความคลาดเคลื่อน ณ เวลา t ที่มีการแจกแจงปกติซึ่งมีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ σ^2

3.2 แบบจำลองอารีแม็กซ์ (ARIMAX (p,d,q,r))

ตัวแบบ ARIMAX (p,d,q,r) มีตัวแบบดังนี้ [5]

$$(1-B)^d(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_P B^P)Y_t=\theta_0+(1-\theta_1 b-\theta_2 B^2-\ldots-\theta_q B^q)\varepsilon_t+\sum_{i=1}^{r}\alpha_i X_{it} \quad (5)$$

เมื่อ θ_0 คือ ค่าคงที่

d คือ จำนวนครั้งของการหาผลต่าง

ϕ_i คือ พารามิเตอร์ของการถดถอยในตัว (autoregressive parameter)

θ_i คือ พารามิเตอร์ของค่าเฉลี่ยเคลื่อนที่ (moving average parameter)

α_i คือ พารามิเตอร์ของ X_t ตัวที่ i : i = 1, 2,..., r

ε_t คือ ค่าความคลาดเคลื่อน ณ เวลา t ที่มีการแจกแจงปกติซึ่งมีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ σ^2

4. เกณฑ์ในการตัดสินใจ

การนำวิธีพยากรณ์โดยตัวแบบอารีมาและตัวแบบอารีแม็กซ์มาเปรียบเทียบกันว่าวิธีใดเหมาะสมกับข้อมูลที่นำมาศึกษา โดยพิจารณาจากค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error: MAPE) และพิจารณาจากค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE)

4.1 ค่าเฉลี่ยของเปอร์เซ็นต์ความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Percentage Error: MAPE)

$$MAPE=\frac{1}{n}\sum_{i=1}^{n}\frac{|y_t-\hat{y}_t|}{y_t}\times 100 \quad (6)$$

4.2 ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE)

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t-\hat{y}_t)^2} \quad (7)$$

โดยที่ y_t คือ ค่าจริงของข้อมูลอนุกรมเวลา ณ เวลาที่ t

$\hat{y}_t$ คือ ค่าพยากรณ์ของข้อมูลอนุกรมเวลา ณ เวลาที่ t

n คือ จำนวนข้อมูลทั้งหมด

เกณฑ์ในการเลือกตัวแบบพิจารณาจากค่า MAPE และ RMSE ของการพยากรณ์ในแต่ละวิธีที่มีค่าต่ำที่สุด ซึ่งแสดงว่าวิธีการพยากรณ์ดังกล่าวเหมาะสมกับข้อมูลอนุกรมเวลาชุดที่ใช้สร้างตัวแบบนั้น

5. ขั้นตอนการพยากรณ์

การพยากรณ์อนุกรมเวลาโดยแบบจำลองอารีมาและอารีแม็กซ์มีดังต่อไปนี้

1) การกำหนดรูปแบบ (identification) นำอนุกรมเวลาที่มีความนิ่งมาดำเนินการศึกษาหารูปแบบ ARIMA (p,d,q) ที่เหมาะสมให้กับอนุกรมเวลา โดยที่ค่า autocorrelation มีค่าอยู่ใช่วง [1,-1] โดยพิจารณา (r_k) ของอนุกรมตัวอย่างที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ดังสูตรต่อไปนี้

$$r_k = \frac{\sum_{t=1}^{n-k}(Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2} \qquad (8)$$

เมื่อ Y_t คือ ข้อมูลหรือค่าสังเกต ณ เวลา t

k คือ จำนวนช่วงข้อมูลที่ห่างกัน

Partial Autocorrelation: r_{kk} คือ การวัดความสัมพันธ์ของแต่ละช่วงเวลา โดยมีช่วงเวลาที่ย้อนหลังไป k หน่วยเวลาโดยพิจารณาค่า Partial Autocorrelation (r_{kk}) ของอนุกรมเวลาตัวอย่างที่มีช่วงเวลาย้อนหลังไป k หน่วยเวลา ซึ่งมีสูตรดังนี้

$$r_{kk} = \begin{cases} r_1 \\ \dfrac{r_k - \sum_{j=1}^{k-1} r_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k=1} r_{k-1,j} r_j} \end{cases} \qquad (9)$$

ในการพิจารณาแต่ละรูปแบบ ต้องพิจารณา r_k, r_{kk} พร้อมกันหลาย ๆ ค่า จึงมักพิจารณาจากรูปที่เรียกว่า “คอเรลโลแกรม” (correlogram) ที่ได้จากการพล็อต r_k, r_{kk}

