

ตัวแบบพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม A Forecasting Model for the Number of Establishments Registered with the Social Security Fund

ดร.วารังคณา เรียนสุทธิ์ (Dr.Warangkhana Riansut)^{1*} ศุภมิตร วิริยกุลโอภาส (Supamit Wiriakulopast)**

ปวิตร อัครวนิชกุล (Pawit Akarawanichakun)*** ปิยศักดิ์ คงทอง (Piyasak Khongthong)***

(Received: November 15, 2022; Revised: December 29, 2022; Accepted: January 4, 2023)

บทคัดย่อ

วัตถุประสงค์ของการศึกษาค้นคว้าครั้งนี้คือ การสร้างตัวแบบพยากรณ์ที่เหมาะสมของจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม โดยใช้วิธีการทางสถิติ 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแดมป์ ข้อมูลรายเดือนที่นำมาวิเคราะห์ได้จากเว็บไซต์ของสำนักงานประกันสังคม ตั้งแต่เดือนมกราคม 2555 ถึงเดือนกันยายน 2565 จำนวน 129 เดือน การตรวจสอบความแม่นยำของตัวแบบพยากรณ์ใช้เกณฑ์ร้อยละค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (MAPE) และเกณฑ์รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ผลการศึกษาพบว่า ตัวแบบพยากรณ์ของวิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์มีความแม่นยำมากที่สุด มีค่า MAPE = 0.6639 และ RMSE = 3,720 ตัวแบบพยากรณ์ คือ

$$\hat{Y}_{t+m} = 490,283.3277 + 691.56405 \left[(m-1) + \frac{1}{0.53102} \right] \text{ เมื่อ } m = 1 \text{ แทนเดือนมกราคม 2565}$$

ABSTRACT

The objective of this study is to construct an appropriate forecasting model for the number of establishments registered with the social security fund. We used four statistical methods: the Box-Jenkins method, Holt's exponential smoothing method, Brown's exponential smoothing method, and the damped trend exponential smoothing method to construct the models. We used monthly data from the Social Security Office website from January 2012 to September 2022 (129 months) to analyze. The forecasting model's accuracy was checked using the mean absolute percentage error (MAPE) and root mean square error (RMSE). The study found that the forecasting model from Brown's exponential smoothing method was the most accurate, with MAPE = 0.6639 and RMSE = 3,720. The forecasting model was $\hat{Y}_{t+m} = 490,283.3277 + 691.56405 \left[(m-1) + \frac{1}{0.53102} \right]$ where $m = 1$ represented January 2022.

คำสำคัญ: การพยากรณ์ อนุกรมเวลา การเปรียบเทียบความแม่นยำ SPSS

Keywords: Forecast, Time series, Accuracy comparison, SPSS

¹Corresponding author: warang27@gmail.com

*รองศาสตราจารย์ สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยทักษิณ

Associate Professor, Department of Mathematics and Statistics, Faculty of Science, Thaksin University

