

ตัวแบบพยากรณ์รายได้รัฐบาลของประเทศไทยหลังจากฟื้นตัวของสถานการณ์  
การแพร่ระบาดโควิด-19  
Forecasting Model for Thailand's Government Revenue During the  
Recovery of the COVID-19 Situation

วฐา มินเสน<sup>1\*</sup> และ พรารถนา มินเสน<sup>2</sup>

Watha Minsan<sup>1\*</sup> and Pradthana Minsan<sup>2</sup>

<sup>1</sup>ศูนย์วิจัยวิทยาการข้อมูล ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่ จ.เชียงใหม่ ประเทศไทย

<sup>2</sup>ภาควิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏเชียงใหม่ จ.เชียงใหม่ ประเทศไทย

<sup>1</sup>Data Science Research Center, Department of Statistics, Faculty of Science, Chiang Mai University,  
Chiang Mai, Thailand

<sup>2</sup>Department of Mathematics and Statistics, Faculty of Science and Technology,  
Chiang Mai Rajabhat University, Chiang Mai, Thailand

วันที่ส่งบทความ : 22 มิถุนายน 2566 วันที่แก้ไขบทความ : 20 มีนาคม 2567 วันที่ตอบรับบทความ : 23 มีนาคม 2567

Received: 22 June 2023, Revised: 20 March 2024, Accepted: 23 March 2024

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อหาตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลารายได้รัฐบาลของประเทศไทยทั้ง 4 ส่วน ได้แก่ กรมสรรพากร กรมสรรพสามิต กรมศุลกากร และหน่วยงานอื่น เช่น รัฐวิสาหกิจและกรมธนารักษ์ เป็นต้น หลังจากฟื้นตัวของสถานการณ์การแพร่ระบาดโควิด-19 โดยใช้ข้อมูลเปิดเผยของกระทรวงการคลังตั้งแต่เดือนตุลาคม พ.ศ. 2555 ถึงเดือนเมษายน พ.ศ. 2566 จำนวน 127 เดือน แบ่งข้อมูลออกเป็น 2 ชุด คือ ชุดข้อมูลฝึกฝนใช้ข้อมูลในช่วงเดือนตุลาคม พ.ศ. 2555 ถึงเดือนเมษายน พ.ศ. 2565 จำนวน 115 เดือน และชุดข้อมูลทดสอบในช่วงเดือนพฤษภาคม พ.ศ. 2565 ถึงเดือนเมษายน พ.ศ. 2566 จำนวน 12 เดือน สร้างตัวแบบพยากรณ์ทางสถิติ 5 วิธี ได้แก่ วิธีแยกส่วนประกอบ แนวโน้มเส้นตรงฤดูกาลรูปแบบบวก วิธีแยกส่วนประกอบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวก วิธีการทำให้เรียบของโฮลต์-วินเทอร์รูปแบบบวก วิธีบอกซ์-เจนกินส์ และวิธีพยากรณ์รวมด้วยการถ่วงน้ำหนักด้วยสัมประสิทธิ์การถดถอย มาเปรียบเทียบค่าความคลาดเคลื่อนโดยใช้เกณฑ์ร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยเป็นตัววัดความแม่นยำเพื่อหาตัวแบบที่เหมาะสมในชุดข้อมูลทดสอบ ผลการศึกษาพบว่า ตัวแบบพยากรณ์โดยวิธีการทำให้เรียบของโฮลต์-วินเทอร์รูปแบบบวกเป็นตัวแบบที่เหมาะสมที่สุด สำหรับกรมสรรพากร กรมสรรพสามิต และหน่วยงานอื่น ส่วนตัวแบบพยากรณ์โดยวิธีบอกซ์-เจนกินส์ตัวแบบ SARIMA(2,1,0)(0,0,2)<sub>12</sub> เป็นตัวแบบที่เหมาะสมที่สุดสำหรับกรมศุลกากร

คำสำคัญ : รายได้รัฐบาลของประเทศไทย การแยกส่วนประกอบ โฮลต์-วินเทอร์ บอกซ์-เจนกินส์ การพยากรณ์รวม

\*ที่อยู่ติดต่อ E-mail address: wathaminsan@gmail.com

<https://doi.org/10.55003/scikmitl.2024.259486>

## Abstract

The objective of this research is to identify an appropriate forecasting model for government revenue time series data across four sectors in Thailand: the Revenue Department, the Excise Department, the Customs Department, and the other agencies such as the Stat Enterprises and the Treasury Department. The study utilizes publicly available data from the website Ministry of Finance Open Data, covering the period from October 2012 to April 2023, totaling 127 months. The data set is divided into two subsets: a training data set spanning from October 2012 to April 2022 (115 months) and a testing data set from May 2022 to April 2023 (12 months). Five statistical forecasting methods are employed: additive trend-based linear with seasonal decomposition, additive trend-based logarithmic with seasonal decomposition, smoothing with Holt-Winters' additive method, the Box-Jenkins method, and a combined forecasting approach using regression analysis. The performance of these models is evaluated using the Mean Absolute Percentage Error (MAPE) metric applied to the testing data set. The findings of the study reveal that smoothing with Holt-Winters' additive method is the most suitable model for forecasting the Revenue Department, the Excise Department, and the other agencies. Additionally, the Box-Jenkins method with a SARIMA(2,1,0)(0,0,2)<sub>12</sub> model is identified as the most appropriate choice for forecasting the Customs Department.

**Keywords:** Government Revenue of Thailand, Decomposition, Holt-Winters, Box-Jenkins, Combined forecasting

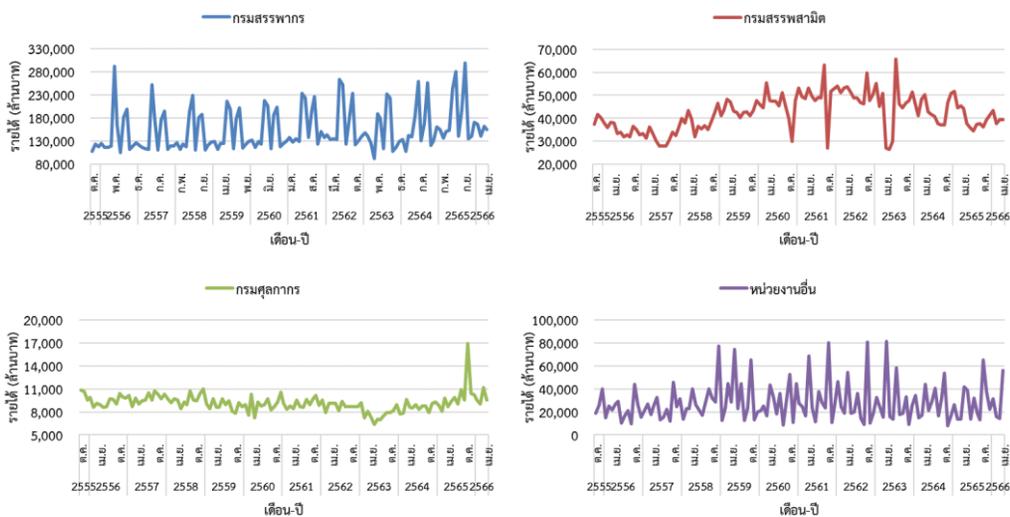
## 1. บทนำ

การคาดการณ์รายได้รัฐบาลที่แม่นยำเป็นสิ่งสำคัญสำหรับการวางแผนเศรษฐกิจและการกำหนดนโยบายที่ดีในประเทศไทย ซึ่งรายได้หลักของประเทศไทยประกอบด้วย 4 ส่วน ได้แก่ กรมสรรพากร กรมสรรพสามิต กรมศุลกากร และหน่วยงานอื่น (ส่วนราชการอื่น กรมธนารักษ์ เงินส่วนเกินจากการจำหน่ายพันธบัตรรัฐบาล แปรรูปรัฐวิสาหกิจ และรัฐวิสาหกิจ) โดยกรมสรรพากร กรมสรรพสามิตและกรมศุลกากรมีรายได้หลักจากการเก็บภาษีในรูปแบบต่าง ๆ จากการดำเนินกิจกรรมที่เกิดขึ้นผ่านการบริหารจัดการของประเทศ รายได้หลัก 5 อันดับแรกของกรมสรรพากรปีงบประมาณ พ.ศ. 2565 ได้แก่ ภาษีมูลค่าเพิ่ม ภาษีเงินได้นิติบุคคล ภาษีเงินได้บุคคลธรรมดา ภาษีเงินได้ปิโตรเลียม และภาษีธุรกิจเฉพาะ คิดเป็นรายได้ประมาณ 2,148,567 ล้านบาท และถ้ารวมรายได้ทุกรายการคิดเป็นรายได้ประมาณ 2,166,430 ล้านบาท กรมสรรพสามิตมีรายได้จากการเก็บภาษี 5 อันดับแรกของปีงบประมาณ พ.ศ. 2565 ได้แก่ ภาษีน้ำมัน ภาษिरถยนต์ ภาษีเบียร์ ภาษีสุรา และภาษียาสูบ คิดเป็นรายได้ประมาณ 469,101 ล้านบาท และถ้ารวมรายได้ทุกรายการคิดเป็นรายได้ประมาณ 503,465 ล้านบาท กรมศุลกากรจัดเก็บอากรขา

เข้าและอากรขาออกจากการส่งสินค้าเข้าหรือออกจากราชอาณาจักรไทย โดยรายได้หลักมาจากอากรขาเข้า ซึ่งปีงบประมาณ พ.ศ. 2565 มีรายได้ประมาณ 108,896 แสนล้านบาท และถ้ารวมรายได้ทุกรายการคิดเป็นรายได้ประมาณ 110,453 แสนล้านบาท และในหน่วยงานอื่นรายได้หลักมาจาก 2 แหล่ง คือ รัฐวิสาหกิจและส่วนราชการอื่น มีรายได้รวมทุกรายการทั้งหมดประมาณ 290,924 แสนล้านบาท [1]-[2]