2) การประมาณค่าพารามิเตอร์ (estimation) ของแบบจำลอง โดยการหาค่าประมาณจากวิธีกำลังสองน้อยที่สุด

3) การตรวจแบบจำลอง (diagnostic checking) เมื่อกำหนดรูปแบบและประมาณค่าพารามิเตอร์ในแบบจำลอง จะต้องทำการตรวจสอบทุกครั้งว่ารูปแบบที่กำหนดนั้นมีความเหมาะสมจริงหรือไม่ โดยการตรวจสอบสามารถทำได้หลายวิธี ได้แก่ การพิจารณาคอเรลโลแกรมของ r_k ของค่าความคลาดเคลื่อน และการทดสอบความเหมาะสมของแบบจำลองโดยการทดสอบของ Box และ Liung ซึ่งพิจารณาจากค่า Q-statistic ดังสมการนี้

$$Q_m = n(n+2)\sum_{k=1}^{m}\left\{\frac{r_k^2(e_t)}{(n-k)}\right\} \qquad (10)$$

โดยที่ e_t คือ ค่าความคลาดเคลื่อนของการพยากรณ์ ณ เวลา t

n คือ จำนวนอนุกรมเวลาของ e_t ทั้งหมด

m คือ ช่วงเวลาที่ห่างกันมากที่สุดของ e_t ทั้งหมด

ซึ่งค่า Q ที่ได้มีการแจกแจงแบบ Chi-square และมีองศาความเป็นอิสระ (degree of freedom) เท่ากับ m-n โดยมีสมมติฐานว่างเป็นพจน์ของความคลาดเคลื่อนที่ได้จากการประมาณที่มีลักษณะเป็น white noise หมายถึง แบบจำลองที่ไม่มี อัตสหสัมพันธ์ ถ้าหากแบบจำลองที่ได้ไม่มีอัตสหสัมพันธ์ให้ใช้แบบจำลองนี้สำหรับการพยากรณ์ต่อไป

4) การพยากรณ์ (forecasting) เมื่อได้แบบจำลองที่เหมาะสม สามารถทำการพยากรณ์ได้ทั้งแบบจุด (point forecast) และแบบช่วง (interval forecast)

ผลการวิจัย

1. ผลการทดสอบ unit root test

ในการศึกษาครั้งนี้ได้แปลงข้อมูลราคาหุ้น BBL ให้อยู่ในรูปของ log ก่อนที่จะนำไปทดสอบความนิ่งของข้อมูลราคาหุ้น BBL เพื่อลดความแปรปรวนของข้อมูล ซึ่งจากการทดสอบ Augmented Dickey-Fuller (ADF) ที่ระดับ Level พบว่า ชุดข้อมูลทั้ง 3 มีลักษณะแบบ nonstationary จึงทำการหาผลต่างลำดับที่ 1 (1st Difference) และผลต่างลำดับที่ 2 (2nd Difference) (ตารางที่ 1)

ตารางที่ 1 ผลการทดสอบ unit root ของหุ้น BBL

ระดับ Level							
		None		Intercept		Trend and Intercept	
Stock Price	lag	ADF Statistic	% critical value	ADF Statistic	% critical value	ADF Statistic	% critical value
Log(SP)	0	-0.141	1% - 2.605 5% - 1.946 10% -1.613	-2.316	1% - 3.546 5% - 2.912 10% -2.594	-3.969	1% - 4.121 5% - 3.488 10% -3.172
ผลต่างลำดับที่ 1 (1st Difference)							
D(logSP)	0	-8.257*	1% - 2.605 5% - 1.947 10% -1.613	-8.190*	1% - 3.548 5% - 2.913 10% -2.594	-8.145*	1% - 4.124 5% - 3.489 10% -3.173
ผลต่างลำดับที่ 2 (2nd Difference)							
D(logSP,2)	0	-6.690*	1% - 2.608 5% - 1.947 10% - 1.613	-6.631*	1% - 3.557 5% - 2.917 10% -2.596	-6.575*	1% - 4.137 5% - 3.495 10% -3.177