**ครูโรงเรียนวิทยาศาสตร์จุฬารามราชวิทยาลัย นครศรีธรรมราช

Teacher, Chulabhorn Rajavidyalaya Science School, Nakhon Si Thammarat

***นักเรียนชั้นมัธยมศึกษาปีที่ 5 โรงเรียนวิทยาศาสตร์จุฬารามราชวิทยาลัย นครศรีธรรมราช

Student, Mathayom 5, Chulabhorn Science College School, Nakhon Si Thammarat

บทนำ

ประกันสังคม คือ การสร้างหลักประกันในการดำรงชีวิตในกลุ่มของสมาชิกที่มีรายได้ และจ่ายเงินสมทบเข้ากองทุนประกันสังคมเพื่อรับผิดชอบเฉลี่ยความเสี่ยงที่อาจเกิดขึ้นจากการเจ็บป่วย คลอดบุตร ทูพพลภาพ ตาย สงเคราะห์ชราภาพ และการว่างงาน เพื่อให้ได้รับการรักษาพยาบาลและมีการทดแทนรายได้อย่างต่อเนื่อง [1] การศึกษาครั้งนี้ให้ความสนใจกับการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม ซึ่งสถานประกอบการหมายถึง สถานประกอบการทั้งหมด (แห่ง) ที่มีที่ตั้งในประเทศไทยและลงทะเบียนกับกองทุนประกันสังคมตั้งแต่เดือนมกราคม 2555 ถึงเดือนกันยายน 2565 โดยเป็นสถานประกอบการของหน่วยงานภาคเอกชน รวมถึงหน่วยงานราชการที่จะต้องขึ้นทะเบียนให้กับลูกจ้างที่ไม่ใช่ข้าราชการ เช่น ลูกจ้างเหมา ลูกจ้างรายปี และพนักงานราชการ เป็นต้น จากการศึกษาอนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมในช่วงเวลาดังกล่าว พบว่ามีแนวโน้มเพิ่มขึ้น ยกเว้นในช่วงสถานการณ์โควิด 19 ช่วงเดือนมีนาคม 2563 ถึงเดือนเมษายน 2564 ที่มีแนวโน้มลดลง [2] จึงอาจมีคำถามว่าในอนาคตจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมจะเปลี่ยนแปลงไปอย่างไร การพยากรณ์ทางสถิติเป็นเครื่องมือที่สามารถตอบคำถามนี้ได้ เมื่อทำการทบทวนวรรณกรรมที่เกี่ยวข้อง พบว่า ยังไม่เคยมีการศึกษาเกี่ยวกับการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม แต่พบเพียงการศึกษาเกี่ยวกับการประกันสังคม เช่น การศึกษาของ Sirijarupat และ Pimdee [3] ได้ศึกษาตัวแปรที่ส่งผลกับเจตคติต่อระบบประกันสังคมของแรงงานในจังหวัดสมุทรปราการ การศึกษาของ Keerativibool [4] ได้ศึกษาตัวแบบพยากรณ์จำนวนผู้ประกันตนภาคสมัครใจ (มาตรา 39) และการศึกษาของสถาบันวิจัยเพื่อการพัฒนาประเทศไทย [5] ได้ศึกษาโครงการการออกแบบระบบการประกันสังคมที่รองรับการเปลี่ยนแปลงของรูปแบบการทำงานในอนาคต ด้วยเหตุผลดังกล่าว ผู้วิจัยจึงมีความสนใจที่จะพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมด้วยวิธีการทางสถิติ 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแคมป์ โดยวิธีบ็อกซ์-เจนกินส์มีความเหมาะสมกับอนุกรมเวลาที่มีแนวโน้ม และ/หรือ มีส่วนประกอบของความผันแปรตามฤดูกาล มีความแม่นยำในการพยากรณ์สูง เนื่องจากการกำหนดตัวแบบโดยการตรวจสอบคุณสมบัติของฟังก์ชันสหสัมพันธ์ในตัวเอง (Autocorrelation Function: ACF) และฟังก์ชันสหสัมพันธ์ในตัวเองบางส่วน (Partial Autocorrelation Function: PACF) ซึ่งพิจารณาภายใต้อนุกรมเวลาที่คงที่ (Stationary) หรืออนุกรมเวลาที่มีค่าเฉลี่ยและความแปรปรวนเท่ากันทุกช่วงเวลา [6] วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์มีความเหมาะสมกับอนุกรมเวลาที่มีแนวโน้มเป็นเส้นตรงและไม่มีส่วนประกอบของความผันแปรตามฤดูกาล มีค่าคงตัวการทำให้เรียบ 2 ตัว คือ ค่าคงตัวการทำให้เรียบของค่าระดับ (Level: α) และค่าคงตัวการทำให้เรียบของค่าความชัน (Trend: γ) วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์มีความเหมาะสมกับอนุกรมเวลาที่มีแนวโน้มและไม่มีส่วนประกอบของความผันแปรตามฤดูกาล เช่นเดียวกับวิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ แต่มีการกำหนดให้ค่าคงตัวการทำให้เรียบของค่าระดับและของค่าความชันเท่ากัน (α) ดังนั้นอาจกล่าวได้ว่าวิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์เป็นกรณีพิเศษของวิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ [7] และวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแคมป์มีความเหมาะสมกับอนุกรมเวลาที่มีแนวโน้มและไม่มีส่วนประกอบของความผันแปรตามฤดูกาล โดยอัตราการเปลี่ยนแปลงของแนวโน้มไม่ว่าจะเพิ่มขึ้นหรือลดลงจะช้ากว่าอัตราการเปลี่ยนแปลงของแนวโน้มที่เป็นเส้นตรง นั่นคือ ความชันมีค่าลดลงตามเวลา มีค่าคงตัวการทำให้เรียบ 3 ตัว คือ ค่าคงตัวการทำให้เรียบของค่าระดับ (α) ค่าคงตัวการทำให้เรียบของค่าความชัน (γ) และค่าคงตัวการทำให้เรียบของค่าความชันแบบแคมป์ (Damped Trend) (ϕ) [8]