จากข้อมูลรายได้รัฐบาลทั้งหมดมียอดรายได้รวมทุกส่วนปีงบประมาณ พ.ศ. 2565 เป็นเงิน 3,071,272 ล้านล้านบาท ซึ่งเป็นจำนวนเงินที่สูงมากและสำคัญต่อประเทศ ดังนั้นเพื่อให้ได้การประมาณรายได้ที่ใกล้เคียงและถูกต้อง การใช้ตัวแบบพยากรณ์ทางสถิติมาพยากรณ์จึงมีความสำคัญเพื่อใช้ผลการวิเคราะห์และค่าพยากรณ์ไปวางแผนการจัดเก็บหรือนำไปลงทุนต่อไป วิธีการทางสถิตินั้นมีด้วยกันหลายวิธี วิธีการหนึ่งที่เป็นที่นิยมคือการวิเคราะห์อนุกรมเวลา (Time series analysis) เป็นกระบวนการทางสถิติที่ใช้ในการศึกษาและการวิเคราะห์แบบลึกลับเกี่ยวกับลักษณะและแบบแผนของข้อมูลที่เปลี่ยนแปลงตามเวลา การวิเคราะห์นั้นใช้เครื่องมือทางสถิติเพื่อตรวจสอบแนวโน้ม ฤดูกาล รูปแบบ และความสัมพันธ์ระหว่างค่าของข้อมูลในช่วงเวลาที่ต่างกัน ยกตัวอย่างเช่น วิธีการแยกส่วนประกอบ วิธีการทำให้เรียบ (Smoothing method) ของโฮลต์-วินเทอร์ วิธีบ็อกซ์-เจนกินส์ และวิธีพยากรณ์รวม (Combined forecasting method) เป็นวิธีการทางสถิติที่นิยมใช้ในการวิเคราะห์อนุกรมเวลา เนื่องจากเป็นวิธีการพยากรณ์ที่สามารถจัดการกับความซับซ้อนของข้อมูลอนุกรมเวลาได้ ทำให้เกิดการใช้ประโยชน์จากแบบจำลองการพยากรณ์ อาทิเช่น Benvenuto และคณะ [3] ได้ทำนายการแพร่ระบาดของโควิด-19 ที่เป็นภัยคุกคามระดับโลก ได้มีการประยุกต์ตัวแบบรวมการถดถอยในตัวกับการเคลื่อนที่ (Autoregressive Integrated Moving Average: ARIMA) ซึ่งเป็นวิธีบ็อกซ์-เจนกินส์ ที่สามารถใช้ประโยชน์ในการทำนายการแพร่ระบาดของโรคโควิด-19 ในงานวิจัยของ Sen และ Chaudhuri [4] ได้ทำนายค่าดัชนีอนุกรมเวลาของภาคสินค้าอุปโภคบริโภคของประเทศอินเดีย ซึ่งกล่าวว่าการเคลื่อนไหวของราคาหุ้นเป็นแบบสุ่มตามธรรมชาติ การทำนายราคาหุ้นในอดีตด้วยการใช้การวิเคราะห์อนุกรมเวลาเป็นเรื่องที่ยากและท้าทายต่อการศึกษา แต่ต่อมาตัวแบบพยากรณ์ทางสถิติได้มีการพัฒนาอย่างมากทำให้สามารถวิเคราะห์ได้แม่นยำมากขึ้น ในการศึกษาขั้นนี้ได้ใช้ตัวแบบ 6 ตัวแบบ โดยแบ่งเป็น 2 กลุ่มวิธี คือ วิธีการทำให้เรียบของโฮลต์-วินเทอร์ และวิธีของบ็อกซ์-เจนกินส์ นอกจากนี้ Xiao และคณะ [5] ได้มีการใช้วิธีการพยากรณ์รวม วิธีของบ็อกซ์-เจนกินส์ ตัวแบบอาร์ชโครงข่ายประสาทเทียมแบบแพร่กลับ (ARCH Backpropagation Neural Network) สำหรับการพยากรณ์พลังงานลมกรณีศึกษาในประเทศจีน สมบูรณ์ ศิริสรธริทธิ และคณะ [6] ได้ศึกษาการฟื้นตัวของประเทศไทยสู่ความยั่งยืนทางการคลังภายหลังสถานการณ์การแพร่ระบาดของโรคโควิด-19 โดยใช้การพยากรณ์วิธีแยกส่วนประกอบ (Decomposition method) ซึ่งพิจารณาเฉพาะแนวโน้มเส้นตรงโดยปรับข้อมูลเป็นรายปี พบว่าแนวโน้มด้านรายได้รัฐบาลหรือการคลังมีแนวโน้มลดลงในช่วง พ.ศ. 2563 จำนวนร้อยละ 10.01 เมื่อเทียบกับช่วงปี พ.ศ. 2562 Tibulca [7] ศึกษาผลกระทบของการระบาดใหญ่ของโควิด-19 ต่อรายได้ภาษีในสหภาพยุโรป รัฐบาลต่าง ๆ ต้องเผชิญกับความจำเป็นในการดำเนินนโยบายที่จะตอบโต้ผลกระทบด้านลบทางเศรษฐกิจ และได้มีการวิเคราะห์สถานการณ์ความยั่งยืนของการคลังหลังสถานการณ์โควิด-19 การระบาดใหญ่ของโรคครั้งนี้นำไปสู่วิกฤตเศรษฐกิจที่มีผลกระทบกว้างไกลไปทั่วโลก ในงานวิจัยจึงได้ทำการคาดการณ์หรือพยากรณ์รายรับจากภาษี เพื่อช่วยวางแผนและออกแบบนโยบายของเศรษฐกิจสำหรับประเทศสมาชิกสหภาพยุโรป โดยใช้การวิเคราะห์ถดถอย (Regression analysis) ในการศึกษา

มีวิธีการจำนวนมากในงานวิจัยสำหรับการพยากรณ์ข้อมูลอนุกรมเวลา ดังนั้นการพิจารณาใช้วิธีการที่เหมาะสมในงานวิจัยนี้จะพิจารณาเบื้องต้นจากลักษณะของข้อมูลอนุกรมเวลารายได้รัฐบาล [1] ดังรูปที่ 1 พบว่ารายได้รัฐบาลของกรมสรรพากร มีลักษณะรูปแบบซ้ำกันในแต่ละปีชัดเจนที่สุด กรมสรรพสามิตและหน่วยงานอื่นก็มีรูปแบบซ้ำกันในแต่ละปีให้เห็นเช่นกัน โดยเฉพาะในช่วงแรกของข้อมูล ส่วนกรมศุลกากรรูปแบบซ้ำกันไม่ค่อยชัดเจน และรายได้รัฐบาลทั้ง 4 มีบางช่วงมีลักษณะแนวโน้มเพิ่มขึ้นหรือลดลงตามการเปลี่ยนแปลงของเวลา จึงตัดสินใจในเบื้องต้นควรใช้วิธีการทางสถิติที่สามารถตรวจจับลักษณะแนวโน้มและฤดูกาลได้ นั่นคือ วิธีการแยกส่วนประกอบ วิธีการทำให้เรียบ วิธีการบอกซ์-เจนกินส์ และวิธีการพยากรณ์รวม โดยแต่ละวิธีมีวิธีการทางสถิติในการสร้างตัวแบบแยกย่อยไปตามลักษณะเฉพาะของข้อมูล การวิเคราะห์ลักษณะเฉพาะของข้อมูลจะสามารถคัดเลือกวิธีการแยกย่อยที่เหมาะสมเหล่านั้นได้ การวิเคราะห์ลักษณะเฉพาะเป็นการศึกษาการเคลื่อนไหวของอนุกรมเวลา ได้แก่ ทดสอบองค์ประกอบแนวโน้ม ฤดูกาล และความแปรปรวนของข้อมูลที่จะกล่าวถึงในผลการทดลองและวิจารณ์



รูปที่ 1. ข้อมูลรายเดือนรายได้รัฐบาลแยกตามแหล่งที่มาของรายได้ 4 แหล่ง ตั้งแต่เดือนตุลาคม พ.ศ. 2555 ถึงเดือนเมษายน พ.ศ. 2565

ดังนั้นการเลือกใช้สถิติวิเคราะห์อนุกรมเวลาในการวิจัยครั้งนี้จึงนำเสนอทั้งสิ้น 5 วิธี ได้แก่ วิธีการแยกส่วนประกอบตัวแบบแนวโน้มเส้นตรงฤดูกาลรูปแบบบวก วิธีการแยกส่วนประกอบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวก วิธีการของไฮลด์-วินเทอร์รูปแบบบวก วิธีบอกซ์-เจนกินส์ และวิธีพยากรณ์รวม โดยทั้ง 5 วิธีนี้เหมาะสมกับลักษณะของชุดข้อมูลตามการทดสอบทางสถิติ โดยใช้เกณฑ์เปรียบเทียบเพื่อคัดเลือกตัวแบบที่เหมาะสมจากทั้ง 5 วิธี ด้วยร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) ที่น้อยที่สุด ผลลัพธ์จากงานวิจัยเพื่อเป็นแนวทางสำหรับผู้กำหนดนโยบายและนักวางแผนเศรษฐกิจเกี่ยวกับแนวทางที่ดีที่สุดในการคาดการณ์รายได้รัฐบาลไทยหลังจากการฟื้นตัวของสถานการณ์โค

วิด-19 การค้นพบของบทความนี้อาจจะมีส่วนช่วยในการพัฒนานโยบายและยุทธศาสตร์ทางเศรษฐกิจที่มีประสิทธิภาพมากขึ้นสำหรับประเทศไทย

## 2. วิธีการทดลอง

การศึกษาวิจัยครั้งนี้มีขั้นตอนวิธีการทดลอง ดังนี้

### 2.1 การจัดเตรียมข้อมูล

การวิจัยนี้มุ่งเน้นเพื่อศึกษาวิธีการพยากรณ์ที่เหมาะสมของรายได้รัฐบาลไทย โดยประกอบไปด้วยรายได้หลัก 4 ส่วน คือ กรมสรรพากร กรมสรรพสามิต กรมศุลกากร และหน่วยงานอื่น ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิที่รวบรวมจากส่วนบริหารจัดการข้อมูลสารสนเทศกระทรวงการคลัง ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงการคลัง [1] ซึ่งลักษณะข้อมูลมีการจำแนกเป็นรายเดือน โดยใช้ข้อมูลตั้งแต่ปีงบประมาณเดือนตุลาคม พ.ศ. 2555 ถึง เดือนเมษายน พ.ศ. 2566 ทั้งหมด 127 เดือน จัดแบ่งข้อมูลออกเป็น 2 ชุด ดังนี้ คือ ชุดที่ 1 เป็นชุดข้อมูลฝึกฝน (Training data set) ตั้งแต่เดือนตุลาคม พ.ศ. 2555 ถึงเดือนเมษายน พ.ศ. 2565 จำนวน 115 เดือน เพื่อใช้สำหรับการสร้างตัวแบบพยากรณ์ที่เหมาะสมของแต่ละวิธีการพยากรณ์ ส่วนข้อมูลชุดที่ 2 เป็นชุดข้อมูลทดสอบ (Test data set) ตั้งแต่เดือนพฤษภาคม พ.ศ. 2565 ถึงเดือนเมษายน พ.ศ. 2566 จำนวน 12 เดือน โดยใช้เกณฑ์เปรียบเทียบเพื่อคัดเลือกตัวแบบที่เหมาะสมจากทั้ง 5 วิธี ด้วยร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ยที่น้อยที่สุด

### 2.2 องค์ประกอบของอนุกรมเวลา

ในการวิเคราะห์อนุกรมเวลาจะแยกองค์ประกอบต่าง ๆ ที่ประกอบกันขึ้นเป็นอนุกรมเวลา การทราบองค์ประกอบของอนุกรมเวลาเป็นสิ่งสำคัญในการตัดสินใจเลือกวิธีที่เหมาะสมในการสร้างตัวแบบการพยากรณ์ โดยงานวิจัยนี้องค์ประกอบของอนุกรมเวลาจะใช้เป็นตัวตัดสินใจในการเลือกตัวแบบการพยากรณ์ในวิธีแยกส่วนประกอบและวิธีการทำให้เรียบ โดยพิจารณาองค์ประกอบของการแปรผันของอนุกรมเวลา 3 ส่วน ดังนี้

1. แนวโน้ม (Trend) ใช้สถิติทดสอบรันส์ (Runs test) ในการตรวจสอบ โดยนำข้อมูลอนุกรมเวลา  $Y$  มาหาค่าเฉลี่ยเคลื่อนที่กลางแบบ 12 เดือน เพื่อกำจัดอิทธิพลของฤดูกาล จากนั้นจึงใช้ข้อมูลที่กำจัดอิทธิพลของฤดูกาลแล้ว  $Y^T$  ในการทดสอบ

2. การเปลี่ยนแปลงหรือความแปรผันตามฤดูกาล (Seasonal variation) ใช้สถิติทดสอบ Kruskal-Wallis ในการตรวจสอบ โดยนำข้อมูลอนุกรมเวลาลบข้อมูลที่กำจัดอิทธิพลของฤดูกาลจะเป็นการกำจัดอิทธิพลของแนวโน้ม  $Y^S = Y - Y^T$  จากนั้นใช้ข้อมูล  $Y^S$  ในการทดสอบ

3. ความแปรปรวนของข้อมูลอนุกรมเวลา ใช้สถิติทดสอบของเลวีเน (Levene's test) ถ้าความแปรปรวนของข้อมูลอนุกรมเวลาเท่ากัน (Homogeneity of variance หรือ Homoscedasticity) หรือความแปรปรวนคงที่ จะเลือกใช้สมการรูปแบบบวกและถ้าความแปรปรวนของข้อมูลอนุกรมเวลาไม่เท่ากัน จะใช้สมการรูปแบบคูณในการพยากรณ์ข้อมูลอนุกรมเวลาที่มีส่วนประกอบแนวโน้มและฤดูกาล

## 2.3 การสร้างตัวแบบพยากรณ์

ในการทําวิจัยครั้งนี้ได้กำหนดสัญลักษณ์ไว้ดังนี้

$Y_t$  คือ ข้อมูลอนุกรมเวลา ณ เวลา  $t$

$\hat{Y}_t$  คือ ค่าพยากรณ์ ณ เวลา  $t$

$\varepsilon_t$  คือ ความคลาดเคลื่อนที่มีการแจกแจงปกติและเป็นอิสระกัน มีค่าเฉลี่ยเท่ากับศูนย์ และ  
ความแปรปรวนคงที่ทุกช่วงเวลา

