

การเปรียบเทียบตัวแบบการพยากรณ์ จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย Comparison of Time Series Models for Forecasting Pneumonia Cases in Thailand

วราพร ตั้วทอง และสวพร หิญาชีระนันท์*

ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ มหาวิทยาลัยนเรศวร

มหาวิทยาลัยนเรศวร ตำบลท่าโพธิ์ อำเภอเมือง จังหวัดพิษณุโลก 65000

Waraporn Thuathong and Sawaporn Hinsheranant*

Department of Mathematics, Faculty of Science,

Naresuan University, Tha Pho, Muang, Phitsanulok 65000

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบเทคนิคการพยากรณ์เพื่อหาตัวแบบที่เหมาะสมสำหรับพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย ซึ่งเป็นข้อมูลอนุกรมเวลาที่มีแนวโน้มและฤดูกาลที่ชัดเจน เทคนิคการพยากรณ์ที่ใช้ในการศึกษา 3 วิธี คือ วิธีการแยกส่วนประกอบ วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เซิง ผลคูณ และวิธีบ็อกซ์และเจนกินส์ โดยใช้ข้อมูลผู้ป่วยโรคปอดอักเสบในประเทศไทยรายไตรมาส ตั้งแต่ พ.ศ. 2551 ถึง พ.ศ. 2561 เปรียบเทียบตัวแบบพยากรณ์ที่เหมาะสมภายใต้เกณฑ์ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute percentage error, MAPE) ต่ำที่สุด ผลการศึกษาพบว่าวิธีบ็อกซ์-เจนกินส์ให้ค่า MAPE ต่ำที่สุด ตัวแบบที่เหมาะสมในการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบ คือ ตัวแบบรวมการถดถอยในตัวกับการเคลื่อนที่แบบมีฤดูกาล $SARIMA(0,1,1)(0,1,1)_4$

คำสำคัญ : อนุกรมเวลา; เทคนิคการพยากรณ์แบบมีฤดูกาล; วิธีบ็อกซ์และเจนกินส์

Abstract

The objective of this research was to compare forecasting techniques to find an appropriate model for forecasting numbers of pneumonia cases in Thailand, which consists of obvious trend and seasonality in time series. Three forecasting methods were investigated including the classical decomposition method, Winter's multiplicative method and Box-Jenkins method. The numbers of

pneumonia cases data reported quarterly from 2008 to 2018 were used. Compared the suitable forecasting model under the smallest mean absolute percentage error (MAPE) criterion. The results showed that the Box-Jenkins method gave the lowest MAPE. The appropriate model for forecasting the number of pneumonia cases in Thailand was the autoregressive integrated moving average model $SARIMA(0,1,1)(0,1,1)_4$.

Keywords: time series; seasonal forecasting technique; Box-Jenkins method

1. บทนำ

โรคปอดอักเสบหรือเรียกกันว่าปอดบวม เป็น การอักเสบของเนื้อปอด ซึ่งส่งผลให้การทำงานของ ระบบทางเดินหายใจลดลง โรคปอดอักเสบเป็นโรคที่ พบได้บ่อย โดยเฉพาะปอดอักเสบจากการติดเชื้อใน เด็กเล็ก ผู้สูงอายุ และผู้ที่มีภาวะภูมิคุ้มกันต่ำ พบได้ ประมาณร้อยละ 8-10 ของผู้ป่วยที่มีการติดเชื้อ เสียบบัณระบบหายใจ [1] ในประเทศไทยพบผู้ป่วย โรคปอดอักเสบตลอดทั้งปี โดยจำนวนผู้ป่วยจะเพิ่ม มากขึ้นในช่วงฤดูฝนและฤดูหนาว เนื่องจากเป็นช่วงที่ สภาพอากาศเปลี่ยนแปลง อากาศร้อนสลับฝนตกทำให้อากาศมีความชื้นสูง ไวรัสและแบคทีเรียเจริญเติบโต ได้ ซึ่งอาจทำให้ผู้ที่มิสุขภาพไม่แข็งแรงหรือมีภูมิ คุ้มกันต่ำ เจ็บป่วยด้วยโรคติดเชื้อทางระบบ ทางเดินหายใจได้ง่าย โดยเฉพาะโรคปอดอักเสบ [2] นอกจากนี้ยังพบว่าอัตราการป่วยของโรคปอดอักเสบมี แนวโน้มเพิ่มขึ้นอีกด้วย [3] ข้อมูลการจากเผ่าระวังทาง ระบาดวิทยาพบว่าช่วงสิบปีที่ผ่านมาอัตราการป่วยด้วย โรคปอดอักเสบเพิ่มขึ้นอย่างต่อเนื่อง จาก 221.90 ต่อแสนประชากร ในปี พ.ศ. 2551 เป็น 293.42 ต่อแสน ประชากรในปี พ.ศ. 2557 และในปี พ.ศ. 2561 อัตรา การป่วยเพิ่มขึ้นเป็น 419.92 ต่อแสนประชากร [4]

เทคนิคการพยากรณ์เป็นวิธีการทางสถิติที่นำ อนุกรมเวลามาศึกษา เพื่อหาตัวแบบที่เหมาะสมในการ อธิบายลักษณะการเคลื่อนไหวหรือความผันแปรของ

ข้อมูลในอดีต และนำตัวแบบที่ได้ไปใช้ในการพยากรณ์ ข้อมูลในอนาคตต่อไป ภายใต้ข้อสมมติว่าอนุกรมเวลา ในอนาคตมีลักษณะใกล้เคียงกับอนุกรมเวลาในอดีต [5] เทคนิคการพยากรณ์มีหลายวิธีด้วยกัน ซึ่งการเลือก เทคนิคการพยากรณ์ที่เหมาะสมขึ้นอยู่กับหลายปัจจัย เช่น ลักษณะของข้อมูลอนุกรมเวลา ซึ่งอธิบายได้ด้วย ส่วนประกอบ 4 ส่วน คือ แนวโน้ม (trend, T) การผัน แปรตามฤดูกาล (seasonal variation, S) วัฏจักร (cycle, C) และการผันแปรแบบผิดปกติหรือการแปร ผันแบบไม่สม่ำเสมอ (irregular variation, I) ช่วงเวลา ของการพยากรณ์ ได้แก่ ระยะใกล้ ระยะสั้น ระยะ กลาง และระยะยาว ความแม่นยำของการพยากรณ์จะ พิจารณาจากความคลาดเคลื่อนที่ได้จากการพยากรณ์ โดยการวัดความแม่นยำทำได้หลายวิธี ได้แก่ การใช้ค่า ส่วนเบี่ยงเบนสัมบูรณ์เฉลี่ย (mean absolute deviation) ค่าคลาดเคลื่อนกำลังสองเฉลี่ย (mean square error) และร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute percentage error) เป็นต้น โดย ตัวแบบที่มีความแม่นยำจะมีความคลาดเคลื่อนในการ พยากรณ์ต่ำ นอกจากความแม่นยำแล้ว การนำตัวแบบ ไปใช้ในการพยากรณ์ควรคำนึงถึงความเหมาะสมของ ตัวแบบภายใต้ข้อสมมติเบื้องต้นตามหลักการทางสถิติ ซึ่งตัวแบบที่เหมาะสมควรให้ความคลาดเคลื่อนที่ มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเท่ากับศูนย์ ความ แปรปรวนคงที่ และไม่มีสหสัมพันธ์ในตัวเอง