วัตถุประสงค์การวิจัย

เพื่อสร้างตัวแบบพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมด้วยวิธีการทางสถิติ 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแคมป์ จากนั้นคัดเลือกตัวแบบพยากรณ์ที่เหมาะสม สำหรับใช้ในการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม เพื่อเป็นประโยชน์ต่อหน่วยงานต่าง ๆ ทั้งภาครัฐ ภาคธุรกิจเอกชน ภาคประชาชน และภาคส่วนต่าง ๆ ที่เกี่ยวข้องต่อไป

วิธีการวิจัย

1. วิธีการพยากรณ์

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ คือ อนุกรมเวลาจำนวนสถานประกอบการทั้งหมดในประเทศไทย (แห่ง) ที่ลงทะเบียนกับกองทุนประกันสังคม จากเว็บไซต์ของสำนักงานประกันสังคม [2] ตั้งแต่เดือนมกราคม 2555 ถึงเดือนกันยายน 2565 จำนวน 129 เดือน ดำเนินการแบ่งข้อมูลออกเป็น 2 ชุด โดยชุดที่ 1 ควรมีจำนวนข้อมูลร้อยละ 85 – 95 และชุดที่ 2 ควรมีจำนวนข้อมูลร้อยละ 5 – 15 [8] ดังนั้นการศึกษาครั้งนี้ได้แบ่งข้อมูลออกเป็น ชุดที่ 1 คือข้อมูลตั้งแต่เดือนมกราคม 2555 ถึงเดือนธันวาคม 2564 จำนวน 120 เดือน คิดเป็นร้อยละ 93 ใช้สำหรับการสร้างตัวแบบพยากรณ์ด้วยวิธีการทางสถิติ 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแคมป์ ชุดที่ 2 คือ ข้อมูลตั้งแต่เดือนมกราคมถึงเดือนกันยายน 2565 จำนวน 9 เดือน คิดเป็นร้อยละ 7 ใช้สำหรับการตรวจสอบความแม่นยำของตัวแบบพยากรณ์ด้วยเกณฑ์ร้อยละค่าคลาดเคลื่อนสัมบูรณ์เฉลี่ย (Mean Absolute Percentage Error: MAPE) และเกณฑ์รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error: RMSE) ที่ต่ำที่สุด ตัวแบบพยากรณ์ 4 วิธีที่ศึกษาแสดงดังนี้ (ตารางที่ 1)

ตารางที่ 1 ตัวแบบพยากรณ์

วิธีที่	วิธีพยากรณ์	ตัวแบบพยากรณ์
1	บ็อกซ์-เจนกินส์	ARIMA(p, d, q): $\hat{\phi}_p(B)(1-B)^d \hat{\gamma}_q \hat{Y}_t = \hat{\delta} + \hat{\theta}_q(B)e_t$ [9]
2	โฮลต์	$\hat{Y}_{t+m} = a_t + b_t(m)$ [7] โดยที่ $a_t = \alpha Y_t + (1-\alpha)(a_{t-1} + b_{t-1})$, $b_t = \gamma(a_t - a_{t-1}) + (1-\gamma)b_{t-1}$
3	บราวน์	$\hat{Y}_{t+m} = a_t + b_t \left[(m-1) + \frac{1}{\alpha} \right]$ [7] โดยที่ $a_t = \alpha Y_t + (1-\alpha)a_{t-1}$, $b_t = \alpha(a_t - a_{t-1}) + (1-\alpha)b_{t-1}$
4	แคมป์	$\hat{Y}_{t+m} = a_t + b_t \sum_{i=1}^m \phi^i$ [8] โดยที่ $a_t = \alpha Y_t + (1-\alpha)(a_{t-1} + \phi b_{t-1})$, $b_t = \gamma(a_t - a_{t-1}) + (1-\gamma)\phi b_{t-1}$