$t$  คือ ช่วงเวลา มีค่าตั้งแต่ 1 ถึง  $n_1$  หรือ  $n_2$  หรือ  $n$  แล้วแต่กรณี

กรณีอยู่ในขั้นตอนสร้างตัวแบบ  $n_1$  แทนจำนวนข้อมูลในอนุกรมเวลาชุดข้อมูลฝึกฝน 115  
เดือน

กรณีอยู่ในขั้นตอนทดสอบตัวแบบ  $n_2$  แทนจำนวนข้อมูลในอนุกรมเวลาชุดข้อมูลทดสอบ  
12 เดือน

กรณีอยู่ในขั้นตอนการพยากรณ์ล่วงหน้า เดือนพฤษภาคม พ.ศ. 2566 ถึงเดือน เมษายน  
พ.ศ. 2567  $n$  แทนจำนวนข้อมูลทั้งหมด 127 เดือน

$s$  คือ คาบของฤดูกาลมีค่าเท่ากับ 12

### 2.3.1 วิธีแยกส่วนประกอบ

วิธีที่แยกอนุกรมเวลาออกเป็นส่วนประกอบต่าง ๆ [8] คือ แนวโน้ม ฤดูกาล วัฏจักร และ  
ส่วนประกอบไม่ปกติ โดยทั่วไปการพยากรณ์ในระยะสั้น ส่วนประกอบที่มีผลต่อการพยากรณ์จะมี 2  
ส่วนประกอบ คือ แนวโน้มและฤดูกาล

แนวโน้มกำหนดให้มีตัวแบบ 2 ตัวแบบ คือ ตัวแบบแนวโน้มเส้นตรง (Linear trend model) และ  
ตัวแบบแนวโน้มลอการิทึม (Logarithm trend model)

ตัวแบบและตัวแบบพยากรณ์แนวโน้มเส้นตรง ดังสมการที่ (1) และสมการที่ (2) ตามลำดับ ดังนี้

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t \quad (1)$$

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t \quad (2)$$

ตัวแบบและตัวแบบพยากรณ์แนวโน้มลอการิทึม ดังสมการที่ (3) และสมการที่ (4) ตามลำดับ ดังนี้

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

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 \ln(t) \quad (4)$$

เมื่อพารามิเตอร์  $\beta_0$  และ  $\beta_1$  แทนระดับของข้อมูลและความชัน ตามลำดับ ส่วน  $\hat{\beta}_0$  และ  $\hat{\beta}_1$  เป็น  
ตัวประมาณของ  $\beta_0$  และ  $\beta_1$  ตามลำดับ

เมื่อนำส่วนประกอบของข้อมูลอนุกรมเวลามีแนวโน้มและฤดูกาลเข้ารวมกัน จะแบ่งการพิจารณา  
เป็น 2 รูปแบบ ได้แก่ รูปแบบบวก (Additive decomposition) เหมาะกับอนุกรมเวลาที่มีความแปรผัน

ตามฤดูกาลคงที่ และรูปแบบคูณ (Multiplicative decomposition) เหมาะกับอนุกรมเวลาที่มีความแปรผันตามฤดูกาลไม่คงที่ โดยในงานวิจัยนี้พิจารณาเฉพาะรูปแบบบวกเนื่องจากผลการทดสอบตามตารางที่ 1 ที่จะกล่าวถึงในผลการทดลองและวิจารณ์

ตัวแบบแนวโน้มเส้นตรงฤดูกาลรูปแบบบวก ตัวแบบและตัวแบบพยากรณ์ดังสมการที่ (5) และสมการที่ (6) ตามลำดับ

$$Y_t = \beta_0 + \beta_1 t + S_t + \varepsilon_t \quad (5)$$

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 t + \hat{S}_t \quad (6)$$

ตัวแบบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวก ตัวแบบและตัวแบบพยากรณ์ดังสมการที่ (7) และสมการที่ (8) ตามลำดับ

$$Y_t = \beta_0 + \beta_1 \ln(t) + S_t + \varepsilon_t \quad (7)$$

$$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 \ln(t) + \hat{S}_t \quad (8)$$

เมื่อพารามิเตอร์  $S_t$  แทนความแปรผันตามฤดูกาล ณ เวลา  $t$  ส่วน  $\hat{S}_t$  เป็นตัวประมาณของ  $S_t$

### 2.3.2 วิธีการทำให้เรียบ

วิธีการทำให้เรียบ [9] นั้นเป็นวิธีการสร้างสมการพยากรณ์จากค่าสังเกตในอนุกรมเวลาบางส่วนหรือทั้งหมด โดยให้น้ำหนักกับค่าสังเกตในอนุกรมเวลาแตกต่างกันแล้วแต่ส่วนประกอบของอนุกรมเวลา

ในกรณีที่ส่วนประกอบของข้อมูลอนุกรมเวลาที่มีแนวโน้มและฤดูกาล จะสร้างตัวแบบพยากรณ์ด้วยวิธีการทำให้เรียบของโฮลต์-วินเทอร์รูปแบบบวกหรือแบบคูณ วิธีของโฮลต์-วินเทอร์ใช้ค่าทำให้เรียบ 3 ค่า ได้แก่  $\alpha$ ,  $\gamma$  และ  $\delta$  ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1 สำหรับค่าระดับ (Level parameter) ค่าที่ใกล้เคียงกับ 1 บ่งชี้ว่าตัวแบบปรับให้เข้ากับการสังเกตล่าสุดอย่างรวดเร็ว ในขณะที่ค่าที่ใกล้ 0 จะให้ความสำคัญกับการสังเกตในอดีตมากกว่า ค่าแนวโน้ม (Trend parameter) ค่าสูงช่วยให้ตัวแบบปรับตัวเข้ากับการเปลี่ยนแปลงของแนวโน้มได้อย่างรวดเร็ว ในขณะที่ค่าต่ำจะทำให้ตัวแบบตอบสนองต่อการเปลี่ยนแปลงของแนวโน้มล่าช้าลง และค่าฤดูกาล (Seasonal parameter) เป็นอิทธิพลของฤดูกาลหรือดัชนีฤดูกาล ค่าที่สูงขึ้นส่งผลให้มีความยืดหยุ่นและความไวต่อความผันผวนของฤดูกาลมากขึ้น ในขณะที่ค่าที่ต่ำกว่าจะทำให้ตัวแบบมีความทนทานต่อการเปลี่ยนแปลงตามฤดูกาลในระยะสั้นมากขึ้น โดยในงานวิจัยนี้พิจารณาเฉพาะรูปแบบบวกเนื่องจากผลการทดสอบตามตารางที่ 1 ที่จะกล่าวถึงในผลการทดลองและวิจารณ์

รูปแบบบวก ตัวแบบและตัวแบบพยากรณ์ดังสมการที่ (5) และสมการที่ (9) ตามลำดับ

$$\hat{Y}_{t+p} = \hat{T}_{t+p} + \hat{S}_{t+p-s} \quad \text{สำหรับ } p=1,2,\dots \quad (9)$$

เมื่อ  $\hat{T}_{t+p} = \hat{T}_t + p\hat{\beta}_t$

$\hat{Y}_{t+p}$  คือ ค่าพยากรณ์ ณ เวลา  $t+p$

$\hat{T}_t$  คือ ตัวประมาณระดับ  $\hat{\beta}_t$  คือ ตัวประมาณแนวโน้ม และ  $\hat{S}_t$  คือ ตัวประมาณฤดูกาล ณ เวลา  $t$   
โดยสามารถคำนวณได้ดังสมการที่ (10) - (12)

$$\hat{T}_t = \alpha(Y_t - \hat{S}_{t-s}) + (1-\alpha)(\hat{T}_{t-1} + \hat{\beta}_{t-1}) \quad (10)$$

$$\hat{\beta}_t = \gamma(\hat{T}_t - \hat{T}_{t-1}) + (1-\gamma)\hat{\beta}_{t-1} \quad (11)$$

$$\hat{S}_t = \delta(Y_t - \hat{T}_t) + (1-\delta)\hat{S}_{t-s} \quad (12)$$

### 2.3.3 วิธีบอกซ์-เจนกินส์

วิธีบอกซ์-เจนกินส์ [10] เป็นวิธีที่ให้ค่าพยากรณ์ที่มีความถูกต้องสูงกว่าค่าพยากรณ์จากวิธีการอื่นในการพยากรณ์ระยะสั้น การกำหนดตัวแบบพยากรณ์ทำได้โดยการตรวจสอบฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) ของอนุกรมเวลาแบบคงที่ (Stationary time series) หรืออนุกรมเวลาที่มีค่าเฉลี่ยของความแปรปรวนคงที่ [9] ตัวแบบทั่วไปวิธีบอกซ์-เจนกินส์ คือ Seasonal Autoregressive Integrated Moving Average หรือ SARIMA(  $p, d, q$  )(  $P, D, Q$  )<sub>s</sub> ดังที่แสดงในสมการที่ (13)

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Y_t = \delta + \theta_q(B)\Theta_q(B^s)\varepsilon_t \quad (13)$$

เมื่อ  $\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 \dots - \phi_p B^p$  แทนการถดถอยในตัวแบบไม่มีฤดูกาลอันดับ  $p$  (Non-seasonal autoregressive operator of order  $p$ : AR( $p$ ))

$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \Phi_3 B^{3s} \dots - \Phi_p B^{Ps}$  แทนการถดถอยในตัวแบบมีฤดูกาลอันดับ  $P$

(Seasonal autoregressive operator of order  $P$ : SAR( $P$ ))

$(1-B)^d$  คือ อันดับของผลต่างแบบไม่มีฤดูกาล (Non-seasonal difference)

$(1-B^s)^D$  คือ อันดับของผลต่างแบบมีฤดูกาล (Seasonal difference)

$\delta = \phi_p(B)\Phi_p(B^s)\mu$  คือ ค่าคงตัว โดยที่  $\mu$  คือ ค่าเฉลี่ยของอนุกรมเวลาที่คงที่

$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 \dots - \theta_q B^q$  แทนค่าเฉลี่ยเคลื่อนที่แบบไม่มีฤดูกาลอันดับ  $q$  (Non-seasonal moving average operator of order  $q$ : MA( $q$ ))

$\Theta_q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \Theta_3 B^{3s} \dots - \Theta_q B^{Qs}$  แทนค่าเฉลี่ยเคลื่อนที่แบบมีฤดูกาลอันดับ  $Q$

(Seasonal moving average operator of order  $Q$ : SMA( $Q$ ))

$B$  คือ ตัวดำเนินการถดถอยหลัง (Backward operator) โดยที่  $B^s Y_t = Y_{t-s}$

$d$  และ  $D$  คือ ลำดับที่ของการหาผลต่างและผลต่างฤดูกาลของอนุกรมเวลา ตามลำดับ การสร้างตัวแบบพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์ มีขั้นตอนดังนี้

1) การตรวจสอบข้อมูลเพื่อพิจารณาว่าอนุกรมเวลาคงที่หรือไม่ โดยพิจารณาจากกราฟอนุกรมเวลา และพิจารณาจากกราฟ ACF และ PACF และการตรวจสอบ Unit root ด้วยการทดสอบ Dickey-Fuller (DF) ในงานวิจัยนี้กำหนดลำดับมากที่สุดของ Lag = 12 เนื่องมาจากการเลือกลำดับ Lag ที่เหมาะสมคือค่าเกณฑ์ข้อสนเทศของอะกะอิเกะ (Akaike's Information Criteria: AIC) ที่ต่ำที่สุด หากพบว่าอนุกรมเวลาไม่คงที่ (Non-stationary) ต้องแปลงอนุกรมเวลาให้คงที่ก่อนที่ดำเนินการในขั้นตอนต่อไป

- การแปลงข้อมูลด้วยการหาผลต่าง (Regular difference) เมื่อพิจารณากราฟ ACF และ PACF แล้วพบว่าอนุกรมเวลามีแนวโน้ม ค่า  $d$  จะถูกกำหนดจาก 0 เป็น 1 หรือ 2 ตามลักษณะของแนวโน้ม