เนื่องจากข้อมูลจำนวนผู้ป่วยโรคปอดอักเสบมีแนวโน้มเพิ่มขึ้นทุกปีและมีการเปลี่ยนแปลงตามฤดูกาล ในช่วงฤดูฝนและฤดูหนาว จึงมีผู้ศึกษาตัวแบบที่เหมาะสมเพื่อใช้ในการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทยหลายงานวิจัย โดยข้อมูลที่น่าสนใจ คือ จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย จำแนกตามพื้นที่ (ข้อมูลภาพรวมของประเทศ [6] หรือ ข้อมูลรายจังหวัด [7-9]) ช่วงเวลาและลักษณะการเก็บรวบรวมข้อมูล (ข้อมูลรายเดือน [7-9] หรือข้อมูลรายไตรมาส [6]) โดยใช้เทคนิคการพยากรณ์ที่ต่างกัน เช่น วิธีแยกองค์ประกอบ (classical decomposition method) [8] วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลบวก (Winter's additive method) [6,8] วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณ (Winter's multiplicative method) [7] และวิธีบ็อกซ์และเจนกินส์ (Box-Jenkins method) [6-9] วิธีการพยากรณ์รวม (combined forecasting method) [6,8] และวิธีสัดส่วนต่อแนวโน้ม (ratio-to-trend method) [7] อย่างไรก็ตาม ไม่มีงานวิจัยใดที่ให้ผลสอดคล้องไปในทิศทางเดียวกัน ทั้งนี้เนื่องจากความผันแปรของข้อมูลที่นำมาศึกษาซึ่งมีความแตกต่างกันในด้านพื้นที่ ช่วงเวลา และลักษณะการเก็บรวบรวมข้อมูล

ดังที่กล่าวมาข้างต้น ผู้วิจัยจึงมีความสนใจเปรียบเทียบเทคนิคการพยากรณ์เพื่อหาตัวแบบที่เหมาะสมในการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบ โดยใช้ข้อมูลรายไตรมาสของจำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย ปี พ.ศ. 2551-2561 เนื่องจากข้อมูลโรคปอดอักเสบมีการเคลื่อนไหวเปลี่ยนแปลงตามฤดูกาล มีจำนวนผู้ป่วยเพิ่มขึ้นในช่วงฤดูฝนและฤดูหนาวของทุกปี ศึกษาโดยแบ่งข้อมูลเป็น 2 ส่วน คือ ส่วนที่ 1 ใช้เพื่อหาตัวแบบที่เหมาะสม และส่วนที่ 2 ใช้เพื่อทดสอบแม่นยำของการพยากรณ์ เนื่องจากข้อมูล

จำนวนผู้ป่วยโรคปอดอักเสบมีแนวโน้มและฤดูกาลที่ชัดเจน ผู้วิจัยจึงเลือกเทคนิคการพยากรณ์ 3 วิธี ได้แก่ วิธีแยกองค์ประกอบ วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณ และวิธีบ็อกซ์และเจนกินส์ มาใช้ในการหาตัวแบบเพื่อพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบ เนื่องจากทั้ง 3 วิธี สามารถใช้ได้กับอนุกรมเวลาที่มีทั้งองค์ประกอบของแนวโน้มและฤดูกาล [10] ผู้วิจัยเปรียบเทียบความแม่นยำของการพยากรณ์โดยใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (mean absolute percentage error, MAPE) โดยตัวแบบที่มีค่า MAPE ต่ำที่สุดจะเป็นตัวแบบที่มีความเหมาะสมในการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย

2. วิธีการวิจัย

2.1 ข้อมูลที่ใช้ในการวิจัย

ข้อมูลที่ใช้ในการศึกษา คือ จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทยจากระบบฐานข้อมูลด้านสังคมและคุณภาพชีวิต [11] ซึ่งเป็นข้อมูลรายไตรมาส ปี พ.ศ. 2551-2561 (รูปที่ 1) โดยแบ่งข้อมูลเป็น 2 ส่วน คือ ส่วนที่ 1 ได้แก่ ข้อมูลตั้งแต่ไตรมาสที่ 1 ปี พ.ศ. 2551 ถึงไตรมาสที่ 4 ปี พ.ศ. 2560 จำนวน 40 ไตรมาส เพื่อใช้ในการสร้างตัวแบบการพยากรณ์ และส่วนที่ 2 ได้แก่ ข้อมูลตั้งแต่ไตรมาสที่ 1 ถึงไตรมาสที่ 4 ในปี พ.ศ. 2561 จำนวน 4 ไตรมาส (ร้อยละ 9.1 ของข้อมูลทั้งหมด) เพื่อทดสอบความแม่นยำของการพยากรณ์ เนื่องจากข้อมูลแสดงให้เห็นถึงแนวโน้มและอิทธิพลของฤดูกาลที่ชัดเจน ผู้วิจัยจึงเลือกใช้เทคนิคการพยากรณ์ที่เหมาะสมกับลักษณะของข้อมูล 3 วิธี ได้แก่ วิธีแยกองค์ประกอบ วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์ และวิธีของบ็อกซ์และเจนกินส์ ซึ่งมีรายละเอียดของแต่ละวิธี ดังนี้

2.2 วิธีแยกองค์ประกอบ

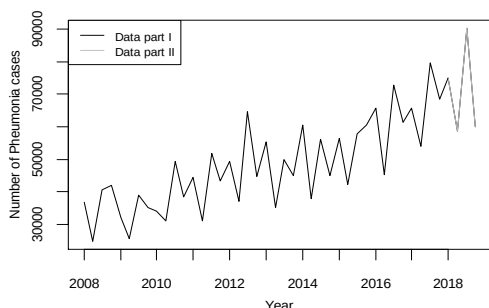


Figure 1 Numbers of pneumonia cases in Thailand during 2008-2018.