ความหมายของสัญลักษณ์ต่าง ๆ ในตารางที่ 1 แสดงดังนี้

$\hat{Y}_t$ และ $\hat{Y}_{t+m}$ แทนค่าพยากรณ์ ณ เวลา t และเวลา t + m ตามลำดับ โดยที่ m แทนจำนวนช่วงเวลาที่ต้องการพยากรณ์ไปข้างหน้า

e_t แทนค่าคลาดเคลื่อนจากการพยากรณ์ ณ เวลา t

$\hat{\mu} = \hat{\mu}_p(B)$ แทนค่าคงตัว (Constant) โดยที่ $\hat{\mu}$ แทนค่าเฉลี่ยของอนุกรมเวลาที่คงที่

$\hat{\phi}_p(B) = 1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2 - \dots - \hat{\phi}_p B^p$ แทนตัวดำเนินการถดถอยในตัวอันดับที่ p กรณีไม่มีฤดูกาล (Non-Seasonal Autoregressive Operator of Order p : AR(p))

$\hat{\theta}_q(B) = 1 - \hat{\theta}_1 B - \hat{\theta}_2 B^2 - \dots - \hat{\theta}_q B^q$ แทนตัวดำเนินการเฉลี่ยเคลื่อนที่อันดับที่ q กรณีไม่มีฤดูกาล (Non-Seasonal Moving Average Operator of Order q : MA(q))

t แทนช่วงเวลา ซึ่งมีค่าตั้งแต่ 1 ถึง n_1 โดยที่ n_1 แทนจำนวนข้อมูลในอนุกรมเวลาชุดที่ 1 ($n_1 = 120$)

d แทนลำดับของการหาผลต่าง

B^j แทนตัวดำเนินการถอยหลัง (Backward Operator) ไป j ช่วงเวลา นั่นคือ $B^j Y_t = Y_{t-j}$

a_t และ b_t แทนค่าประมาณ ณ เวลา t แสดงระยะตัดแกน Y และความชันของแนวโน้ม ตามลำดับ

α , γ และ ϕ แทนค่าคงตัวการทำให้เรียบ โดยที่ $0 < \alpha < 1$, $0 < \gamma < 1$ และ $0 < \phi < 1$

วิธีบ็อกซ์-เจนกินส์ขั้นตอนการพิจารณาตัวแบบพยากรณ์ดังนี้

1) พิจารณาอนุกรมเวลาว่ามีลักษณะคงที่หรือไม่ โดยพิจารณาจากกราฟของอนุกรมเวลาเทียบกับเวลา (Y_t, t) กราฟ ACF และกราฟ PACF ถ้าอนุกรมเวลาไม่คงที่ (Non-Stationary) ต้องแปลงอนุกรมเวลาให้คงที่ก่อนที่จะทำขั้นตอนต่อไป เช่น การแปลงข้อมูลด้วยการหาผลต่างเพื่อกำจัดแนวโน้ม ($d = 1, 2, \dots$) หรือการแปลงข้อมูลด้วยการหาผลต่างฤดูกาลเพื่อกำจัดส่วนประกอบของฤดูกาล ($D = 1, 2, \dots$) [10]

2) กำหนดตัวแบบพยากรณ์ที่เป็นไปได้จากกราฟ ACF และ PACF ของอนุกรมเวลาที่มีลักษณะคงที่ นั่นคือ กำหนดค่า p, q, P และ Q พร้อมทั้งประมาณค่าพารามิเตอร์ของตัวแบบ โดยค่าของ q และ Q จะพิจารณาจากกราฟ ACF และค่าของ p และ P จะพิจารณาจากกราฟ PACF ซึ่งค่าของ p และ q คือ จำนวนแท่งสัมพันธ์สหสัมพันธ์แท่งแรก ๆ ที่มีค่าเกินจากขอบเขตความเชื่อมั่นร้อยละ 95 ขณะที่ค่าของ P และ Q คือ จำนวนแท่งสัมพันธ์สหสัมพันธ์ตามฤดูกาลที่มีค่าเกินจากขอบเขตความเชื่อมั่นร้อยละ 95 สำหรับการศึกษาครั้งนี้ พบว่า อนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมไม่มีฤดูกาล จึงพิจารณาเพียงค่าของ p และ q เท่านั้น

3) ตัดพารามิเตอร์ที่ไม่มีนัยสำคัญออกจากตัวแบบพยากรณ์ครั้งละ 1 ตัว แล้วประมาณค่าพารามิเตอร์ใหม่จนกว่าจะได้ตัวแบบพยากรณ์ที่ประกอบด้วยพารามิเตอร์ที่มีนัยสำคัญทั้งหมด