- การแปลงข้อมูลด้วยการหาผลต่างแบบมีฤดูกาล เมื่อพิจารณากราฟ ACF และ PACF แล้วพบว่าอนุกรมเวลามีฤดูกาล ค่า  $D$  จะถูกกำหนดจาก 0 เป็น 1 หรือ 2 ตามลักษณะของฤดูกาล

- เทคนิควิธีการแปลงของบ็อกซ์-ค็อกซ์ (Box-Cox transformation) แก้ไขปัญหาเมื่อความแปรปรวนของอนุกรมเวลาไม่คงที่ ทำการแปลงข้อมูล  $Y$  เป็น  $\ln(Y)$

2) การกำหนดตัวแบบพยากรณ์ที่เป็นไปได้จากอนุกรมเวลาใหม่ที่คงที่ นั่นคือ กำหนดค่า  $p$ ,  $q$ ,  $P$  และ  $Q$  พร้อมทั้งประมาณค่าพารามิเตอร์ของตัวแบบด้วยวิธีกำลังสองน้อยที่สุด (Ordinary least squares method)

ในงานวิจัยนี้การกำหนดขอบเขตค่า  $p$ ,  $q$ ,  $P$  และ  $Q$  ตั้งแต่ 0 ถึง 3 ในการค้นหาพารามิเตอร์ที่เหมาะสมจึงใช้วิธีการค้นหาแบบกริด (Grid search) แทนการพิจารณาจากกราฟ ACF และ PACF ที่พิจารณาได้ยาก ถ้าพารามิเตอร์ที่เป็นไปได้มีความเป็นไปได้หลายตัวแบบ (ในกรณีนี้มีตัวแบบที่เป็นไปได้  $4^4$  หรือ 256 ตัวแบบ)

การค้นหาแบบกริดเป็นเทคนิคที่ใช้ในการเรียนรู้ของเครื่อง (Machine Learning) เพื่อค้นหาไฮเปอร์พารามิเตอร์ (Hyperparameters) ที่เหมาะสมที่สุดสำหรับตัวแบบที่กำหนด ไฮเปอร์พารามิเตอร์คือพารามิเตอร์ที่ตั้งค่าไว้ก่อนกระบวนการเรียนรู้จะเริ่มต้น ในกรณีนี้ค่าพารามิเตอร์สามารถกำหนดตัวแบบ SARIMA( $p, d, q$ )( $P, D, Q$ ), ขึ้นมาได้ ซึ่งประกอบด้วย  $p$ ,  $q$ ,  $P$  และ  $Q$  ที่มีค่าตั้งแต่ 0 ถึง 3 ดังนั้นชุดค่าผสมต่าง ๆ เกิดจากผลคูณคาร์ทีเซียน (Cartesian) ของค่าพารามิเตอร์อย่างเป็นระบบเพื่อระบุตัวแบบที่เหมาะสมที่สุดตามเกณฑ์ที่กำหนด

ขั้นตอนที่สำคัญของการค้นหาแบบกริดจึงอธิบายได้ดังนี้

1. กำหนดชุดของค่าที่เป็นไปได้สำหรับแต่ละพารามิเตอร์  $p$ ,  $q$ ,  $P$  และ  $Q$  ที่มีค่าตั้งแต่ 0 ถึง 3
2. สร้างกริดโดยใช้ผลคูณคาร์ทีเซียนของค่าที่เป็นไปได้สำหรับแต่ละพารามิเตอร์
3. สำหรับแต่ละจุดในกริด สร้างตัวแบบ SARIMA( $p, d, q$ )( $P, D, Q$ ), ตามพารามิเตอร์ในกริดที่กำหนดนั้น
4. คำนวณค่าประเมินความเหมาะสมของตัวแบบ โดยในงานวิจัยนี้ใช้ค่า AIC เป็นเกณฑ์ที่กำหนด
5. ทำซ้ำขั้นตอนที่ 3 และ 4 ทุกจุดของกริด

6. เรียงลำดับตัวแบบเหมาะสมตามค่า AIC จากน้อยไปมาก ดังนั้นจะได้ค่าพารามิเตอร์ของ  $p$ ,  $q$ ,  $P$  และ  $Q$  และ AIC ทุกตัวแบบที่เป็นไปตามแต่ละจุดในกริด

3) การตรวจสอบความเหมาะสมของตัวแบบพยากรณ์ที่กำหนดได้จากขั้นตอนที่ 2 โดยทดสอบตัวประมาณค่าพารามิเตอร์ของตัวแบบโดยสถิติการทดสอบที (t-test) และตรวจสอบข้อสมมติเกี่ยวกับความคลาดเคลื่อนจากการพยากรณ์ ( $e_t$ ) ที่ต้องมีคุณลักษณะความเป็นอิสระกันและมีการแจกแจงปรกติด้วยค่าเฉลี่ยเท่ากับศูนย์และมีความแปรปรวนคงที่ทุกช่วงเวลา สรุปรูปการทดสอบแต่ละสถิติทดสอบดังนี้

- ทดสอบตัวประมาณค่าพารามิเตอร์ของตัวแบบทุกตัว โดยสถิติการทดสอบที
- ความเป็นอิสระกัน โดยสถิติการทดสอบ Ljung-Box เมื่อกำหนดจำนวน Lag ที่พิจารณาเท่ากับ 24 (LB Lag24)
- การแจกแจงปรกติ โดยสถิติการทดสอบคอลโมโกรอฟ-สมิร์นอฟ (Kolmogorov-Smirnov test: KS)

- ค่าเฉลี่ยเท่ากับศูนย์ โดยสถิติการทดสอบที

- ความแปรปรวนคงที่ทุกช่วงเวลา โดยสถิติการทดสอบของเลวีเน (Levene's test: Levene)

4) ในกรณีที่ตัวแบบไม่ผ่านขั้นตอนที่ 3 จะเลือกตัวแบบที่มีค่า AIC ต่ำที่สุดตัวถัดไปจากขั้นตอนที่ 3 มาทดสอบ และดำเนินการจนกว่าจะได้ตัวแบบที่ผ่านขั้นตอนที่ 3 ทุกเงื่อนไข

5) พยากรณ์อนุกรมเวลาโดยใช้ตัวแบบพยากรณ์ที่เหมาะสมที่สุดจากขั้นตอนที่ 4 ยกเว้นไม่สามารถหาตัวแบบที่ผ่านทุกเงื่อนไขในขั้นตอนที่ 3 จะไม่ใช้ตัวแบบ SARIMA( $p, d, q$ )( $P, D, Q$ ) ในการพยากรณ์

### 2.3.4 การสร้างตัวแบบพยากรณ์โดยวิธีพยากรณ์รวม

การพยากรณ์รวมเป็นวิธีการประยุกต์ที่มีการรวมค่าพยากรณ์จากวิธีการพยากรณ์เดี่ยวตั้งแต่ 2 วิธีขึ้นไปเพื่อให้ได้ค่าพยากรณ์ใหม่ที่มีความคลาดเคลื่อนน้อยที่สุด การพยากรณ์รวมมีด้วยกันหลายวิธี เช่น วิธีถ่วงน้ำหนักความคลาดเคลื่อนกำลังสองเฉลี่ยผกผัน (Inverse of mean squares error) วิธีถ่วงน้ำหนักความคลาดเคลื่อนสมบูรณ์เฉลี่ยต่ำสุด (Minimum mean absolute error) วิธีหาค่าเหมาะที่สุดแบบวาฬ (Whale optimization algorithm) และวิธีถ่วงน้ำหนักด้วยสัมประสิทธิ์การถดถอย (Regression: REG) การพยากรณ์รวมด้วยวิธี REG เป็นวิธีการที่เหมาะสมสำหรับหลายงานวิจัยดังนี้ อินทิตรา เนขุนทด และเขตสิริ คำขอด [11] ศึกษาตัวแบบพยากรณ์อนุกรมเวลาผู้ป่วยโรคปอดอักเสบ นอกจากนี้ จิโรจน์ แซ่ยี่ง และคณะ [12] ศึกษาตัวแบบพยากรณ์อนุกรมเวลาปริมาณน้ำไหลเข้าอ่างเก็บน้ำของเขื่อนการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย (กฟผ.) 11 แห่ง อีกทั้ง ดิณฑณภัทร ตุ่นแก้ว และคณะ [13] หาตัวแบบพยากรณ์ที่เหมาะสมสำหรับพยากรณ์อัตราแลกเปลี่ยนสกุลเงินต่างประเทศ 10 สกุลที่คนไทยนิยมไปท่องเที่ยวมากที่สุดเทียบกับค่าเงินบาทพบว่า วิธีพยากรณ์รวม REG เป็นวิธีที่เหมาะสมกับข้อมูลอนุกรมเวลาเป็นส่วนใหญ่ในทุกงานวิจัยดังกล่าว ดังนั้นในงานวิจัยนี้จึงเลือกใช้วิธีพยากรณ์รวม REG ด้วยลักษณะการวิเคราะห์ถดถอยพหุคูณ (Multiple regression analysis) ในสมการที่ (14)

$$\hat{Y}_t = b_{Con} + b_{Dec\_T} \hat{Y}_{1t} + b_{Dec\_L} \hat{Y}_{2t} + b_{Smo} \hat{Y}_{3t} + b_{Box} \hat{Y}_{4t} \quad (14)$$

เมื่อ  $\hat{Y}_t$  คือ ค่าพยากรณ์รวม ณ เวลา  $t$  และ  $\hat{Y}_{1t}, \hat{Y}_{2t}, \hat{Y}_{3t}$  และ  $\hat{Y}_{4t}$  แทนค่าพยากรณ์เดี่ยว ณ เวลา  $t$  จากวิธีแยกส่วนประกอบแนวโน้มเส้นตรงฤดูกาลรูปแบบบวก วิธีแยกส่วนประกอบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวก วิธีการทำให้เรียบของโฮลต์-วินเทอร์รูปแบบบวก และวิธีบอกซ์-เจนกินส์ ตามลำดับ

$b_{Con}$  คือ ตัวประมาณค่าคงตัวของตัวแบบการวิเคราะห์ถดถอย

$b_{Dec-T}, b_{Dec-L}, b_{Smo}$  และ  $b_{Box}$  คือ ตัวประมาณค่าถ่วงน้ำหนักของแต่ละวิธีพยากรณ์เดี่ยว วิธีแยกส่วนประกอบแนวโน้มเส้นตรงฤดูกาลรูปแบบบวก วิธีแยกส่วนประกอบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวก วิธีการทำให้เรียบของโฮลต์-วินเทอร์รูปแบบบวก และวิธีบอกซ์-เจนกินส์ ตามลำดับ ด้วยวิธีกำลังสองน้อยที่สุด (Least squares method) [14]

### 2.3.5 การเปรียบเทียบความถูกต้องของค่าพยากรณ์ด้วยร้อยละความคลาดเคลื่อนสัมบูรณ์เฉลี่ย

ความถูกต้องของการพยากรณ์เป็นสิ่งที่ผู้ใช้ค่าพยากรณ์ต้องการ ความถูกต้องมีมากหรือน้อยขึ้นอยู่กับค่าความคลาดเคลื่อนของการพยากรณ์ ซึ่งเป็นผลต่างของค่าจริงและค่าพยากรณ์ ความคลาดเคลื่อนจะน้อยถ้าค่าพยากรณ์ใกล้เคียงกับค่าจริง และจะมากถ้าค่าจริงห่างจากค่าพยากรณ์มาก ในที่นี้เลือกการพิจารณาค่าความแม่นยำของตัวแบบพยากรณ์ใช้ MAPE ซึ่งเป็นค่าที่ไม่มีหน่วย เหมาะที่จะใช้เปรียบเทียบอนุกรมเวลาหลายชุดเมื่อใช้วิธีพยากรณ์เดียวกัน หรือเปรียบเทียบวิธีการพยากรณ์หลายวิธีเมื่อใช้อนุกรมเวลาชุดเดียวกัน ดังแสดงในสมการที่ (15)

$$MAPE = \frac{100}{n_2} \sum_{t=1}^{n_2} \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \quad (15)$$