ผลการทดสอบราคาหุ้น BBL ที่ระดับ level จากการพิจารณาเปรียบเทียบค่าสถิติ ADF กับค่า Mackinnon critical ที่ระดับ 1, 5 และ 10 เปอร์เซ็นต์ ของอนุกรมเวลาปรากฏว่าข้อมูลอนุกรมเวลานั้นไม่มีความนิ่ง เนื่องจากค่าสถิติ ADF มีค่ามากกว่า Mackinnon critical แสดงว่าข้อมูลไม่มีความนิ่งที่ระดับ level ทำให้พยากรณ์ไม่ได้ ดังนั้นจึงต้องทำการหาผลต่างลำดับที่ 1 และ 2 ตามลำดับ แล้วทำการเปรียบเทียบค่าสถิติ ADF กับค่า Mackinnon critical ที่ระดับต่าง ๆ พบว่า ค่าสถิติ ADF น้อยกว่าค่า Mackinnon critical แสดงว่าข้อมูลอนุกรมเวลานั้นมีความนิ่งแล้ว

2. การกำหนดรูปแบบ (identification)

เมื่อพิจารณาจาก correlogram ของข้อมูลอนุกรมเวลา ณ ผลต่างลำดับที่ 1 (1st Difference) พบว่าข้อมูลมีลักษณะนิ่งแบบ white noise กล่าวคือ ราคาหุ้น BBL ขึ้นอยู่กับค่าความคลาดเคลื่อนแบบสุ่ม ซึ่งไม่สามารถกำหนดรูปแบบของตัวแบบได้ จึงได้ทำการพิจารณา correlogram ของผลต่างลำดับที่ 2 (2nd Difference) ของข้อมูล (ตารางที่ 2) พบว่ามีลักษณะนิ่งและสามารถหารูปแบบของตัวแบบได้โดยการกำหนดตัวแบบเพื่อหาอันดับที่ p ของตัวแบบ autoregressive และอันดับที่ q ของตัวแบบ moving average ซึ่งจะพิจารณาจากค่า Autocorrelation Function (ACF) และค่า Partial Autocorrelation (PACF) ในการสร้างตัวแบบ ARIMA(p,d,q) โดยพิจารณาว่า ACF และ PACF ที่เกินออกมานอกช่วงความเชื่อมั่น 95 เปอร์เซ็นต์ ซึ่งจากตารางที่ 1 จะได้ตัวแบบคือ ARIMA(2,1)

ตารางที่ 2 Correlogram ณ ผลต่างลำดับที่ 2 (2^{nd} Difference)

Sample: 2012M01 2016M12
Included observations: 58

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.437	-0.437	11.675	0.001
		2	-0.064	-0.315	11.929	0.003
		3	0.055	-0.164	12.119	0.007
		4	-0.097	-0.227	12.727	0.013
		5	0.109	-0.069	13.510	0.019
		6	-0.054	-0.089	13.702	0.033
		7	0.149	0.160	15.215	0.033
		8	-0.274	-0.184	20.441	0.009
		9	0.166	0.001	22.403	0.008
		10	0.021	0.012	22.436	0.013
		11	-0.120	-0.067	23.500	0.015
		12	0.068	-0.103	23.855	0.021
		13	-0.043	-0.073	24.000	0.031
		14	0.077	-0.000	24.469	0.040
		15	0.006	0.107	24.472	0.058
		16	0.016	0.080	24.494	0.079
		17	-0.080	0.026	25.037	0.094
		18	-0.017	-0.013	25.061	0.123
		19	0.087	0.011	25.736	0.138
		20	-0.005	0.052	25.738	0.175
		21	0.057	0.136	26.047	0.205
		22	-0.174	-0.087	28.964	0.146
		23	0.100	0.033	29.958	0.151
		24	-0.037	-0.048	30.100	0.181
		25	0.037	0.024	30.245	0.215
		26	0.024	0.009	30.310	0.255
		27	-0.171	-0.156	33.585	0.178
		28	0.140	-0.072	35.860	0.146
		29	0.030	0.083	35.966	0.175
		30	-0.037	-0.041	36.133	0.204
		31	-0.064	-0.071	36.658	0.223
		32	0.041	-0.028	36.886	0.253
		33	0.045	0.039	37.172	0.283
		34	-0.029	0.034	37.291	0.320
		35	0.031	-0.058	37.433	0.358
		36	-0.071	-0.052	38.235	0.368

3. การประมาณค่าพารามิเตอร์ (estimation)