4) คัดเลือกตัวแบบพยากรณ์ที่มีค่าสถิติ Ljung-Box Q ไม่มีนัยสำคัญ ซึ่งจะหมายถึง ตัวแบบพยากรณ์ที่มีความเหมาะสม และตรวจสอบอนุกรมเวลาของค่าคลาดเคลื่อน (e_t) มีการเคลื่อนไหวเป็นอิสระกัน โดยพิจารณาจากกราฟ ACF และ PACF ของ e_t ควรมีค่าสัมพันธ์สหสัมพันธ์ในตัวและสัมพันธ์สหสัมพันธ์ในตัวบางส่วน ค่าคลาดเคลื่อนตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 95 โดยเฉพาะช่วงเวลาที่ 1 - 6

5) นำตัวแบบพยากรณ์ที่ได้จากขั้นตอนที่ 4 ไปพยากรณ์อนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมในอนาคต ตั้งแต่เดือนตุลาคม 2565 ถึงเดือนธันวาคม 2566

ขั้นตอนการสร้างตัวแบบพยากรณ์ด้วยโปรแกรม SPSS version 17 มีรายละเอียดดังนี้

1) ใช้คำสั่ง Analyze → Forecasting → Create Models...

2) ใน Tab Variables นำตัวแปรจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมใส่ในช่อง Dependent Variables: และเลือกวิธีที่จะสร้างตัวแบบพยากรณ์ในช่อง Method: โดยเลือกระหว่าง Exponential Smoothing (การทำให้เรียบด้วยเลขชี้กำลัง) หรือ ARIMA (บ็อกซ์-เจนกินส์) จากนั้นระบุ Criteria... การศึกษาครั้งนี้เลือก Method: เป็น Exponential Smoothing และระบุ Criteria... เป็น Holt's linear trend คือ วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ Brown's linear trend คือ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ Damped trend คือ วิธีการทำ

ให้เรียงด้วยเลขชี้กำลังที่มีแนวโน้มแบบแคมป์ และเลือก Method: เป็น ARIMA ระบุ Criteria... เป็นตัวแบบเริ่มต้นในส่วน Nonseasonal หมายถึง ค่า p, d, q และในส่วน Seasonal หมายถึง ค่า P, D, Q สำหรับการศึกษาครั้งนี้ พบว่าอนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมไม่มีฤดูกาล จึงพิจารณาเพียงค่าของ p, d, q ในส่วน Nonseasonal เท่านั้น โดยที่ค่า p พิจารณาจากกราฟ PACF ค่า d คือ ลำดับที่ของการหาผลต่างเพื่อกำจัดแนวโน้ม และค่า q พิจารณาจากกราฟ ACF

- 3) ใน Tab Statistics เลือก Parameter estimates เพื่อเก็บค่าประมาณพารามิเตอร์
- 4) ใน Tab Save ระบุที่อยู่ที่จะบันทึกค่าประมาณพารามิเตอร์ในช่อง Export Model File
- 5) ใน Tab Options ระบุเวลาสุดท้ายที่ต้องการให้โปรแกรมพยากรณ์ไปถึงในช่อง Forecast Period และเลือก First case after end of estimation period through a specified date สำหรับการศึกษาครั้งนี้ คือ Year 2023 และ Month 12 (ธันวาคม 2566)

2. การตรวจสอบความแม่นยำของตัวแบบพยากรณ์

ตรวจสอบความแม่นยำของตัวแบบพยากรณ์ โดยการเปรียบเทียบจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมชุดที่ 2 ตั้งแต่เดือนมกราคมถึงเดือนกันยายน 2565 ($n_2 = 9$) กับค่าพยากรณ์จากวิธีการพยากรณ์ทั้ง 4 วิธีที่ศึกษา เพื่อคำนวณค่า MAPE และ RMSE โดยตัวแบบพยากรณ์ที่ให้ค่า MAPE และ RMSE ต่ำที่สุด จัดเป็นตัวแบบที่มีความแม่นยำมากที่สุด เนื่องจากให้ค่าพยากรณ์ที่มีความแตกต่างกับข้อมูลจริงน้อยที่สุด สูตร MAPE และ RMSE แสดงดังนี้ [7]

$$MAPE = \frac{100}{n_2} \sum_{t=1}^{n_2} \left| \frac{e_t}{Y_t} \right| \text{ และ } RMSE = \sqrt{\frac{1}{n_2} \sum_{t=1}^{n_2} e_t^2}$$