เมื่อ  $n_2$  คือ จำนวนชุดข้อมูลในอนุกรมเวลาชุดข้อมูลทดสอบ 12 เดือน

## 3. ผลการทดลองและวิจารณ์

การพิจารณาลักษณะการเคลื่อนไหวของอนุกรมเวลารายได้รัฐบาล 4 ส่วน ชุดข้อมูลฝึกฝนตั้งแต่เดือนตุลาคม พ.ศ. 2555 ถึงเดือนเมษายน พ.ศ. 2565 จำนวน 115 ค่า ได้ผลการพิจารณาการเคลื่อนไหวของอนุกรมเวลาและการวิเคราะห์ด้วยวิธีการพยากรณ์ 5 วิธี มีรายละเอียดดังนี้

### 3.1. การเคลื่อนไหวของอนุกรมเวลา

ผลการวิเคราะห์ลักษณะการเคลื่อนไหวของอนุกรมเวลา ทดสอบองค์ประกอบแนวโน้ม ฤดูกาล และความแปรปรวนของข้อมูล โดยการใช้สถิติทดสอบรันส์ ทดสอบ Kruskal-Wallis และทดสอบของเลวินตามลำดับ แสดงในตารางที่ 1 พบว่าสถิติทดสอบรันส์ และ Kruskal-Wallis ค่าพี (p-value) น้อยกว่าระดับนัยสำคัญที่กำหนด แสดงว่ารายได้รัฐบาลทั้ง 4 ส่วนมีอิทธิพลจากแนวโน้มและฤดูกาล สถิติทดสอบของเลวินค่าพีมากกว่าระดับนัยสำคัญที่กำหนดแสดงว่ารายได้รัฐบาลทั้ง 4 มีความแปรปรวนคงที่ ดังนั้นวิธีแยก

ส่วนประกอบแนวโน้มเส้นตรงฤดูกาลรูปแบบบวก วิธีแยกส่วนประกอบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวก วิธีของโฮสต์-วินเทอร์รูปแบบบวกเหมาะสมที่จะนำมาใช้สร้างตัวแบบกับข้อมูลรายได้รัฐบาลทั้ง 4 ส่วน

ตารางที่ 1. สถิติทดสอบและค่าพีของสถิติทดสอบรันส์ Kruskal-Wallis และเลวิน แยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

รายได้รัฐบาล	ทดสอบรันส์		ทดสอบ Kruskal-Wallis		ทดสอบของเลวิน	
	Statistic	p-value	Statistic	p-value	Statistic	p-value
กรมสรรพากร	-2.2578	0.000*	87.089	0.000*	0.148	0.701
กรมสรรพสามิต	-7.5261	0.000*	37.386	0.000*	0.338	0.562
กรมศุลกากร	-1.8815	0.000*	38.877	0.000*	0.216	0.643
หน่วยงานอื่น	1.8815	0.000*	44.570	0.000*	2.777	0.098

หมายเหตุ \* ปฏิเสธสมมติฐานว่างที่ระดับนัยสำคัญ 0.05

### 3.2. วิธีแยกส่วนประกอบ

การศึกษาครั้งนี้แยกการศึกษาวิธีแยกส่วนประกอบเป็น 2 วิธี คือ วิธีแยกส่วนประกอบแนวโน้มเส้นตรงฤดูกาลรูปแบบบวกและวิธีแยกส่วนประกอบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวก การคำนวณหาค่าประมาณตัวแบบแนวโน้มเส้นตรงและแนวโน้มลอการิทึมได้ผลดังตารางที่ 2 และการคำนวณหาความแปรผันฤดูกาลจากตัวแบบแนวโน้มเส้นตรงได้ผลดังตารางที่ 3 และความแปรผันฤดูกาลจากตัวแบบแนวโน้มลอการิทึมดังตารางที่ 4 (ทั้งสองตารางนี้ถ้าไม่มีอิทธิพลฤดูกาลจะมีค่าเท่ากับ 0) ดังนั้นจากตารางที่ 3 และ 4 อิทธิพลของฤดูกาลของแต่ละรายได้รัฐบาลนั้นแตกต่างกัน กรมสรรพากรมีช่วงรายได้มากในเดือน พฤษภาคม มิถุนายน สิงหาคม และกันยายน กรมสรรพสามิตมีช่วงรายได้มากในเดือน ธันวาคม มกราคม มีนาคม และเมษายน กรมศุลกากรมีช่วงรายได้มากในเดือนตุลาคม พฤศจิกายน ธันวาคม มกราคม มีนาคม และสิงหาคม และรายได้หน่วยงานอื่นมีช่วงรายได้มากในเดือนตุลาคม ธันวาคม เมษายน พฤษภาคม และกรกฎาคม

ตารางที่ 2. ค่าประมาณพารามิเตอร์  $\hat{\beta}_0$  และ  $\hat{\beta}_1$  ของตัวแบบเส้นตรงและลอการิทึมแยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

ตัวแบบ		กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
เส้นตรง	$\hat{\beta}_0$	141,561.64	35,270.36	9,805.87	25,925.12
	$\hat{\beta}_1$	198.68	125.54	-13.49	35.73
ลอการิทึม	$\hat{\beta}_0$	129,086.17	26,128.54	10,658.62	19,298.83
	$\hat{\beta}_1$	6,354.28	4,348.39	-432.98	2,303.19

ตารางที่ 3. ค่าประมาณความแปรผันตามฤดูกาลด้วยวิธีแยกส่วนประกอบแนวโน้มเส้นตรงฤดูกาลรูปแบบบวกแยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

ความแปรผันตามฤดูกาล	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
ตุลาคม	-37,841	-2,424	118	22,782
พฤศจิกายน	-27,049	-474	493	-11,120
ธันวาคม	-21,175	3,186	530	3,624
มกราคม	-18,282	3,369	208	-2,904
กุมภาพันธ์	-31,793	-249	-820	-5,434
มีนาคม	-24,063	3,956	415	-7,465
เมษายน	-29,619	441	-420	17,455
พฤษภาคม	74,238	-2,743	-383	2,644
มิถุนายน	57,496	-2,825	-265	-10,094
กรกฎาคม	-35,682	210	-88	5,612
สิงหาคม	33,122	-1,519	308	-9,131
กันยายน	60,648	-927	-96	-5,968

หมายเหตุ ค่าน้อย  ค่ามากสำหรับแต่ละส่วนของรายได้รัฐบาล

ตารางที่ 4. ค่าประมาณความแปรผันตามฤดูกาลด้วยวิธีแยกส่วนประกอบตัวแบบแนวโน้มลอการิทึมฤดูกาลรูปแบบบวกแยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

ความแปรผันตามฤดูกาล	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
ตุลาคม	-36,930	-1,770	56	23,220
พฤศจิกายน	-26,520	-92	457	-10,856
ธันวาคม	-20,841	3,425	507	3,781
มกราคม	-18,062	3,519	193	-2,824
กุมภาพันธ์	-31,642	-156	-830	-5,416
มีนาคม	-23,951	4,011	407	-7,497
เมษายน	-29,525	474	-426	17,380
พฤษภาคม	73,751	-3,055	-350	2,540
มิถุนายน	57,006	-3,150	-232	-10,235
กรกฎาคม	-36,162	-118	-55	5,438
สิงหาคม	32,662	-1,844	340	-9,334
กันยายน	60,216	-1,244	-66	-6,197

หมายเหตุ ค่าน้อย  ค่ามากสำหรับแต่ละส่วนของรายได้รัฐบาล

### 3.3 การสร้างตัวแบบพยากรณ์โดยวิธีการทำให้เรียบ

การสร้างตัวแบบพยากรณ์โดยวิธีการทำให้เรียบของโฮลต์-วินเทอร์รูปแบบบวก สามารถประมาณค่าพารามิเตอร์ของอนุกรมเวลารายได้รัฐบาลทั้ง 4 ส่วนได้ค่า  $\alpha$ ,  $\gamma$  และ  $\delta$  ดังตารางที่ 5 ซึ่งตัว  $\delta$  ที่เป็นตัวปรับเรียบฤดูกาลมีค่าค่อนข้างมากในรายได้รัฐบาลกรมสรรพากร กรมศุลกากร และหน่วยงานอื่น คือ 0.53, 0.43 และ 0.41 ตามลำดับ นั้นแสดงว่าน้ำหนักของฤดูกาลในปีก่อนมีผลประมาณร้อยละ 40-50 และปีถัด ๆ ไปจะลดลงแบบเลขชี้กำลัง (Exponential)

ตารางที่ 5. ค่าพารามิเตอร์  $\alpha$ ,  $\gamma$  และ  $\delta$  วิธีของโฮลต์-วินเทอร์รูปแบบบวกแยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

พารามิเตอร์	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
$\alpha$	0.09	0.18	0.28	0.03
$\gamma$	0.00	0.00	0.00	0.00
$\delta$	0.53	0.13	0.43	0.41

หมายเหตุ ค่าน้อย  ค่ามากสำหรับแต่ละส่วนของรายได้รัฐบาล  
0.00 ใช้แสดงแทนค่าจริงเมื่อค่าเข้าใกล้ศูนย์เกินทศนิยมตำแหน่งที่ 2

### 3.4 การสร้างตัวแบบพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์

ในการพิจารณาสถิติทดสอบ DF ตารางที่ 6 พบว่า ค่าที่มีค่ามากกว่าระดับนัยสำคัญที่กำหนด แสดงว่าข้อมูลรายได้รัฐบาลทั้ง 4 ส่วนเป็นอนุกรมเวลาไม่คงที่ ควรทำการหาผลต่างโดยพิจารณาเพิ่มเติม ACF จากรูปที่ 2 ซึ่งพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ในตัวมีค่ามากและเป็นบวกเมื่อ Lag มีค่าน้อยและลดลงอย่างช้า ๆ เมื่อ Lag มีค่ามาก ดังนั้นจึงพิจารณาได้ว่าข้อมูลมีแนวโน้มควรกำหนด  $d=1$  ทุกข้อมูลรายได้รัฐบาลทั้ง 4 ส่วน และจากรูปที่ 3 เป็นกราฟ ACF จากข้อมูลที่กำลังจัดแนวโน้มพบว่า กรมสรรพากร และหน่วยงานอื่น Lag 12, 24, 36 และ 48 มีค่ามากและลดลงอย่างช้า ๆ ดังนั้นจึงพิจารณาได้ว่าข้อมูลมีฤดูกาลควรกำหนด  $D=1$  ในส่วนของกรมสรรพสามิตกราฟ ACF ตรวจสอบไม่พบลักษณะฤดูกาลดังนั้นจึงกำหนด  $D=0$  และสุดท้ายกรมศุลกากร Lag 12, 24, 36 และ 48 มีค่าลดลงเร็วดังนั้นจึงกำหนด  $D=0$  แต่ให้สังเกต  $P$  หรือ  $Q$  น่าจะมีค่ามากกว่า 0 ในขั้นตอนพิจารณาตัวแบบ