3.1 แบบจำลองอารีมา

ตัวแบบแสดงดังนี้

$$Z_t = -0.000202 + 2Z_{t-1} - Z_{t-2} + 0.972921\varepsilon_{t-1} + \varepsilon_t$$

เมื่อ $Z_t = \log(Y_t)$

ตารางที่ 3 ค่าสัมประสิทธิ์และค่าสถิติของตัวแบบ ARIMA(2,1)

Variable	Coefficient	t-statistic	Prob.
c	-0.000202	-0.350280	0.7274
MA(1)	-0.972921	-22.12921	0.0000
R-squared		0.534845	
Durbin-Watson stat		2.076677	
Akaike information criterion		-3.010006	
Schwarz information criterion		-2.938956	
Residual sum of square		0.156233	

จากตารางที่ 3 พบว่าค่าสัมประสิทธิ์ของ MA(1) มีค่าเท่ากับ -0.972921 ซึ่งมีค่า Prob. ของ c มีค่ามากกว่า 0.05 ดังนั้น จึงต้องตัดค่า c ออกจากตัวแบบ เมื่อตัดค่า c ออกแล้วจะได้รูปแบบจำลองที่ศึกษาดังนี้

$$Z_t = 2Z_{t-1} - Z_{t-2} + 0.9753881\varepsilon_{t-1} + \varepsilon_t$$

เมื่อ $Z_t = \log(Y_t)$

ตารางที่ 4 ค่าสัมประสิทธิ์และค่าสถิติของตัวแบบ ARIMA(2,1) เมื่อตัดค่าคงที่

Variable	Coefficient	t-statistic	Prob.
MA(1)	-0.975388	-27.01075	0.0000
R-squared		0.533355	
Durbin-Watson stat		2.064991	
Akaike information criterion		-3.041292	
Schwarz information criterion		-3.005767	
Residual sum of square		0.156733	

จากตารางที่ 4 พบว่าค่าสัมประสิทธิ์ของ MA(1) มีค่าเท่ากับ -0.975388 ซึ่งมีค่า t-statistic แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 โดยมีค่าสถิติที่สำคัญ คือค่า Akaike information criterion (AIC) เท่ากับ -3.041292 ค่า Schwarz information criterion (SIC) เท่ากับ -3.005767 และค่า Residual sum of square เท่ากับ 0.156733

3.2 แบบจำลองอารีแม็กซ์ (ARIMAX)

ตัวแบบแสดงดังนี้

$$Z_t = -0.030564 + 2Z_{t-1} - Z_{t-2} + 0.999689\varepsilon_{t-1} + 0.000926X_t + \varepsilon_t$$

เมื่อ $Z_t = \log(Y_t)$

ตารางที่ 5 ค่าสัมประสิทธิ์และค่าสถิติของตัวแบบ ARIMAX(2,1,1)

Variable	Coefficient	t-statistic	Prob.
c	-0.030564	-2.525888	0.0145
X	0.000926	2.555137	0.0134
MA(1)	-0.999689	-14.62690	0.0000
R-squared		0.548713	
Durbin-Watson stat		2.077405	
Akaike information criterion		-3.005792	
Schwarz information criterion		-2.899218	
Residual sum of square		0.151575	

จากตารางที่ 5 ค่าสัมประสิทธิ์ของ c, X, MA(1) มีค่าเท่ากับ -0.030564, 0.000926, -0.999689 ตามลำดับ ซึ่งมีค่า t-statistic แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 โดยมีค่าสถิติที่สำคัญค่า Akaike information criterion (AIC) เท่ากับ -3.005792 ค่า Schwarz information criterion (SIC) เท่ากับ -2.899218 และค่า Residual sum of square เท่ากับ 0.151575

4. การตรวจสอบความถูกต้อง (diagnostics checking)

ขั้นตอนการตรวจสอบความถูกต้องนั้นจะพิจารณาค่า Q-statistic เพื่อทดสอบคุณสมบัติการเป็น white noise ของความคลาดเคลื่อนที่ พบว่า Q-statistic ที่มีความล่าช้าของช่วงเวลาที่ 36 ของตัวแบบ ARIMA(2,1) มีค่า probability ไม่แตกต่างจากศูนย์อย่างมีนัยสำคัญที่ระดับ 0.05 แสดงว่าค่าความคลาดของตัวแบบมีลักษณะเป็น white noise หรือค่าความคลาดเคลื่อนมีการกระจายแบบปกติ (normal distribution) ค่าเฉลี่ย (mean) เท่ากับศูนย์และค่าความแปรปรวนคงที่ ซึ่งหมายความว่า ตัวแบบได้ผ่านการตรวจสอบความถูกต้องแล้วว่ามีความเหมาะสมที่จะใช้ในการพยากรณ์ต่อไป (ตารางที่ 6)