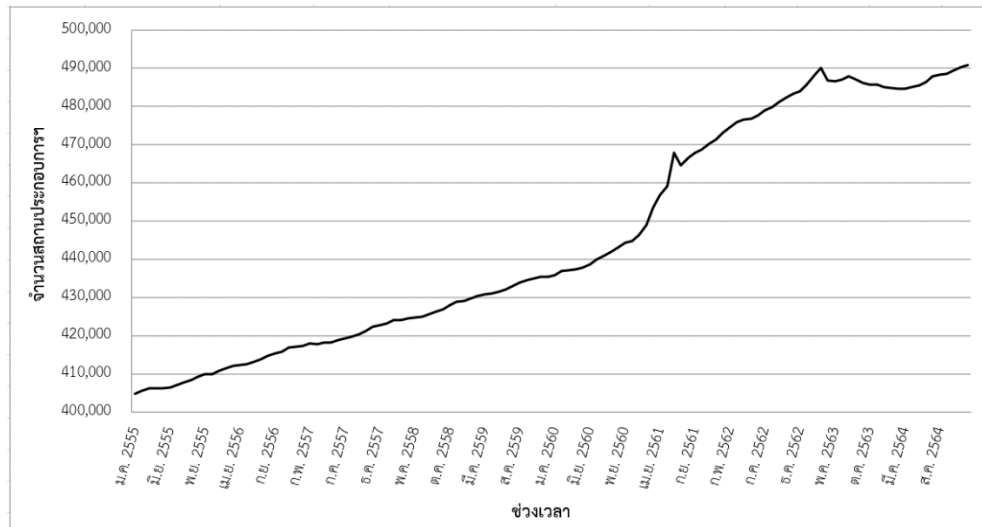
เมื่อ $e_t = Y_t - \hat{Y}_t$ แทนค่าคลาดเคลื่อนจากการพยากรณ์ ณ เวลา t