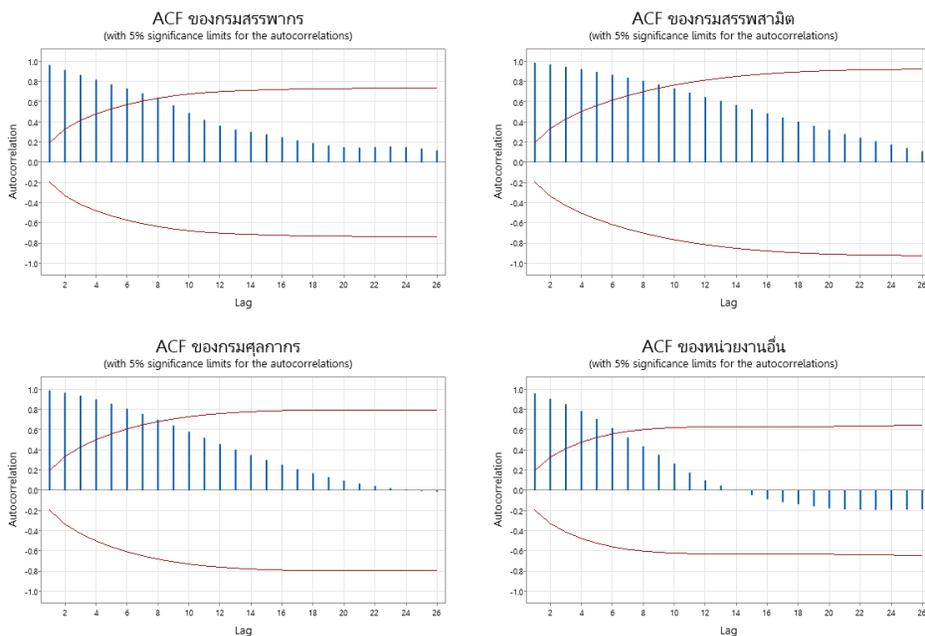
จากตารางที่ 6 และ 7 ตัวแบบพยากรณ์โดยวิธีบ็อกซ์-เจนกินส์ของอนุกรมเวลารายได้รัฐบาลทั้ง 4 ส่วน จากการค้นหาแบบกริดที่เป็นเทคนิคการเรียนรู้ของเครื่อง ข้อมูลอนุกรมเวลารายได้รัฐบาลกรมศุลกากรได้ตัวแบบ SARIMA(2,1,0)(0,0,2)<sub>12</sub> ที่เหมาะสมผ่านทุกเงื่อนไข คือ ทดสอบตัวประมาณค่าพารามิเตอร์ของตัวแบบโดยสถิติการทดสอบที่ ค่าพิน้อยกว่าระดับนัยสำคัญที่กำหนดทุกตัวประมาณที่ทดสอบ การตรวจสอบข้อสมมติเกี่ยวกับความคลาดเคลื่อนจากการพยากรณ์คือ การทดสอบความเป็นอิสระกันด้วยสถิติการทดสอบ LB Lag24 การแจกแจงปกติด้วยสถิติการทดสอบ KS ค่าเฉลี่ยเท่ากับศูนย์ด้วยสถิติการทดสอบที่ และมีความแปรปรวนคงที่ทุกช่วงเวลาด้วยสถิติทดสอบของเลวิน ค่าพิน่ากวาระดับนัยสำคัญที่กำหนดทุกตัวสถิติทดสอบ สำหรับข้อมูลอนุกรมเวลารายได้รัฐบาลกรมสรรพากรและกรมสรรพสามิตไม่ผ่านความคลาดเคลื่อนมีการแจกแจงปกติและความแปรปรวนคงที่ รายได้รัฐบาล

หน่วยงานอื่นไม่ผ่านความคลาดเคลื่อนมีความเป็นอิสระและมีการแจกแจงปกติ ดังนั้นมีเพียงตัวแบบรายได้รัฐบุคลากรกรมศุลกากรจะนำไปใช้ในการพยากรณ์ชุดข้อมูลทดสอบเท่านั้น

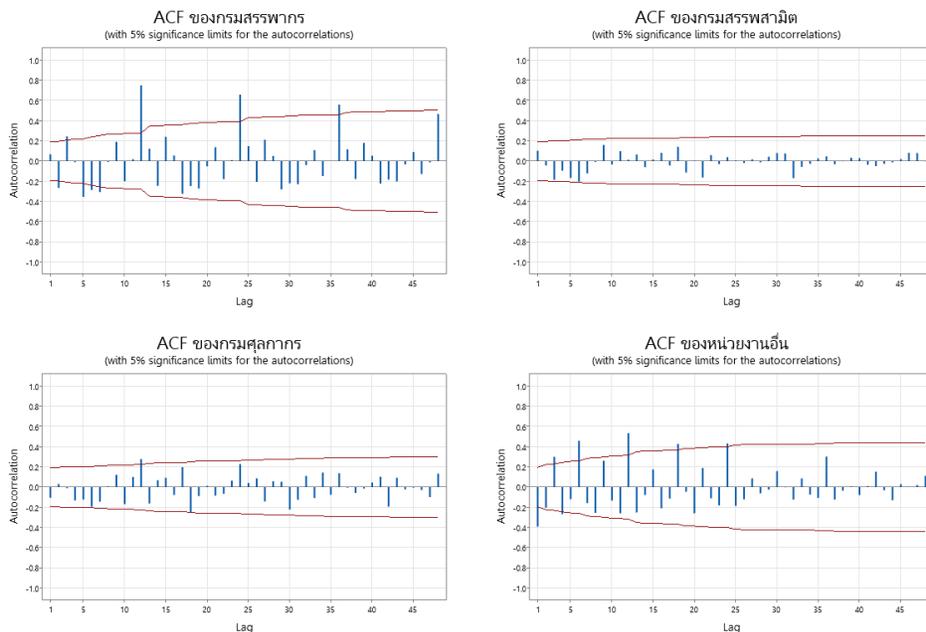
ตารางที่ 6. ตัวแบบบอซซ์-เจนกินส์ และการตรวจสอบความเหมาะสมของตัวแบบแยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

วิธีการตรวจสอบ	กรมสรรพากร		กรมสรรพสามิต		กรมศุลกากร		หน่วยงานอื่น	
	Statistic	p-Value	Statistic	p-value	Statistic	p-value	Statistic	p-Value
$SARIMA(p,d,q)(P,D,Q)_s$	SARIMA(0,1,1)(1,1,0) <sub>12</sub>		SARIMA(0,1,1)(0,0,0) <sub>12</sub>		SARIMA(2,1,0)(0,0,2) <sub>12</sub>		SARIMA(0,1,2)(2,1,0) <sub>12</sub>	
DF	-0.774	0.827	-2.447	0.129	-1.630	0.467	-2.699	0.074
LB Lag24	20.465	0.554	32.066	0.099	27.940	0.111	37.643	0.010*
KS	0.187	<0.010*	0.112	<0.010*	0.055	>0.150	0.097	0.026*
t-test	0.420	0.675	0.535	0.594	-0.134	0.893	-1.237	0.219
Levene	9.305	0.003*	9.695	0.002*	0.075	0.785	0.643	0.425

หมายเหตุ \* ปฏิเสธสมมติฐานว่างที่ระดับนัยสำคัญ 0.05



รูปที่ 2. ACF ของข้อมูลที่กำจัดอิทธิพลของฤดูกาล  $Y^T$  แยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน



รูปที่ 3. ACF ของข้อมูลที่กำจัดอิทธิพลของแนวโน้ม  $Y^S$  แยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

ตารางที่ 7. ทดสอบตัวประมาณค่าพารามิเตอร์ของตัวแบบโดยสถิติการทดสอบที่ของ AR, MA, SAR และ SMA ของตัวแบบบอกซ์-เจนกินส์แยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

ตัวแบบ	กรมสรรพากร		กรมสรรพสามิต		กรมศุลกากร		หน่วยงานอื่น	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
AR(1)	-	-	-	-	-8.30	0.000*	-	-
AR(2)	-	-	-	-	-3.22	0.002*	-	-
MA(1)	15.16	0.000*	13.97	0.000*	-	-	17.10	0.000*
MA(2)	-	-	-	-	-	-	-3.72	0.000*
SAR(1)	-5.18	0.000*	-	-	-	-	-7.58	0.000*
SAR(2)	-	-	-	-	-	-	-4.13	0.000*
SMA(1)	-	-	-	-	-2.94	0.004*	-	-
SMA(2)	-	-	-	-	-3.95	0.000*	-	-
Ln	No		No		No		No	
d=1	Yes		Yes		Yes		Yes	
D=1	Yes		No		No		Yes	

หมายเหตุ \* ปฏิเสธสมมติฐานว่างที่ระดับนัยสำคัญ 0.05

### 3.5 การสร้างตัวแบบการพยากรณ์โดยวิธีพยากรณ์รวม

รายได้รัฐบาลกรมสรรพากร กรมสรรพสามิตและหน่วยงานอื่นใช้ตัวแบบพยากรณ์ 3 วิธียกเว้นตัวแบบบอกรี-เจนกินส์มาสร้างตัวแบบการพยากรณ์รวม และรายได้รัฐบาลกรมศุลกากรใช้ตัวแบบพยากรณ์ 4 วิธีมาสร้างตัวแบบการพยากรณ์รวมวิธี REG สามารถหาค่าถ่วงน้ำหนักของแต่ละวิธีได้ผลในตารางที่ 8

ตารางที่ 8. การประมาณค่าพารามิเตอร์  $b_{Con}$ ,  $b_{Dec\_T}$ ,  $b_{Dec\_L}$ ,  $b_{Smo}$  และ  $b_{Box}$  ของวิธี REG แยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลฝึกฝน

พารามิเตอร์	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
$b_{Con}$	2,985.48	-18,520.91	-4,093.40	-3,210.38
$b_{Dec\_T}$	1.36	-0.64	-0.12	-1.84
$b_{Dec\_L}$	-0.43	1.62	1.10	2.74
$b_{Smo}$	0.04	0.45	-0.18	0.20
$b_{Box}$	-	-	0.66	-

### 3.6 การเปรียบเทียบวิธีการพยากรณ์

จากการพยากรณ์รายได้รัฐบาลทั้ง 4 ส่วน ของประเทศไทย ทั้ง 5 วิธีในหัวข้อ 3.2 ถึง 3.5 นำตัวแบบเหล่านั้นพยากรณ์รายเดือน ตั้งแต่เดือนพฤษภาคม พ.ศ. 2565 ถึงเดือนเมษายน พ.ศ. 2566 แล้วนำมาคำนวณค่า MAPE แต่ละวิธีเพื่อเปรียบเทียบตัวแบบพยากรณ์ที่เหมาะสมได้ผลในตารางที่ 9 เลือกวิธีการพยากรณ์ที่ให้ค่า MAPE ต่ำที่สุด สรุปผลได้ว่าส่วนใหญ่แล้วตัวแบบที่มีความเหมาะสมจะใช้โฮลต์-วินเทอร์รูปแบบบวก คือ กรมสรรพากร กรมสรรพสามิต และหน่วยงานอื่น มีค่าร้อยละ MAPE ที่ 8.08, 14.73 และ 27.46 ตามลำดับ และตัวแบบพยากรณ์ที่เหมาะสมของกรมศุลกากรคือ ตัวแบบพยากรณ์บอกรี-เจนกินส์ มีค่า MAPE ที่ 10.34

ตารางที่ 9. การเปรียบเทียบค่า MAPE ของ 5 วิธีการพยากรณ์ทางสถิติแยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลทดสอบ

วิธีการพยากรณ์	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
แยกส่วนประกอบแนวโน้ม เส้นตรงฤดูกาลรูปแบบบวก	9.72	32.05	20.10	35.68
แยกส่วนประกอบแนวโน้ม ลอการิทึมฤดูกาลรูปแบบบวก	13.13	22.83	16.01	35.50
โฮลต์-วินเทอร์รูปแบบบวก	8.08*	14.73*	13.09	27.46*
บอกรี-เจนกินส์	-	-	10.34*	-
Combined REG	8.51	16.28	13.23	30.39