การแยกองค์ประกอบด้วยวิธีการอัตราส่วนต่อแนวโน้ม เป็นการพยากรณ์โดยการแยกส่วนประกอบของข้อมูลอนุกรมเวลาในอดีตเป็น 4 ส่วน ได้แก่ แนวโน้ม (T) การผันแปรตามฤดูกาล (S) วัฏจักร (C) และการผันแปรแบบผิดปกติ (I) โดยกำหนดตัวแบบอนุกรมเวลาดังนี้

$$Y_t = T_t \times S_t \times C_t \times I_t \quad (1)$$

เมื่อ Y_t แทน อนุกรมเวลา ณ เวลา t การพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบจะใช้การพยากรณ์ระยะสั้น ซึ่งประกอบด้วยค่าประมาณแนวโน้ม ($\hat{T}_t$) และดัชนีฤดูกาล ($\hat{S}_t$) ณ เวลา t โดยมีสมการพยากรณ์ดังนี้

$$\hat{Y}_t = \hat{T}_t \times \hat{S}_t \quad (2)$$

เมื่อ $\hat{Y}_t$ แทน ค่าพยากรณ์ของอนุกรมเวลา ณ เวลา t

การคำนวณหาอิทธิพลของแต่ละส่วนประกอบมีดังนี้

ขั้นตอนที่ 1 หารูปแบบสมการแนวโน้มที่เหมาะสม โดยใช้วิธีกำลังสองน้อยที่สุด (ordinary least square method) เพื่อคำนวณค่าประมาณแนวโน้ม ($\hat{T}_t$)

ขั้นตอนที่ 2 จัดแนวโน้มออกจากอนุกรมเวลา (Y_t) โดยหารอนุกรมเวลาด้วยค่าประมาณแนวโน้มจากขั้นตอนที่ 1

ขั้นตอนที่ 3 คำนวณดัชนีฤดูกาล ($\hat{S}_t$) โดยนำค่าที่ได้จากขั้นตอนที่ 2 มาคำนวณหาค่ามัธยฐานของแต่ละไตรมาส แล้วนำปรับค่าให้ผลรวมของแต่ละไตรมาสเท่ากับขนาดฤดูกาล

ขั้นตอนที่ 4 คำนวณค่าพยากรณ์จากค่าประมาณแนวโน้มและดัชนีฤดูกาล ณ เวลา t ตามสมการ (2)

2.3 วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์

วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์ เป็นวิธีที่เหมาะสมสำหรับอนุกรมเวลาที่แสดงให้เห็นถึงการเคลื่อนไหวของแนวโน้มและฤดูกาล โดยที่แนวโน้มมีลักษณะเป็นเชิงเส้นตรง วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เขียนได้ 2 รูปแบบ คือ ตัวแบบเชิงบวก (additive model) และตัวแบบเชิงคูณ (multiplicative model) โดยตัวแบบเชิงบวกเหมาะสำหรับอนุกรมเวลาที่ความผันแปรของฤดูกาลคงที่ มีการเคลื่อนไหวในลักษณะเดิมไม่เปลี่ยนแปลง ส่วนตัวแบบเชิงคูณเหมาะสำหรับอนุกรมเวลาที่ความผันแปรของฤดูกาลไม่คงที่ มีการเคลื่อนไหวเพิ่มขึ้นหรือลดลงตามเวลาที่เปลี่ยนแปลงไป [5] การศึกษาครั้งนี้ผู้วิจัยเลือกใช้ตัวแบบเชิงคูณ เนื่องจากความผันแปรของฤดูกาลไม่คงที่ ซึ่งเขียนตัวแบบเชิงคูณและสมการพยากรณ์ได้ดังนี้

$$Y_t = (\beta_0 + \beta_1 t) S_t \varepsilon_t \quad (3)$$

$$\hat{Y}_{t+m} = (a_t + b_t m) \hat{S}_{t+m-s} \quad (4)$$

เมื่อ Y_t และ $\hat{Y}_t$ แทนอนุกรมเวลาและค่าพยากรณ์อนุกรมเวลา ณ เวลา t ; β_0, β_1 และ S_t แทนพารามิเตอร์แสดงค่าระดับ ความชันของแนวโน้ม และความผันแปรของฤดูกาล ณ เวลา t ; a_t, b_t และ $\hat{S}_t$ แทนค่าประมาณของพารามิเตอร์ β_0, β_1 และ S_t ; m และ s แทน ช่วงเวลาที่ต้องการพยากรณ์ล่วงหน้า และคาบของฤดูกาล โดยในที่นี้กำหนด $s = 4$ โดยที่

$$a_t = \alpha \frac{Y_t}{\hat{S}_{t-s}} + (1-\alpha)(a_{t-1} + b_{t-1}) \quad (5)$$

$$b_t = \gamma(a_t + a_{t-1}) + (1-\gamma)b_{t-1} \quad (6)$$

$$\hat{S}_t = \delta \frac{Y_t}{a_{t-1}} + (1-\delta)\hat{S}_{t-s} \quad (7)$$

เมื่อ α , γ และ δ แทน ค่าคงที่ของการปรับให้เรียบระดับ ค่าแนวโน้ม และค่าฤดูกาล ตามลำดับ โดยที่ $0 < \alpha, \gamma, \delta < 1$ พิจารณาเลือกค่า α , γ และ δ ที่ทำให้ความคลาดเคลื่อนของการพยากรณ์มีค่าต่ำที่สุด จากนั้นคำนวณค่าต่าง ๆ ณ เวลา t ตามสมการ (5) ถึง (7) และค่าพยากรณ์ตามสมการ (4) [12]

2.4 วิธีบ็อกซ์และเจนกินส์

วิธีบ็อกซ์และเจนกินส์เป็นวิธีหนึ่งที่ได้รับ ความนิยม การหาตัวแบบการพยากรณ์จะพิจารณาจากฟังก์ชันสหสัมพันธ์ในตัวเอง (autocorrelation function, ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (partial autocorrelation function, PACF) ของอนุกรมเวลาที่มีความเสถียร ซึ่งมีค่าเฉลี่ยและความแปรปรวนคงที่ วิธีของบ็อกซ์และเจนกินส์ใช้ได้กับทั้งอนุกรมเวลาไม่มีอิทธิพลของฤดูกาลและมีอิทธิพลของฤดูกาล โดยตัวแบบที่มีอิทธิพลของฤดูกาลเขียนแทนด้วย SARIMA(p, d, q)(P, D, Q)_s มีรูปแบบทั่วไปดังนี้

$$\Phi_s(B^s)\phi(B)(1-B)^d(1-B^s)^pY_t = \Theta_s(B^s)\theta(B)\varepsilon_t \quad (8)$$

เมื่อ B แทนตัวดำเนินการถอยหลัง (backshift operator) โดยที่ $BY_t = Y_{t-1}$ และ $B^sY_t = Y_{t-s}$; $\phi(B)$ แทนตัวดำเนินการถดถอยในตัวแบบที่ไม่มีฤดูกาลอันดับที่ p หรือ AR(p) โดยที่ $\phi(B) = 1 - \phi_1B - \dots - \phi_pB^p$; $\Phi_s(B^s)$ แทนตัวดำเนินการถดถอยในตัวแบบที่มีฤดูกาลอันดับที่ P หรือ SAR(P) โดยที่ $\Phi_s(B^s) = 1 - \Phi_1B^s - \dots - \Phi_PB^{Ps}$; $\theta(B)$ แทนตัวดำเนินการเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับที่ q หรือ MA(q) โดยที่ $\theta(B) = 1 - \theta_1B - \dots - \theta_qB^q$; $\Theta_s(B^s)$ แทนตัว

ดำเนินการเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับที่ Q หรือ SMA(Q) โดยที่ $\Theta_s(B^s) = 1 - \Theta_1B^s - \dots - \Theta_QB^{Qs}$; s แทนคาบของฤดูกาล; d และ D แทนลำดับการหาผลต่างและผลต่างฤดูกาลของอนุกรมเวลา

การกำหนดตัวแบบที่เหมาะสมด้วยวิธีบ็อกซ์และเจนกินส์ดังนี้

ขั้นตอนที่ 1 ตรวจสอบความเสถียรของอนุกรมเวลา โดยอนุกรมเวลาจะต้องมีค่าเฉลี่ยและความแปรปรวนคงที่ ถ้าอนุกรมเวลาไม่มีความแปรปรวนไม่คงที่ จะต้องแปลงข้อมูลด้วยวิธีที่เหมาะสม เพื่อให้อนุกรมเวลาไม่มีความแปรปรวนคงที่ และหากอนุกรมเวลามีค่าเฉลี่ยไม่คงที่ แสดงให้เห็นถึงแนวโน้ม จะต้องหาผลต่างลำดับที่ d และ/หรือผลต่างลำดับที่ D สำหรับอนุกรมเวลาที่มีฤดูกาล เพื่อให้อนุกรมเวลาไม่มีความเสถียรต่อค่าเฉลี่ย

ขั้นตอนที่ 2 เมื่ออนุกรมเวลาไม่มีความเสถียรแล้ว กำหนดตัวแบบที่เหมาะสมโดยพิจารณาจากกราฟฟังก์ชันสหสัมพันธ์ในตัวเอง (autocorrelation function, ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (partial autocorrelation function, PACF) หลังจากระบุตัวแบบที่ใช้สำหรับอนุกรมเวลาแล้ว ประมาณค่าพารามิเตอร์ของตัวแบบที่ได้

ขั้นตอนที่ 3 ตรวจสอบความเหมาะสมของตัวแบบด้วยการทดสอบสมมติฐานเกี่ยวกับพารามิเตอร์ในตัวแบบ โดยค่าพารามิเตอร์ทุกตัวจะต้องมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 และทดสอบภาพรวมของส่วนเหลือ (residual) ด้วยกราฟต่าง ๆ และทดสอบความคลาดเคลื่อนสุ่ม (random error) ด้วยสถิติทดสอบของ Ljung-Box Q ถ้าทดสอบแล้วพบว่าตัวแบบที่กำหนดยังไม่มีความเหมาะสมจะต้องเริ่มขั้นตอนที่ 2 ใหม่อีกครั้ง

ขั้นตอนที่ 4 คำนวณค่าพยากรณ์ของตัวแบบที่ได้

2.5 การตรวจสอบความเหมาะสมและเปรียบเทียบประสิทธิภาพของตัวแบบ

เมื่อได้ตัวแบบของแต่ละวิธีแล้วจะตรวจสอบความเหมาะสมของตัวแบบที่ได้โดยพิจารณาจากกราฟของส่วนเหลือในรูปแบบต่าง ๆ ได้แก่ (1) กราฟตรวจสอบการแจกแจงปกติ (normal QQ plot) (2) กราฟของส่วนเหลือกับช่วงเวลา (residual plot) เพื่อตรวจสอบการกระจายและค่าเฉลี่ยของส่วนเหลือ (3) กราฟของส่วนเหลือกับค่าพยากรณ์ (plot of residual against fitted value) เพื่อตรวจสอบความคงที่ของความแปรปรวน (4) กราฟ ACF (5) กราฟ PACF ของส่วนเหลือ เพื่อตรวจสอบความเป็นอิสระกันของส่วนเหลือ และ (6) กราฟของค่า p-value ของสถิติทดสอบ Ljung-Box Q เพื่อทดสอบสหสัมพันธ์ในตัวของความคลาดเคลื่อนสุ่มที่ช่วงเวลาที่ห่างออกไป ถ้าค่า p-value ณ ช่วงเวลาใด ๆ มากกว่าระดับนัยสำคัญ 0.05 (หรืออยู่เลยเส้นประ) แสดงว่าสหสัมพันธ์ในตัวของความคลาดเคลื่อนสุ่มมีค่าไม่แตกต่างจากศูนย์ที่ระดับนัยสำคัญ 0.05 นั่นคือความคลาดเคลื่อนอย่างสุ่มไม่มีสหสัมพันธ์ในตัวหรือเป็นอิสระกันนั่นเอง [5,10] เมื่อได้ตัวแบบที่เหมาะสมแล้ว คำนวณค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) จากสมการดังนี้

$$MAPE = \frac{\sum_{t=1}^n |e_t/Y_t|}{n} \times 100$$

(9)

เมื่อ e_t แทนส่วนเหลือที่ได้จากการพยากรณ์ ($e_t = Y_t - \hat{Y}_t$); Y_t และ $\hat{Y}_t$ แทนอนุกรมเวลาและค่าพยากรณ์อนุกรมเวลา ณ เวลา t ; n แทนจำนวนค่าพยากรณ์

จากนั้นเปรียบเทียบความแม่นยำของการพยากรณ์ของตัวแบบโดยพิจารณาจากค่า MAPE โดยตัวแบบที่มีความเหมาะสมคือตัวแบบที่มีค่า MAPE ต่ำที่สุด

2.6 โปรแกรมที่ใช้ในการวิจัย

การศึกษาค้นคว้าครั้งนี้ ผู้วิจัยใช้โปรแกรม R โดยใช้แพ็คเกจ forecast ในการประมวลผลและวิเคราะห์ข้อมูล [13]

3. ผลการวิจัย

3.1 การพยากรณ์ด้วยวิธีแยกองค์ประกอบ

พบว่าวิธีกำลังสองน้อยที่สุดได้สมการแนวโน้มที่เหมาะสมกับอนุกรมเวลาในช่วงนี้ คือ $\hat{T}_t = 29790.104 + 877.045t$ เมื่อ $t = 1, \dots, 40$ โดยมีค่าสัมประสิทธิ์การตัดสินใจ (R^2) ของสมการแนวโน้ม 0.772 นำค่าประมาณแนวโน้มไปหาค่าดัชนีฤดูกาลของแต่ละไตรมาส (ตารางที่ 1) แล้วคำนวณค่าพยากรณ์ระยะสั้นโดยใช้ค่าประมาณแนวโน้มและค่าดัชนีฤดูกาลที่ได้ตามสมการที่ (2) ได้ค่าพยากรณ์ดังแสดงในรูปที่ 2

Table 1 Seasonal index from the classical decomposition method

Quarters	Unadjusted seasonal index	Adjusted seasonal index ($\hat{S}_t$)
Quarter 1 (Jan - Mar)	1.092	1.085
Quarter 2 (Apr - Jun)	0.755	0.751

Quarter 3 (Jul - Sep)	1.206	1.198
Quarter 4 (Oct - Dec)	0.972	0.966
Total	4.025	4.000