ตารางที่ 6 การตรวจสอบตัวแบบ ARIMA และ ARIMAX ของหุ้น BBL

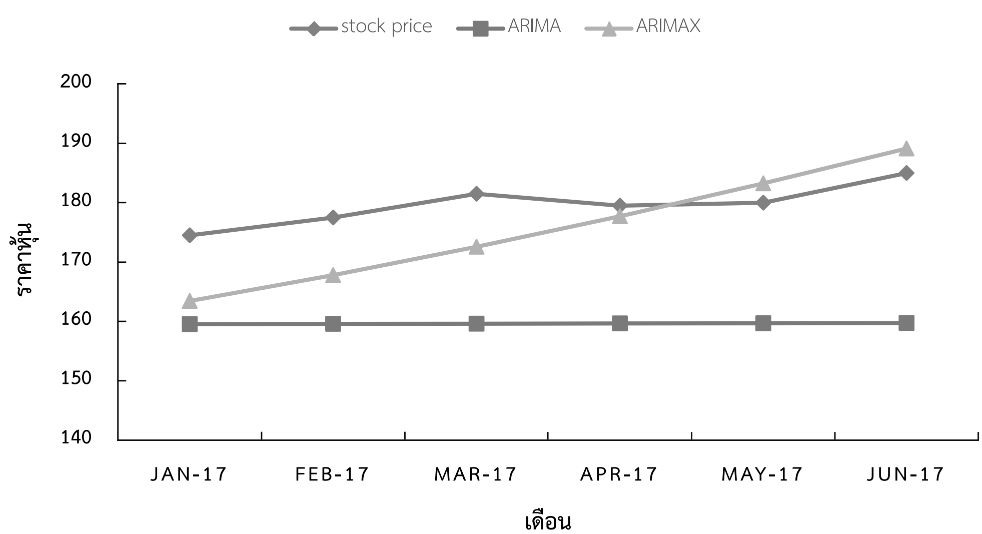
ตัวแบบที่เหมาะสม	**ค่าสถิติ**	
	Q-statistic (Lag 36)	**Probability (Lag 36)**
ARIMA (2,1)	27.919	0.797
ARIMAX (2,1,1)	24.536	0.907

5. การเปรียบเทียบประสิทธิภาพวิธีการพยากรณ์

จากการศึกษาการพยากรณ์ราคาหุ้น BBL โดยใช้ตัวแบบจำลอง 2 ตัวแบบ คือ ตัวแบบ ARIMA และตัวแบบ ARIMAX โดยมีผลการเปรียบเทียบความแม่นยำในการพยากรณ์ สำหรับการพยากรณ์ล่วงหน้า 12 เดือน ดังแสดงในตารางที่ 7

ตารางที่ 7 การเปรียบเทียบค่าพยากรณ์ราคาหุ้น BBL โดยใช้ตัวแบบ ARIMA(2,1) และ ARIMAX(2,1,1) สำหรับการพยากรณ์ล่วงหน้า 12 เดือน

เดือน	ราคาหุ้น BBL	ARIMA(2,1)	ARIMAX(2,1,1)
Jan-17	174.5	159.5371	163.4456
Feb-17	177.5	159.5743	167.7999
Mar-17	181.5	159.6114	172.5720
Apr-17	179.5	159.6486	177.7167
May-17	180.0	159.6857	183.2588
Jun-17	185.0	159.7229	189.1466
Jul-17	179.5	159.7601	195.3564
Aug-17	184.5	159.7973	201.8166
Sep-17	186.5	159.8344	208.5173
Oct-17	193.0	159.8716	215.4887
Nov-17	199.5	159.9089	222.6754
Dec-17	202.0	159.9461	230.0281
Mean Absolute Percentage Error: MAPE		13.6046	7.4111
Root Mean Squared Error: RMSE		26.7931	16.3081