Y_t และ $\hat{Y}_t$ แทนอนุกรมเวลาและค่าพยากรณ์ ณ เวลา t ตามลำดับ

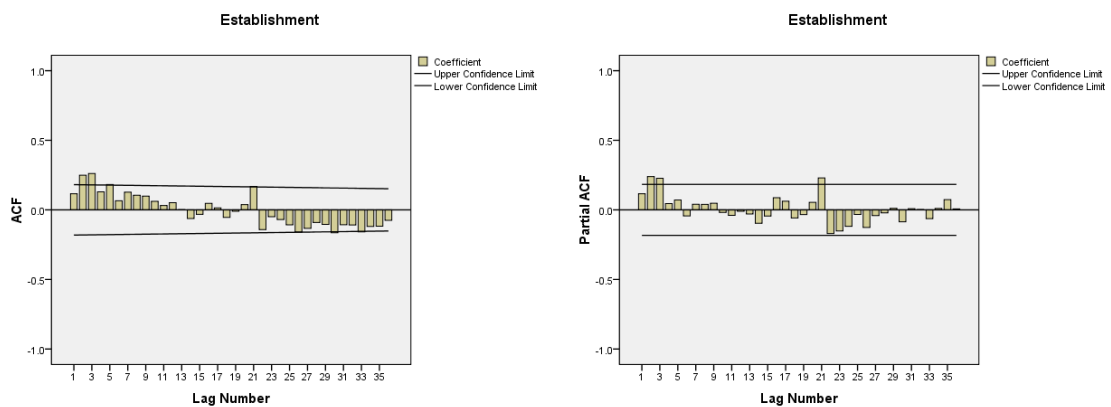
ผลการวิจัย

1. ผลการพยากรณ์

ผลการพิจารณาลักษณะการเคลื่อนไหวของอนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมชุดที่ 1 ตั้งแต่เดือนมกราคม 2555 ถึงเดือนธันวาคม 2564 จำนวน 120 เดือน พบว่า จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมมีแนวโน้มเพิ่มขึ้นอย่างชัดเจน และไม่มีอิทธิพลของฤดูกาล (ภาพที่ 1) ดังนั้นผู้วิจัยจึงแปลงข้อมูลด้วยการหาผลต่างลำดับที่ 1 ($d = 1$) เพื่อกำจัดแนวโน้ม จากนั้นสร้างกราฟ ACF และ PACF ของอนุกรมเวลาที่แปลงข้อมูลแล้ว เพื่อคัดเลือกตัวแบบพยากรณ์ที่เป็นไปได้เริ่มต้นของวิธีบ็อกซ์-เจนกินส์ พบว่า อนุกรมเวลามีลักษณะคงที่ (ภาพที่ 2) จึงกำหนดตัวแบบเริ่มต้น คือ ตัวแบบ ARIMA(2, 1, 2) เนื่องจากกราฟ ACF มีค่าสัมประสิทธิ์สหสัมพันธ์ในตัวของแท่งที่ 2 และ 3 เกินจากขอบเขตความเชื่อมั่นร้อยละ 95 จึงกำหนดให้ $q = 2$ และจากกราฟ PACF มีค่าสัมประสิทธิ์สหสัมพันธ์ในตัวบางส่วนของแท่งที่ 2 และ 3 เกินจากขอบเขตความเชื่อมั่นร้อยละ 95 จึงกำหนดให้ $p = 2$ (ไม่พิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ในตัวบางส่วนของแท่งที่ 21 เนื่องจากอนุกรมเวลาชุดนี้ไม่มีส่วนประกอบของฤดูกาล) เมื่อตัดพารามิเตอร์ที่ไม่มีนัยสำคัญออกจากตัวแบบครั้งละ 1 ตัว แล้วประมาณค่าพารามิเตอร์ใหม่ จะได้ตัวแบบที่ประกอบด้วยพารามิเตอร์ทุกตัวที่มีนัยสำคัญที่ระดับ 0.05 คือ ตัวแบบ ARIMA(1, 1, 1) (ตารางที่ 2)



ภาพที่ 1 ลักษณะการเคลื่อนไหวของอนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม ตั้งแต่เดือนมกราคม 2555 ถึงเดือนธันวาคม 2564



ภาพที่ 2 กราฟ ACF และ PACF ของอนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม เมื่อแปลงข้อมูลด้วยผลต่างลำดับที่ 1

ตารางที่ 2 ค่าประมาณพารามิเตอร์ของตัวแบบ ARIMA(1, 1, 1)

พารามิเตอร์	μ	ϕ_1	θ_1
ค่าประมาณ	710.5966	0.89251	0.75596
t	3.189	9.021	5.211
p-value	0.0018	< 0.0001	< 0.0001

จากตารางที่ 1 สามารถเขียนตัวแบบของวิธีบ็อกซ์-เจนกินส์ที่มีตัวแบบ ARIMA(1, 1, 1) ได้ดังนี้

$$\begin{aligned}
 (1 - \phi_1 B)(1 - B)Y_t &= \delta + (1 - \theta_1 B)\varepsilon_t \\
 (1 - B - \phi_1 B + \phi_1 B^2)Y_t &= (1 - \phi_1 B)\mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} \\
 Y_t &= (1 - \phi_1)\mu + (1 + \phi_1)Y_{t-1} - \phi_1 Y_{t-2} + \varepsilon_t - \theta_1 \varepsilon_{t-1}^*
 \end{aligned}$$

* เนื่องจาก μ เป็นค่าคงตัว ทำให้ $B\mu = \mu$ ดังนั้น $(1 - \phi_1 B)\mu = (1 - \phi_1)\mu$

เมื่อแทนค่าประมาณพารามิเตอร์จากโปรแกรม SPSS จะได้ตัวแบบพยากรณ์ของวิธีบ็อกซ์-เจนกินส์ที่มีตัวแบบ ARIMA(1, 1, 1) วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแคมป์ (ตารางที่ 3)

ตารางที่ 3 ผลการสร้างตัวแบบพยากรณ์

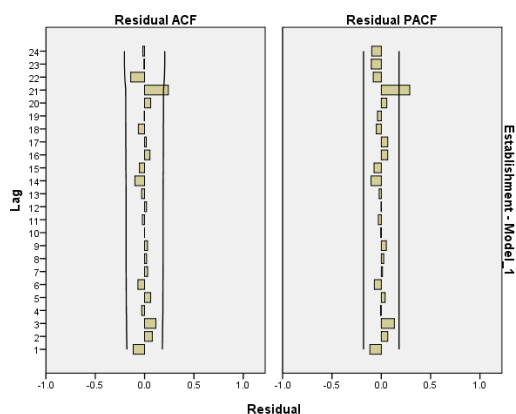
วิธีที่	วิธีพยากรณ์	ตัวแบบพยากรณ์
1	บ็อกซ์-เจนกินส์ ตัวแบบ ARIMA(1, 1, 1)	$\hat{Y}_t = 76.38203 + 1.89251Y_{t-1} - 0.89251Y_{t-2} - 0.75596e_{t-1}$
2	โฮลต์	$\hat{Y}_{t+m} = 490,871.76292 + 617.82857(m)$
3	บราวน์	$\hat{Y}_{t+m} = 490,283.3277 + 691.56405 \left[(m-1) + \frac{1}{0.53102} \right]$
4	แคมป์	$\hat{Y}_{t+m} = 490,865.90948 + 604.34837 \sum_{i=1}^m (0.95661)^i$