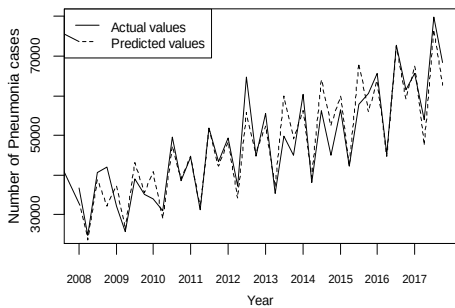


Figure 2 Actual and predicted values from the classical decomposition method

เมื่อตรวจสอบส่วนเหลือ (รูปที่ 3) โดยพิจารณากราฟของส่วนเหลือกับช่วงเวลา (residual

plot) พบว่าส่วนเหลือมีการกระจายอย่างสม่ำเสมอค่า ศูนย์และความแปรปรวนคงที่ และค่า p-value ของ สถิติทดสอบ Ljung-Box Q มากกว่าระดับนัยสำคัญ 0.05 แสดงว่าความคลาดเคลื่อนสุ่มเป็นอิสระกันและ ไม่มีสหสัมพันธ์ในตัว อย่างไรก็ตาม เมื่อพิจารณาจาก กราฟตรวจสอบการแจกแจงปกติ (normal QQ plot) พบว่าส่วนเหลือส่วนใหญ่เกาะอยู่บนเส้นตรง และมีบางส่วนที่ปลายหางทั้งสองข้างเบี่ยงเบนไปจาก เส้นตรง แสดงว่าส่วนเหลือที่ได้ อาจมีการแจกแจงที่ เบี่ยงเบนไปจากการแจกแจงปกติเล็กน้อย ดังนั้น ตัวแบบที่ได้จึงอาจไม่มีความเหมาะสม คำนวณค่า MAPE ได้ 7.096

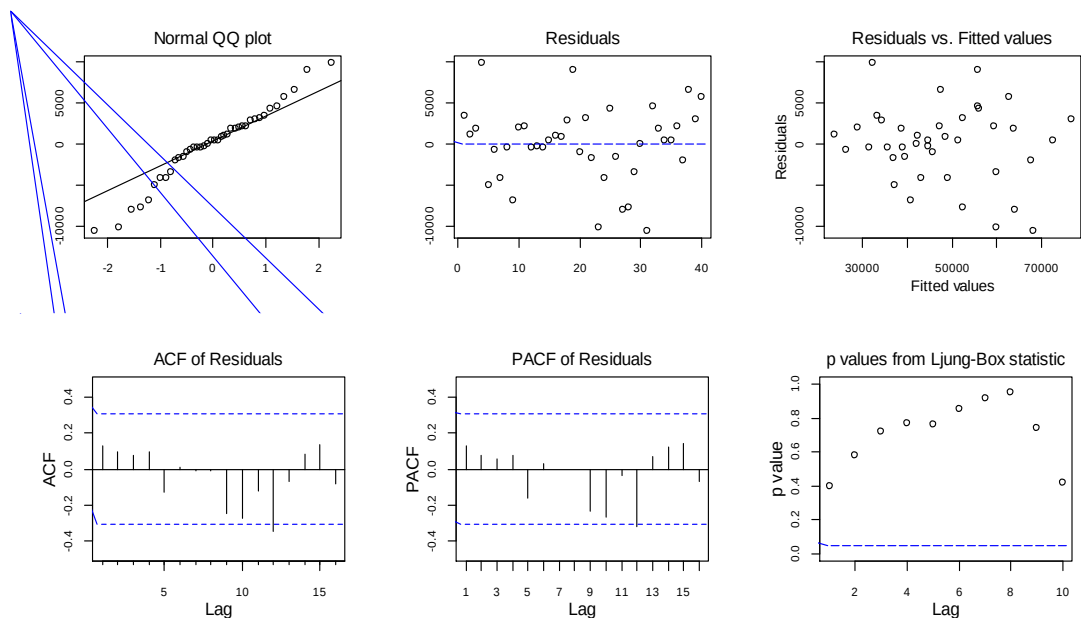


Figure 3 Residual plots of the classical decomposition method

3.2 การพยากรณ์ด้วยวิธีปรับเรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์

วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณจะต้องใช้ค่าพารามิเตอร์ในการปรับให้เรียบ 3 ค่า คือ ค่าระดับ (α) ค่าแนวโน้ม (γ) และค่าฤดูกาล (δ) โดยค่าของการปรับให้เรียบที่เหมาะสมคือ $\alpha = 0.2421$, $\gamma = 0.0004$ และ $\delta = 0.0001$ เนื่องจากให้ค่าความคลาดเคลื่อนต่ำที่สุด นำค่าปรับให้เรียบที่ได้คำนวณค่าต่าง ๆ ใน ณ ช่วงเวลา t ตามสมการที่ (5) ถึง (7) จากนั้นคำนวณค่าพยากรณ์ตามสมการที่ (4) ได้ค่าพยากรณ์ดังแสดงในรูปที่ 4

ตรวจสอบส่วนเหลือ (Figure 5) พบว่าส่วนเหลือมีการแจกแจงใกล้เคียงการแจกแจงปกติ มีค่าเฉลี่ยรอบค่าศูนย์และความแปรปรวนคงที่ และค่า p

value ของสถิติทดสอบ Ljung-Box Q มากกว่าระดับนัยสำคัญ 0.05 แสดงว่า ความคลาดเคลื่อนสุ่มเป็นอิสระกันและไม่มีสหสัมพันธ์ในตัว ดังนั้นตัวแบบที่ได้จึงมีความเหมาะสม ค่าวนควค่า MAPE ได้เท่ากับ 7.822

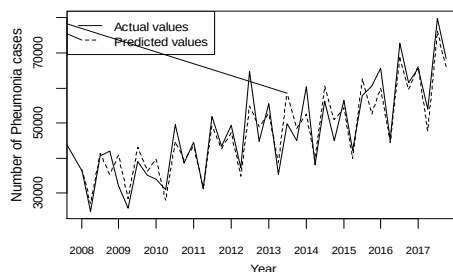


Figure 4 Actual and predicted values from Winter's multiplicative method.