หมายเหตุ \* ค่า MAPE ต่ำที่สุด

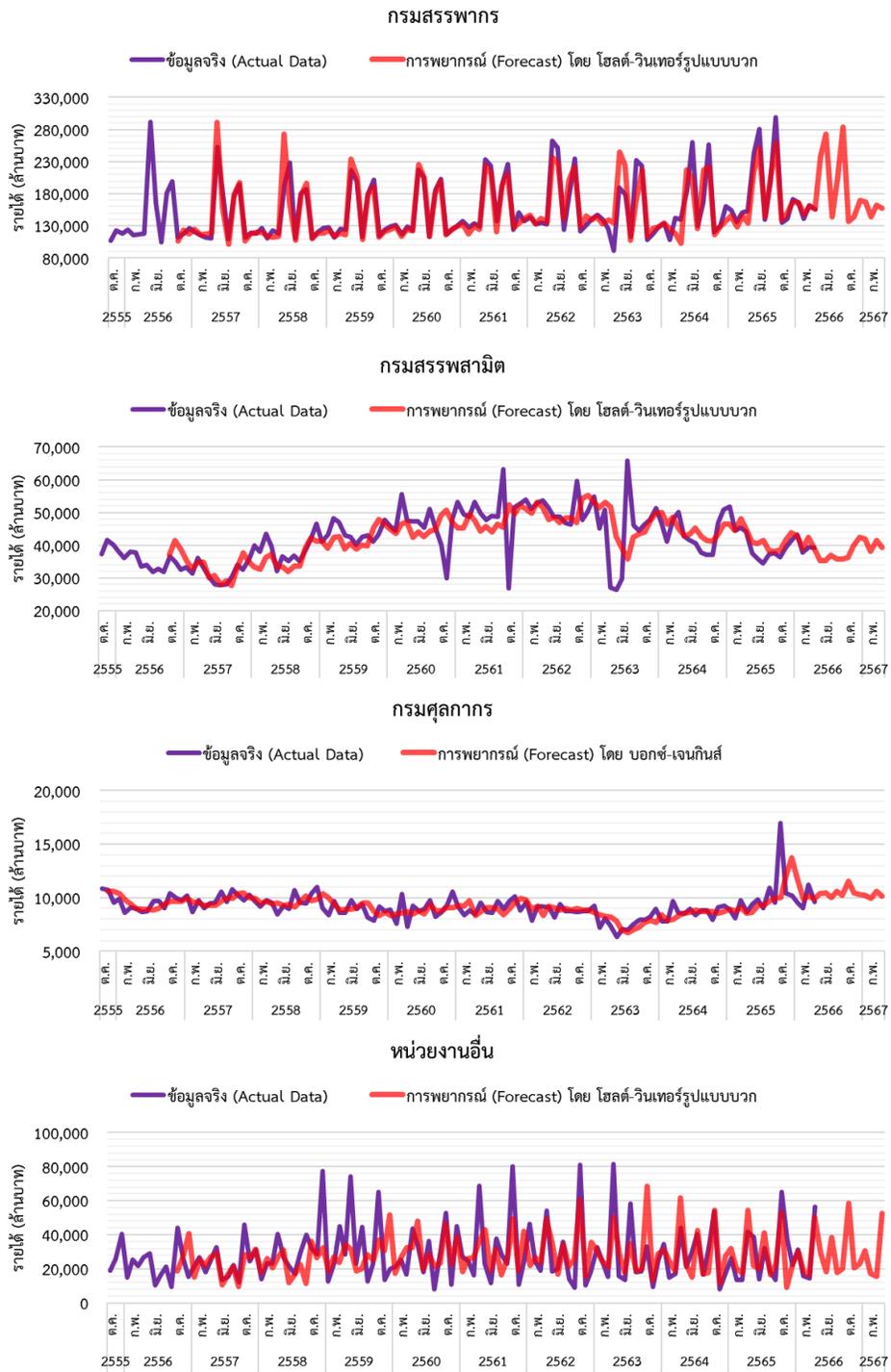
### 3.7 การพยากรณ์รายได้รัฐบาล

ในการพยากรณ์รายได้รัฐบาลทั้ง 4 ส่วน รายเดือนในเดือนพฤษภาคม พ.ศ. 2566 ถึงเดือนเมษายน พ.ศ. 2567 โดยใช้วิธีการพยากรณ์ที่ตัวแบบพยากรณ์ให้ค่า MAPE น้อยที่สุดกับข้อมูลอนุกรมเวลาทั้งหมด ได้ผลการพยากรณ์ในตารางที่ 10 และนำไปเปรียบเทียบกับค่าข้อมูลจริงในรูปที่ 4 ซึ่งจะเห็นว่าข้อมูลจริงและค่าพยากรณ์ของแต่ละรายได้รัฐบาลนั้นสอดคล้องไปด้วยกันดีตามผล MAPE ที่ต่ำซึ่งแสดงในตารางที่ 9 รายได้กรมสรรพากร และหน่วยงานอื่นรูปกราฟเส้นค่าพยากรณ์แสดงให้เห็นรูปแบบอย่างชัดเจนว่ารายได้มีอิทธิพลของฤดูกาล ส่วนกรมสรรพสามิตค่าพยากรณ์ถูกปรับเรียบมาก รูปกราฟเส้นที่แสดงจึงค่อนข้างราบเรียบ และกรมศุลกากรจากตัวแบบพยากรณ์บอกซ์-เจนกินส์ SARIMA(2,1,0)(0,0,2)<sub>12</sub> รูปกราฟเส้นก็ได้รับอิทธิพลของ AR(2) ที่ทำให้ในช่วงเดือนตุลาคม พ.ศ. 2566 มีค่า 11,580 ซึ่งมากกว่าค่าพยากรณ์ในเดือนอื่น ๆ เป็นผลมาจากค่ามากของเดือนตุลาคม พ.ศ. 2565

ตารางที่ 10. ค่าพยากรณ์รายได้รัฐบาลทั้ง 4 ส่วน ตั้งแต่เดือนพฤษภาคม พ.ศ. 2566 ถึงเดือนเมษายน พ.ศ. 2567

เดือน-ปี	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
พฤษภาคม 2566	238,148	35,168	10,353	28,941
มิถุนายน 2566	272,531	35,354	10,429	18,453
กรกฎาคม 2566	144,755	37,022	9,992	38,500
สิงหาคม 2566	204,820	35,853	10,568	18,023
กันยายน 2566	283,958	35,834	10,202	20,064
ตุลาคม 2566	136,348	36,299	11,580	58,570
พฤศจิกายน 2566	144,520	39,825	10,451	20,476
ธันวาคม 2566	170,012	42,376	10,342	22,999
มกราคม 2567	166,773	42,034	10,223	30,558
กุมภาพันธ์ 2567	144,208	38,189	9,969	17,383
มีนาคม 2567	162,416	41,536	10,595	15,803
เมษายน 2567	156,954	39,260	10,184	52,580

หมายเหตุ ค่าน้อย  ค่ามากสำหรับแต่ละส่วนของรายได้รัฐบาล



รูปที่ 4. การเปรียบเทียบค่าจริงกับค่าพยากรณ์โดยวิธีพยากรณ์ที่เหมาะสมรายได้รัฐบาลทั้ง 4 ส่วน

### 3.8 วิจัย

จากผลการวิจัยในการพิจารณาชุดข้อมูลทดสอบค่าร้อยละ MAPE ในตารางที่ 9 ของกรมสรรพากร กรมสรรพสามิต กรมศุลกากร และหน่วยงานอื่นมีค่า 8.08, 14.73, 10.34 และ 27.46 ตามลำดับ ถือว่ามีค่าที่ต่ำโดยเฉพาะรายได้ 3 ส่วนแรก แสดงถึงการพยากรณ์ที่มีความแม่นยำสูง ส่วนหน่วยงานอื่นมีค่าสูงกว่ากรมทั้ง 3 อาจจะเป็นเพราะรายได้ส่วนนี้ได้มาจากหลากหลายส่วนงานที่แตกต่างกันทำให้ความผันผวนของรายได้มีมาก

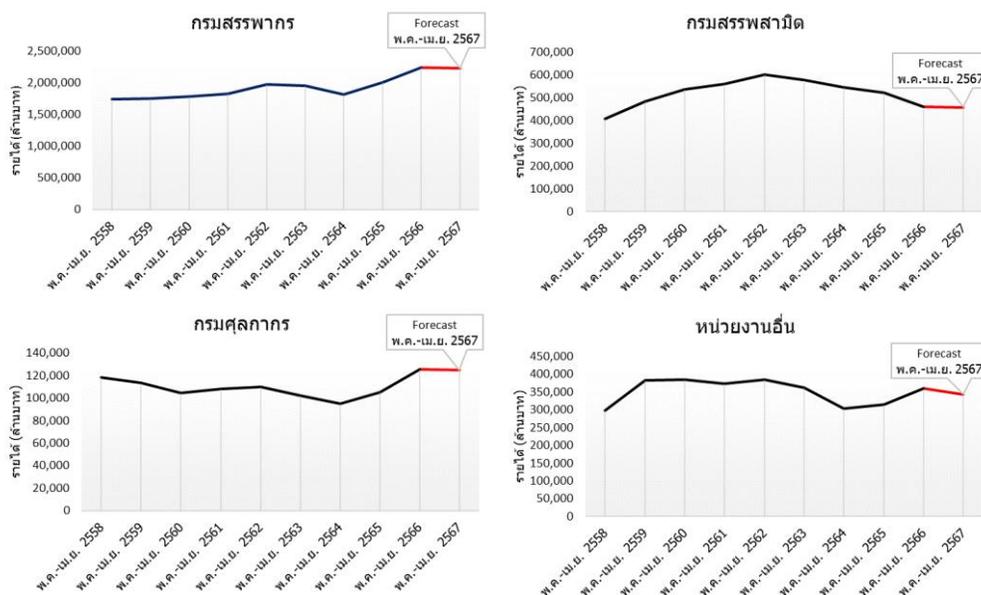
เมื่อนำผลการพยากรณ์มาวิเคราะห์แนวโน้มและฤดูกาลเพิ่มเติมตารางที่ 11 และรูปที่ 5 ทำให้ทราบว่ารายได้รัฐบาลใน 12 เดือน คือ เดือนพฤษภาคม พ.ศ. 2566 ถึงเดือนเมษายน พ.ศ. 2567 มีลักษณะลดลงเล็กน้อย โดยรายได้รัฐบาลของกรมสรรพากร กรมสรรพสามิตและกรมศุลกากร ลดลงไม่ถึงร้อยละ 1 และหน่วยงานอื่นลดลงร้อยละ 4.8 โดยถ้าวิเคราะห์พร้อมทั้งตารางที่ 3 และ 4 ความแปรผันตามฤดูกาลจากวิธีแยกส่วนประกอบจะพบว่า รายได้ของกรมสรรพากรมีโอกาสเก็บได้มากในช่วงเดือนพฤษภาคม มิถุนายน สิงหาคม และกันยายน ซึ่งภาครัฐยังมีโอกาสกระตุ้นเศรษฐกิจในช่วงเวลานี้เพื่อจัดเก็บภาษีให้ได้มากขึ้นในปีงบประมาณ พ.ศ. 2566 ส่วนในปีงบประมาณ พ.ศ. 2567 ช่วงเดือนตุลาคมถึงเดือนเมษายนนั้น กรมสรรพสามิต กรมศุลกากร และรายได้หน่วยงานอื่นเป็นช่วงที่เก็บรายได้ได้จำนวนมาก ภาครัฐควรออกนโยบายกระตุ้นเศรษฐกิจในช่วงนี้เพื่อจัดภาษีและรายได้ให้ได้อย่างน้อยตามเป้าหมาย

ตารางที่ 11. ร้อยละอิทธิพลฤดูกาลและร้อยละการเปลี่ยนแปลงของรายได้รัฐบาลเดือนพฤษภาคม พ.ศ. 2566 ถึงเดือนเมษายน พ.ศ. 2567 เมื่อเทียบกับ 12 เดือนก่อนแยกตามที่มาของรายได้ 4 ส่วน

เดือน-ปี	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
พฤษภาคม 2566	128.41	91.99	99.48	101.44
มิถุนายน 2566	146.95	92.48	100.21	64.68
กรกฎาคม 2566	78.05	96.84	96.01	134.95
สิงหาคม 2566	110.44	93.78	101.55	63.17
กันยายน 2566	153.12	93.73	98.03	70.33
ตุลาคม 2566	73.52	94.95	111.27	205.30
พฤศจิกายน 2566	77.93	104.17	100.42	71.77
ธันวาคม 2566	91.67	110.85	99.37	80.62
มกราคม 2567	89.93	109.95	98.23	107.11
กุมภาพันธ์ 2567	77.76	99.90	95.79	60.93
มีนาคม 2567	87.58	108.65	101.80	55.39
เมษายน 2567	84.63	102.70	97.86	184.30
ร้อยละการเปลี่ยนแปลงเทียบกับ 12 เดือนก่อน	99.72	99.78	99.40	95.20