ภาพที่ 1 การเปรียบเทียบค่าพยากรณ์ราคาหุ้น BBL โดยใช้ตัวแบบ ARIMA และ ARIMAX

จากตารางที่ 7 และภาพที่ 1 พบว่าตัวแบบ ARIMAX (2,1,1) สามารถพยากรณ์ราคาหุ้น BBL ได้ดีกว่าตัวแบบ ARIMA(2,1) เนื่องจากมี ค่า MAPE และ RMSE ต่ำกว่า ทำให้เป็นตัวแบบที่เหมาะสมที่สุดในการใช้พยากรณ์

อภิปรายและสรุปผลการวิจัย

ในการพยากรณ์ราคาหุ้น BBL พบว่าตัวแบบ ARIMAX มีประสิทธิภาพดีกว่าแบบจำลองอารีมา โดยเกณฑ์ที่ใช้ในการเปรียบเทียบประสิทธิภาพคือ MAPE และ RMSE โดยตัวแบบ ARIMAX ให้ค่า MAPE และ RMSE ต่ำสุด เท่ากับ 7.4111 และ 16.3081 ตามลำดับ ทั้งนี้ในงานวิจัยนี้ทำให้เห็นว่า การนำปัจจัยที่ส่งผลต่อข้อมูลที่สนใจมาร่วมในการหาตัวแบบการพยากรณ์ด้วยนั้น สามารถทำให้ค่าพยากรณ์ที่ได้มีความแม่นยำสูงขึ้น จึงสรุปได้ว่าตัวแบบที่เหมาะสมกับการพยากรณ์ราคาหุ้น BBL คือ ตัวแบบ ARIMAX โดยมีสมการการพยากรณ์ คือ

$$\hat{Z}_t = -0.030564 + 2Z_{t-1} - Z_{t-2} + 0.999689\varepsilon_{t-1} + 0.000926X_t$$

เมื่อ $\hat{Z}_t = \log(Y_t)$

จากผลการวิจัยนี้สามารถใช้วิธีการพยากรณ์ในงานวิจัยนี้มาประยุกต์ใช้กับข้อมูลที่มีความผันผวนอื่น ๆ ได้ เช่น ราคาทองคำ อัตราการแลกเปลี่ยนสกุลเงินต่างประเทศ และราคาน้ำมัน เป็นต้น นอกจากตัวแบบดังกล่าวแล้วยังสามารถนำตัวแบบอื่น ๆ มาใช้ในการพยากรณ์ข้อมูลที่มีความผันผวนได้เช่น ARIMA with GARCH, ARIMA with EGARCH และ ARIMA with GARCHM เป็นต้น

เอกสารอ้างอิง

1. ดวงธิดา ไชยวิภาสสาทร. การพยากรณ์ดัชนีราคาเหล็กโดยวิธี ARIMA. วิทยานิพนธ์ปริญญาเศรษฐศาสตรมหาบัณฑิต, บัณฑิตวิทยาลัย มหาวิทยาลัยเชียงใหม่. เชียงใหม่; 2548.
2. ยุพาพิน อติกานต์กุล. การเปรียบเทียบวิธีการพยากรณ์ดัชนีราคาเซ็ท 100 ด้วยวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลอย่างง่าย วิธีการปรับให้เรียบของโฮลท์ และวิธีบ็อกซ์และเจนกินส์. วารสารวิชาการและวิจัย มทร. พระนคร ฉบับพิเศษ 2555;1:68-73.
3. เอกชัย นิตยาเกษตรวัฒน์. การพยากรณ์ราคาทองคำด้วยวิธี ARIMA. วารสารบริหารธุรกิจ นิด้า 2553;1(7): 28-51.
4. ธาตรี จันทรโคลิกา, พอใจ เฉลิมสุข, วรพักตร์ ฐิตะดิลก, อภิญญา ภูมิชัยศักดิ์. การพยากรณ์มูลค่าการส่งออกของประเทศไทยโดยแบบจำลอง ARIMAX. RMUTT Global Business and Economics Review 2554;6(1):35-46.
5. Anggraeni W, Vinarti RX, Kurniawati YD. Performance comparisons between Arima and Arimax method in moslem kids clothes demand forecasting: case study. Procedia Comput Sci 2015;72:630-37.
6. Dickey D, Fuller W. Distribution of the estimators for autoregressive time series with a unit root. J Am Stat Assoc 2012;74(366a):427-31.
7. Dickey D, Said E. Testing for unit roots in autoregressive-moving average models of unknown order. Biometrika 1984;71(3):599-607.
8. Bowerman BL, O'Connell RT. Forecasting and time series: an applied approach. 3rd ed. California: Duxbury Press; 1993.