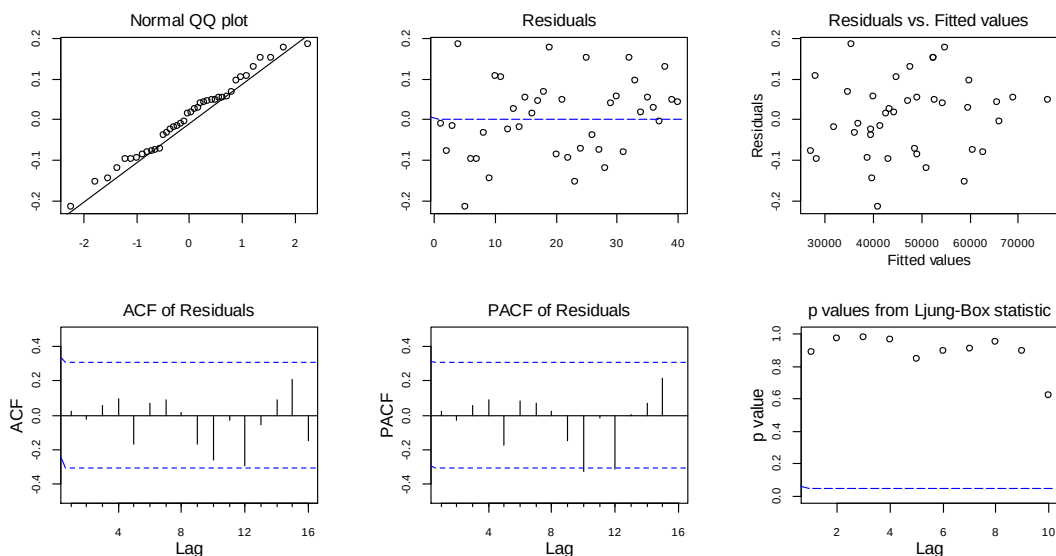


Figure 5 Residual plots of Winter's multiplicative method.

3.3 การพยากรณ์ด้วยวิธีของบ็อกซ์และเจนกินส์

เนื่องจากอนุกรมเวลาในรูปที่ 1 ยังไม่มีความเสถียรต่อค่าเฉลี่ยและความแปรปรวน และแสดงให้เห็นถึงอิทธิพลของฤดูกาลที่ชัดเจน ดังนั้นจึงต้อง

แปลงข้อมูลด้วยรากที่สอง (square root transformation) เพื่อให้อนุกรมเวลามีความแปรปรวนคงที่ จากนั้นจึงผลต่างลำดับที่ 1 เพื่อขจัดแนวโน้มและหาผลต่างลำดับที่ 4 เพื่อกำจัดอิทธิพลของฤดูกาล รูปที่ 6 แสดงกราฟของอนุกรมผลต่างลำดับที่ 1 และลำดับที่ 4

และกราฟ ACF และ PACF ของอนุกรมผลต่างลำดับที่ 1 และลำดับที่ 4 ของอนุกรมเวลาที่แปลงโดยใช้รากที่สอง จากกราฟจะเห็นได้ว่าอนุกรมเวลาที่มีความเสถียรแล้ว จึงกำหนดพารามิเตอร์ของตัวแบบโดยพิจารณาจากกราฟ ACF และ PACF ซึ่งตัวแบบที่เป็นไปได้ 3

ตัวแบบ คือ SARIMA(1,1,1)(1,1,1)₄, SARIMA(0,1,1)(1,1,0)₄ และ SARIMA(0,1,1)(0,1,1)₄ ประมาณค่าและทดสอบค่าพารามิเตอร์ในแต่ละตัวแบบในตารางที่ 2

Table 2 Parameter estimation of SARIMA models

Models	Parameters	Estimate	p-values	MAPE
SARIMA(1,1,1)(1,1,1) ₄	AR(1)	0.3118	0.120	3.939
	MA(1)	-0.9732	<0.001	
	SAR(1)	-0.7333	0.014	
	SMA(1)	0.4513	0.225	
SARIMA(0,1,1)(1,1,0) ₄	MA(1)	-0.6403	<0.001	3.935
	SAR(1)	-0.4504	0.006	
SARIMA(0,1,1)(0,1,1) ₄	MA(1)	-0.7179	<0.001	3.919
	SMA(1)	-0.7537	0.026	

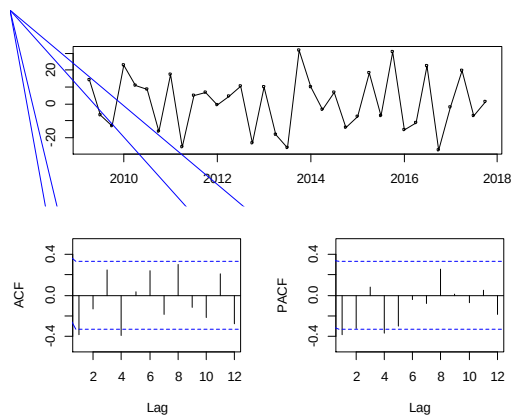


Figure 6 Transformed data after differencing at lags 1 and 4, ACF and PACF plots of transformed data after differencing at lags 1 and 4

ตารางที่ 2 จะเห็นได้ว่าพารามิเตอร์บางตัวของตัวแบบ SARIMA(1,1,1)(1,1,1)₄ มีค่าไม่แตกต่างจากศูนย์ที่ระดับนัยสำคัญ 0.05 ดังนั้นตัวแบบ

ดังกล่าวจึงไม่มีความเหมาะสม และเมื่อพิจารณาตัวแบบ SARIMA(0,1,1)(1,1,0)₄ และ SARIMA(0,1,1)(0,1,1)₄ พบว่าพารามิเตอร์ทุกตัวของตัวแบบมีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับนัยสำคัญ 0.05 แต่ตัวแบบ SARIMA(0,1,1)(0,1,1)₄ มีค่า MAPE ต่ำที่สุด ดังนั้นผู้วิจัยจึงเลือกตัวแบบ SARIMA(0,1,1)(0,1,1)₄ เป็นตัวแบบที่เหมาะสมของวิธีบ็อกซ์และเจนกินส์ ซึ่งสมการพยากรณ์ คือ

$$\hat{X}_t = X_{t-1} + X_{t-4} - X_{t-5} + 0.7179e_{t-1} + 0.7537e_{t-4} + 0.5411e_{t-5} \quad (10)$$

เมื่อ X_t แทนอนุกรมเวลาที่แปลงข้อมูลโดยใช้รากที่สอง ($X_t = \sqrt{Y_t}$); $\hat{Y}_t = \hat{X}_t^2$ แทนค่าพยากรณ์ ณ ช่วงเวลา t โดยค่าพยากรณ์ที่คำนวณได้แสดงในรูปแบบที่ 7

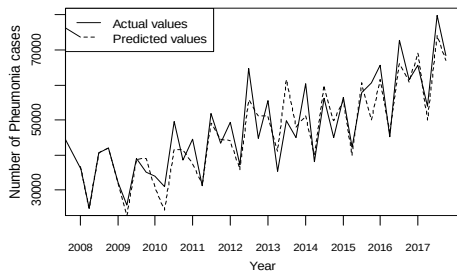


Figure 7 Actual and predicted values from Box-Jenkins method

การตรวจสอบส่วนเหลือ (รูปที่ 8) พบว่า ส่วนเหลือมีการแจกแจงใกล้เคียงการแจกแจงปกติ มีค่าเฉลี่ยรอบค่าศูนย์และความแปรปรวนคงที่ และค่า p-value ของสถิติทดสอบ Ljung-Box Q มากกว่าระดับนัยสำคัญ 0.05 แสดงว่าความคลาดเคลื่อนสุ่มเป็นอิสระกันและไม่มีสหสัมพันธ์ในตัว ดังนั้นตัวแบบที่ได้จึงมีความเหมาะสม คำนวนค่า MAPE ได้ 3.919