หมายเหตุ   ค่าร้อยละที่มีค่าเกิน 100

ถ้าภาครัฐใช้นโยบายเช่นเดิมในการจัดเก็บรายได้ ข้อมูลอนุกรมเวลาในอดีตก็จะสะท้อนการคาดการณ์ล่วงหน้า 12 เดือนตามวิธีการทางสถิติที่มีความแม่นยำสูงจากตารางที่ 9 ดังนั้นแนวโน้มค่าพยากรณ์ในรูปที่ 5 เส้นสีแดง พิจารณาได้ว่าการฟื้นตัวระดับคงที่หรือลดลงเล็กน้อยของเศรษฐกิจในสถานการณ์การแพร่ระบาดของโควิด-19 ยังคงส่งผลกระทบต่อรายได้ภาครัฐต่อเนื่องยาวนานไปถึงกลางปี 2567



รูปที่ 5. แนวโน้มราย 12 เดือน ตั้งแต่เดือนพฤษภาคม พ.ศ. 2557 ถึงเดือนเมษายน พ.ศ. 2567 รายได้รัฐบาลทั้ง 4 ส่วน

#### 4. บทสรุป

การวิจัยครั้งนี้ได้นำเสนอการเลือกตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลาข้อมูลรายได้รัฐบาลทั้ง 4 ส่วน คือ กรมสรรพากร กรมสรรพสามิต กรมศุลกากร และหน่วยงานอื่น (ส่วนราชการอื่น กรมธนารักษ์ เงินส่วนเกินจากการจำหน่ายพันธบัตรรัฐบาล แปรรูปรัฐวิสาหกิจ รัฐวิสาหกิจ) ข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิที่รวบรวมจาก ส่วนบริหารจัดการข้อมูลสารสนเทศกระทรวงการคลัง ศูนย์เทคโนโลยีสารสนเทศและการสื่อสาร สำนักงานปลัดกระทรวงการคลัง [1] ตัวแบบพยากรณ์ที่เหมาะสมของกรมสรรพากรคือ โฮลต์-วินเทอร์รูปแบบบวก มีค่า  $\alpha$ ,  $\gamma$  และ  $\delta$  เท่ากับ 0.13, 0.00 และ 0.58 ตามลำดับ ตัวแบบพยากรณ์ที่เหมาะสมของกรมสรรพสามิต คือ โฮลต์-วินเทอร์รูปแบบบวก มีค่า  $\alpha$ ,  $\gamma$  และ  $\delta$  เท่ากับ 0.20, 0.00 และ 0.12 ตามลำดับ ตัวแบบพยากรณ์ที่เหมาะสมของกรมศุลกากรคือ SARIMA(2,1,0)(0,0,2)<sub>12</sub> และตัวแบบพยากรณ์ที่เหมาะสมของหน่วยงานอื่น คือ โฮลต์-วินเทอร์รูปแบบบวก มีค่า  $\alpha$ ,  $\gamma$  และ  $\delta$  เท่ากับ 0.02, 0.00 และ 0.38 ตามลำดับ จะเห็นว่าวิธีโฮลต์-วินเทอร์รูปแบบบวกค่าพารามิเตอร์ที่ได้จากข้อมูลชุดฝึกฝนตารางที่ 5 แตกต่างเล็กน้อยกับการใช้ข้อมูลทั้งหมดเพื่อการพยากรณ์ล่วงหน้า โดยสรุปเทคนิคในการ

พยากรณ์ที่นำเสนอนี้เป็นเทคนิคการพยากรณ์ทางสถิติที่แบ่งออกเป็น 4 กลุ่ม คือ วิธีการแยกส่วนประกอบ วิธีการทำให้เรียบ วิธีการบอกซ์-เจนกินส์ และวิธีการพยากรณ์รวม

อย่างไรก็ตามวิธีการพยากรณ์อนุกรมเวลานั้นยังสามารถใช้เทคนิคอื่น ๆ ทางด้านวิทยาการข้อมูลที่ใช้การเรียนรู้ของเครื่องช่วยในการหาคำตอบ การใช้โครงข่ายระบบประสาทแบบย้อนกลับ (Recurrent Neural Network: RNN) แบบจำลองหน่วยความจำระยะสั้นแบบยาว (Long Short-term Memory: LSTM) [15] และโครงข่ายประตูกลับ (Gated Recurrent Unit: GRU) [16] เป็นเทคนิคการเรียนรู้ของเครื่องที่เป็นที่นิยมในปัจจุบัน เมื่อนำเทคนิคเหล่านี้มาใช้ในการพยากรณ์รายได้รัฐบาล โดยกำหนดค่าของไฮเปอร์พารามิเตอร์ ดังนี้ ฟังก์ชันความสูญเสีย (Loss function) เป็นวิธีคลาดเคลื่อนกำลังสองเฉลี่ย (Mean square error) ตัวเพิ่มประสิทธิภาพ (Optimizer) เป็นวิธี adam ทำการฝึกสอน (Epoch) จำนวน 100 ครั้ง และจำนวนข้อมูลย้อนหลัง (Look back) เท่ากับ 24 ประมวลผลด้วยไพทอน (Python) ได้ค่าร้อยละ MAPE ดังตารางที่ 12

ตารางที่ 12. การเปรียบเทียบค่า MAPE ของ 3 วิธีการพยากรณ์ทางวิทยาการข้อมูลแยกตามที่มาของรายได้ 4 ส่วนในชุดข้อมูลทดสอบ

วิธีการพยากรณ์	กรมสรรพากร	กรมสรรพสามิต	กรมศุลกากร	หน่วยงานอื่น
RNN	28.27	15.02	19.85	57.13
LSTM	26.66	17.86	11.75	50.51
GRU	25.62	18.06	12.56	51.16

ค่าร้อยละของ MAPE ของกรมสรรพากร กรมสรรพสามิต กรมศุลกากร และหน่วยงานอื่นจากตารางที่ 12 นั้น ได้ผลลัพธ์ที่ดีแต่ยังไม่ได้ผลดีกว่าค่าต่ำที่สุดจากวิธีทางสถิติในตารางที่ 9 ดังนั้นวิธีการทางสถิติที่ได้ในการวิจัยครั้งนี้จึงมีความน่าเชื่อถือและเหมาะสม อย่างไรก็ตามผู้ที่สนใจศึกษาวิทยาการข้อมูลอาจจะศึกษาเชิงลึกและทำการทดลองเพิ่มเติมเรื่องการกำหนดค่าของไฮเปอร์พารามิเตอร์ให้ค่าพยากรณ์แม่นยำมากขึ้น

### เอกสารอ้างอิง (References)

- [1] Ministry of Finance Operation Center. 2023. Ministry of Finance. Available at: [https://dataservices.mof.go.th/menu3?id=2&page=&freq=month&mf=10&yf=2565&mt=4&yt=2566&sort=asc&search\\_text=](https://dataservices.mof.go.th/menu3?id=2&page=&freq=month&mf=10&yf=2565&mt=4&yt=2566&sort=asc&search_text=). Retrieved 1 May 2023.
- [2] Tax Planning Office. 2023. The Excise Department. Available at: <https://www.excise.go.th/cs/groups/public/documents/document/dwnt/ndcw/~edis/p/uatucm470306.pdf>. Retrieved 2 May 2023.
- [3] Benvenuto, D. et al. 2020. Application of the ARIMA model on the COVID-2019 epidemic data set. *Data in Brief*, 29, 1-4, <https://doi.org/10.1016/j.dib.2020.105340>.

- [4] Sen, J. and Chaudhuri, T. 2017. A Predictive Analysis of the Indian FMCG Sector using Time Series Decomposition-based Approach. Available at: <http://dx.doi.org/10.2139/ssrn.2992051>. Retrieved 4 May 2023.
- [5] Xiao, L., Wang, J., Dong, Y. and Wu, J. 2015. Combined forecasting models for wind energy forecasting: A case study in China. *Renewable and Sustainable Energy Reviews*, 44, 271–288, <https://doi.org/10.1016/j.rser.2014.12.012>.
- [6] สมบูรณ์ ศิริสรธริรัณ, สิทธิศักดิ์ ไชยสุข, ยุรีพรรณ วณิชโยบล และกัญญภัคพิมพ์. 2565. การฟื้นตัวของประเทศไทยสู่ความยั่งยืนทางการคลังภายหลังสถานการณ์การแพร่ระบาดของโรคโควิด-19. *วารสารคุณภาพชีวิตกับกฎหมาย*, 18(1), 121-145. [Somboon Sirisunhirun, Sitthisal Chaiyasuk, Yureephan Vanichayobol and Kanphakphim Manoonphol. 2022. Thailand's recovering post-COVID-19 pandemic better for fiscal sustainability. *Quality of Life and Law Journal*, 18(1), 121-145. (in Thai)]
- [7] Tibulca, I.L. 2022. The impact of the COVID-19 pandemic on tax revenues in the EU. *Economic Research-Ekonomska Istraživanja*, 35(1), 2442-2459, <https://doi.org/10.1080/1331677X.2021.1954967>.
- [8] Abraham, B. and Ledolter, J. 1983. *Statistical Method for Forecasting*. John Wiley & Sons, New York.
- [9] ทรงศิริ แต่สมบัติ. 2549. การพยากรณ์เชิงปริมาณ. พิมพ์ครั้งที่ 1, สำนักพิมพ์มหาวิทยาลัยเกษตรศาสตร์, กรุงเทพฯ. [Songsiri Taesombat. 2006. *Quantitative Forecasting*. 1<sup>st</sup> ed., Kasetsart University Press, Bangkok. (in Thai)]
- [10] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. 1994. *Time Series Analysis: Forecasting and Control*. 3<sup>rd</sup> ed., Prentice Hall, New Jersey.
- [11] อินทรา เนขุนทด และเขตสิริ คำชอด. 2555. ตัวแบบพยากรณ์อนุกรมเวลาของอัตราผู้ป่วยโรคปอดอักเสบในจังหวัดลำปาง. การประชุมวิชาการนำเสนอผลงานวิจัยระดับชาติ เครือข่ายบัณฑิตศึกษามหาวิทยาลัยราชภัฏภาคเหนือครั้งที่ 18 และลำปางวิจัยครั้งที่ 4, ลำปาง, 340-358. [Intira Nakunthod and Ketsiri Khamkhod. 2018. Time series modeling for the pneumonia rate of patients in Lampang province. Proceeding of the 18<sup>th</sup> Graduate Studies of Northern Rajabhat, University Network Conference and the 4<sup>th</sup> Lampang Research, Lampang, 340-358. (in Thai)]
- [12] จิโรจน์ แซ่ยั้ง, วฐา มินเสน และพิมพกา ธานินพงค์. 2566. ตัวแบบพยากรณ์ปริมาณน้ำไหลเข้าอ่างเก็บน้ำของการไฟฟ้าฝ่ายผลิตแห่งประเทศไทย (กฟผ.). *วารสารวิจัย มหาวิทยาลัยเทคโนโลยีราชมงคลศรีวิชัย*. 15(2), 494-510. [Jiroge Saeying, Watha Minsan and Phimpaka Taninpong. 2023. Forecasting model for the amount of water flowing into the reservoirs of the electricity generating authority of Thailand (EGAT). *RMUTSV Research Journal*, 15(2). 494-510. (in Thai)]

- [13] ตินห์ณภัทร ตุ่นแก้ว, ปรรารถนา มินเสน, เฉลิมรัช นนทะภา และวฐา มินเสน. 2023. ตัวแบบพยากรณ์ที่เหมาะสมของอัตราแลกเปลี่ยนสกุลเงินต่างประเทศ 10 สกุลที่คนไทยนิยมไปท่องเที่ยวมากที่สุดเทียบกับค่าเงินบาท. *วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์*. 31(3). 1-21. [Tinnaphat Tunkaew, Pradthana Minsan, Chalermrat Nontapa and Watha Minsan. 2023. A suitable forecasting model for exchange rates of the top 10 foreign currencies most preferred by Thai tourists compared to the Thai baht. *Thai Science and Technology Journal*, 31(3). 1-21. (in Thai)]
- [14] Montgomery, D.C., Peck, E.A. and Vining, G.G. 2006. Introduction to Linear Regression Analysis. 4<sup>th</sup> ed., Wiley, New York.
- [15] Hochreiter, S. and Schmidhuber, J. 1997. Long short-term memory. *Neural Comput*, 9(8), 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>.
- [16] Cho, K. et al. 2014. Learning phrase representations using RNN encoder–decoder for statistical machine translation. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 1724–1734, <https://doi.org/10.3115/v1/D14-1179>.