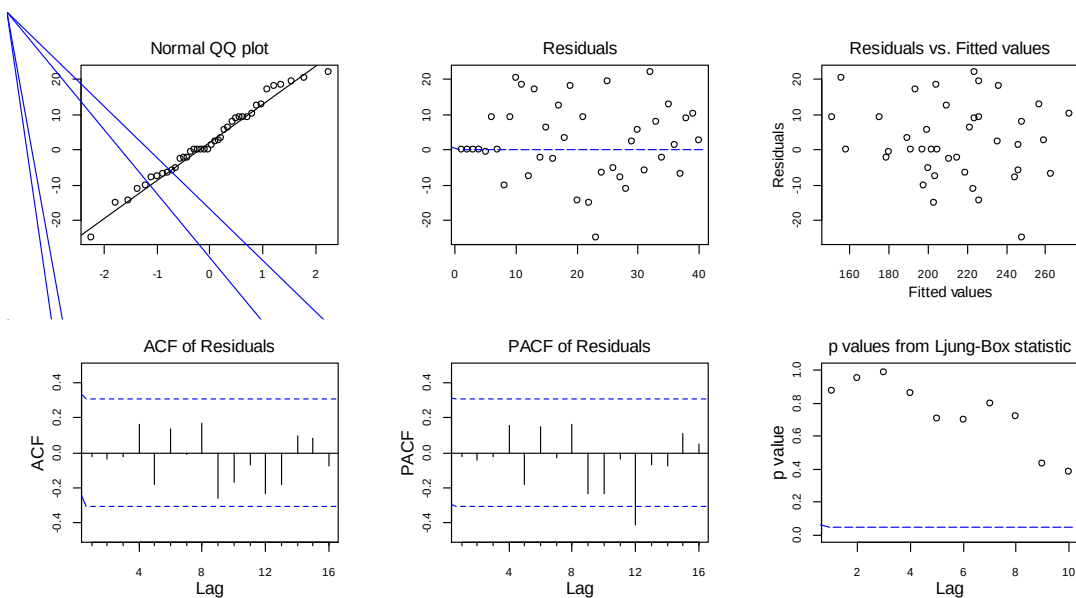


Figure 8 Residual plots of Box-Jenkins method

Table 3 Predicted values of pneumonia cases in 2018

Year 2018	Actual values	Predicted values		
		Decomposition method	Winter's multiplicative method	Box-Jenkins method
Quarter 1	74,881	71,338	72,688	75,024
Quarter 2	58,699	50,036	52,623	57,509
Quarter 3	90,160	80,869	81,279	82,339
Quarter 4	59,861	66,055	69,251	72,708

	MAPE	10.036	9.704	8.088
--	------	--------	-------	-------

3.4 การเปรียบเทียบความแม่นยำของการพยากรณ์

เมื่อได้ตัวแบบที่เหมาะสมของแต่ละวิธี นำตัวแบบที่ได้มาพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในไตรมาสที่ 1 ถึงไตรมาสที่ 4 ปี พ.ศ. 2561 จำนวนค่าพยากรณ์ และค่า MAPE ได้ผลดังตารางที่ 3 พบว่าตัวแบบจากทั้ง 3 วิธี ให้ค่าพยากรณ์ที่สอดคล้องไปในทิศทางเดียวกัน วิธีของบ็อกซ์และเจนกินส์ให้ค่าพยากรณ์ในไตรมาสที่ 1 และไตรมาสที่ 2 ใกล้เคียงกับค่าจริงมากที่สุด (รูปที่ 9) โดยวิธีที่ให้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ต่ำที่สุด คือ วิธีบ็อกซ์และเจนกินส์ ซึ่งให้ค่า MAPE = 8.088 ดังนั้นตัวแบบที่เหมาะสมที่จะใช้ในการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในปี พ.ศ. 2561 คือ ตัวแบบ SARIMA(0,1,1)(0,1,1)₄

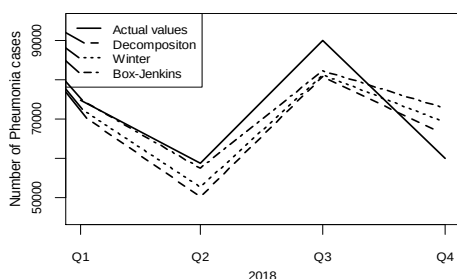


Figure 9 Actual values and predicted values of pneumonia cases in 2018

4. วิจารณ์ผลการวิจัย

การหาตัวแบบที่เหมาะสมเพื่อพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบ ผู้วิจัยได้แบ่งข้อมูลเป็น 2 ส่วน ข้อมูลส่วนแรก จำนวน 40 ไตรมาส ใช้ในการหาตัวแบบที่เหมาะสมของแต่ละวิธี และข้อมูลส่วนที่สอง

จำนวน 4 ไตรมาส ใช้ในการทดสอบความแม่นยำของการพยากรณ์ของตัวแบบที่ได้ การใช้ข้อมูลในส่วนที่ 1 เพื่อหาตัวแบบที่เหมาะสมของแต่ละวิธี พบว่าตัวแบบที่มีประสิทธิภาพดีที่สุด คือ ตัวแบบของวิธีบ็อกซ์และเจนกินส์ ซึ่งให้ค่า MAPE ต่ำที่สุด ส่วนตัวแบบของวิธีแยกองค์ประกอบและวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณมีประสิทธิภาพใกล้เคียงกัน โดยมีค่า MAPE ไม่แตกต่างกันมากนัก เมื่อทดสอบความแม่นยำของการพยากรณ์ของตัวแบบที่ได้โดยใช้ข้อมูลในส่วนที่ 2 พบว่าตัวแบบของวิธีบ็อกซ์และเจนกินส์ยังคงให้ค่าพยากรณ์ในไตรมาสที่ 1 และ 2 ใกล้เคียงกับค่าจริงมากที่สุด ทำให้มีค่า MAPE ต่ำที่สุดเมื่อเปรียบเทียบกับอีก 2 วิธี

ผลการศึกษาจากทั้งสองส่วน พบว่าวิธีบ็อกซ์และเจนกินส์มีประสิทธิภาพในการพยากรณ์ดีกว่าวิธีแยกองค์ประกอบและวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณ ทั้งนี้อาจเป็นเพราะวิธีแยกองค์ประกอบและวิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณพิจารณาองค์ประกอบแนวโน้มในรูปแบบเชิงเส้นตรง (linear trend) ดังนั้นวิธีดังกล่าวอาจไม่มีประสิทธิภาพเท่าที่ควรถ้าอนุกรมเวลาไม่ได้มีองค์ประกอบของแนวโน้มเชิงเส้นตรงที่ชัดเจน ซึ่งรูปที่ 1 จะเห็นได้ว่าภาพรวมการเคลื่อนไหวของแนวโน้มเพิ่มขึ้นคงที่ในช่วงปี พ.ศ. 2551-2558 แต่ในช่วงปี พ.ศ. 2559-2561 แสดงให้เห็นถึงอัตราการเพิ่มขึ้นของแนวโน้ม ซึ่งอาจส่งผลต่อการประมาณค่าแนวโน้มเชิงเส้นตรงในภาพรวม ในทางกลับกัน การกำหนดตัวแบบที่เหมาะสมของวิธีบ็อกซ์และเจนกินส์จะกำหนดจากกระบวนการที่มีความเสถียร กล่าวคือ มีค่าเฉลี่ยและ

ความแปรปรวนคงที่ ดังนั้นองค์ประกอบแนวโน้มของอนุกรมเวลาจะถูกกำจัดออกไปโดยหาผลต่างลำดับที่ 1 ทำให้แนวโน้มไม่ส่งผลต่อการกำหนดพารามิเตอร์ในตัวแบบ ดังนั้นการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบนั้น วิธีบอกซ์และเจนกินส์ให้ตัวแบบที่เหมาะสมกว่าวิธีวิธีแยกองค์ประกอบและวิธีการปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณ เนื่องจากวิธีบอกซ์และเจนกินส์สามารถประยุกต์ใช้กับอนุกรมเวลาที่มีแนวโน้มในรูปแบบต่าง ๆ ดีกว่า และพบว่าวิธีบอกซ์และเจนกินส์นำไปใช้ในการเฝ้าระวังทางระบาดวิทยาในหลายงานวิจัย [13-15] โดย Ruchiraset และ Tantrakarnapa ใช้ตัวแบบรวมการถดถอยในตัวกับการเฉลี่ยเคลื่อนที่แบบมีฤดูกาล (autoregressive integrated moving average model) ร่วมกับตัวแปรนอกระบบ (exogenous variable) เพื่อให้การพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในจังหวัดเชียงใหม่ มีประสิทธิภาพยิ่งขึ้น [9]

5. สรุป

การศึกษาในครั้งนี้เป็นการศึกษาตัวแบบที่เหมาะสมเพื่อใช้ในการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในประเทศไทย โดยใช้เทคนิคการพยากรณ์ทั้งหมด 3 วิธี ได้แก่ วิธีแยกองค์ประกอบ วิธีปรับให้เรียบแบบเอ็กซ์โปเนนเชียลของวินเตอร์เชิงผลคูณ และวิธีบอกซ์และเจนกินส์ ข้อมูลที่ใช้ คือ ข้อมูลจำนวนผู้ป่วยโรคปอดอักเสบรายไตรมาส ปี พ.ศ. 2551-2561 และใช้ค่าร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) ต่ำที่สุดเป็นเกณฑ์ในการเปรียบเทียบความแม่นยำของการพยากรณ์ของตัวแบบ ผลการศึกษาพบว่าวิธีที่ให้ค่า MAPE ต่ำที่สุด คือ วิธีบอกซ์และเจนกินส์ โดยตัวแบบที่เหมาะสมในการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในปี พ.ศ. 2561 คือ ตัวแบบ SARIMA(0,1,1)(0,1,1)₄ ตัวแบบที่ได้สามารถ

นำไปพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบในอนาคต เพื่อให้ผู้ที่เกี่ยวข้องสามารถใช้เป็นข้อมูลในการจัดการหรือวางแผนงานด้านการสาธารณสุขต่อไป อย่างไรก็ตามเมื่อมีข้อมูลปัจจุบันเพิ่มขึ้น ผู้วิจัยควรปรับปรุงและตรวจสอบความเหมาะสมของตัวแบบ เพื่อให้ตัวแบบสอดคล้องกับลักษณะของข้อมูลปัจจุบัน และส่งผลให้การพยากรณ์มีประสิทธิภาพยิ่งขึ้น

6. References

- [1] Bureau of Epidemiology, Pnuemonia, Retrieved, Available Source: <http://www.boe.moph.go.th/boedb/surdata/disease.php?dcontent= def&ds= 31>, September 8, 2018. (in Thai)
- [2] Bureau of Epidemiology, 2019, Pneumonia Situation, Department of Disease Control, Nonthaburi. (in Thai)
- [3] WESR, 2019, Disease situation and fatality from pneumonia in Chaiyaphum province, 2014-2018, Weekly Epidemiol. Surveil. Rep. 50(10): 149-156. (in Thai)
- [4] Bureau of Epidemiology, 2019, National Disease Surveillance, Department of Disease Control of Thailand, Nonthaburi. (in Thai)
- [5] Manmin, M., 2006, Time Series and Forecasting, Prakaypruek Publishing Center, Bangkok. (in Thai)
- [6] Keerativibool, W., 2016, Forecasting model for the number of patients with pneumonia in Thailand, Public Health J. Burapha Univ. 11(1): 24-38. (in Thai)

- [7] Sanguanrungrasirikul, D., Chiewananta vanich, H. and Sangkasem, M., 2015, A comparative study to determine optimal models for forecasting the number of patients having epidemiological-surveillance diseases in Bangkok, KMUTT Res. Develop. J. 38(1): 35-55. (in Thai)
- [8] Nakunthod, I. and Khamkhod, K., 2018, Time series modeling for the pneumonia rate of patients in Lampang province, pp. 340- 358, Proceeding in 18th Graduate Studies of Northern Rajabhat University Network Conference and 4th Lampang Research, Lampang Rajabhat University, Lampang. (in Thai)
- [9] Ruchiraset, A. and Tantrakarnapa, K., 2018, Time series modeling of pneumonia admissions and its association with air pollution and climate variables in Chiang Mai province, Thailand, Environ. Sci. Pollut. Res. 25: 33277-33285.
- [10] Bowerman, B.L., O' Connell, R.T. and Koehler, A.B., 2005, Forecasting, Time Series, and Regression: An Applied Approach, Thomson Brooks/ Cole, Belmont.
- [11] Social and Quality of Life Database System, Number of Cases and Morbidity Rates of Disease Surveillance, Available Source: <http://social.nesdb.go.th/social/Default.aspx?tabid=40>, September 8, 2018.
- [12] Shumway, R.H. and Stoffer, D.S., 2006, Time Series Analysis and Its Applications: With R Examples, Springer, New York.
- [13] R Core Team, 2018, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna.
- [14] Marie, G.C., 2007, Use of ARIMA models for communicable disease surveillance, Revista Cubana de Salud Publica. 33(2).
- [15] Unkel, S., Farrington, C.P., Garthwaite, P.H., Robertson, C. and Andrews, N., 2012, Statistical methods for the prospective detection of infectious disease outbreaks: A review, J. R. Stat. Soc. Ser. A 175: 49-82.
- [16] Zhang, X., Zhang, T., Young, A.A. and Li, X., 2014, Applications and comparisons of four time series models in epidemiological surveillance data, PLoS ONE 9(2): e88075.