โดยที่ Y_{t-j} แทนอนุกรมเวลา ณ เวลา $t-j$, e_{t-1} แทนค่าคลาดเคลื่อนจากการพยากรณ์ ณ เวลา $t-1$ และ $m=1$ แทนเดือนมกราคม 2565

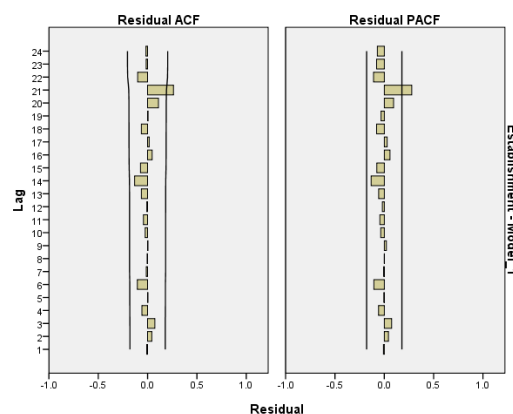
ผลการตรวจสอบค่าสถิติ Ljung-Box Q พบว่า ตัวแบบพยากรณ์ที่สร้างขึ้นทั้ง 4 วิธีมีความเหมาะสมที่ระดับนัยสำคัญ 0.05 ($p\text{-value} > 0.05$) (ตารางที่ 4) และผลการตรวจสอบกราฟ ACF และ PACF ของ e_t พบว่า ค่าคลาดเคลื่อนของทุกวิธีการพยากรณ์เป็นอิสระกัน เนื่องจากค่าสัมประสิทธิ์สหสัมพันธ์ในตัวของค่าคลาดเคลื่อน (กราฟ Residual ACF) และค่าสัมประสิทธิ์สหสัมพันธ์ในตัวบางส่วนของค่าคลาดเคลื่อน (กราฟ Residual PACF) ตกอยู่ในขอบเขตความเชื่อมั่นร้อยละ 95 ยกเว้นช่วงเวลาที่ 21 มีค่าสัมประสิทธิ์สหสัมพันธ์ในตัวและสัมประสิทธิ์สหสัมพันธ์ในตัวบางส่วนของค่าคลาดเคลื่อนเกินจากขอบเขตความเชื่อมั่นร้อยละ 95 ซึ่งไม่มีผลเสียแต่อย่างใด เนื่องจากเป็นความสัมพันธ์ของค่าคลาดเคลื่อนที่ห่างกันถึง 21 ช่วงเวลา นั่นคือ ความสัมพันธ์ของ e_t และ e_{t-21} (ภาพที่ 3-6)

ตารางที่ 4 ค่าสถิติ Ljung-Box Q

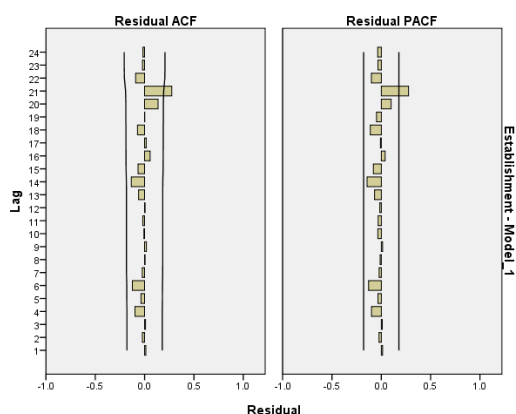
วิธีที่	วิธีพยากรณ์	Ljung-Box Q	p-value
1	บ็อกซ์-เจนกินส์ ตัวแบบ ARIMA(1, 1, 1)	8.388	0.936
2	โฮลต์	7.648	0.959
3	บราวน์	8.500	0.955
4	แคมป์	7.294	0.949



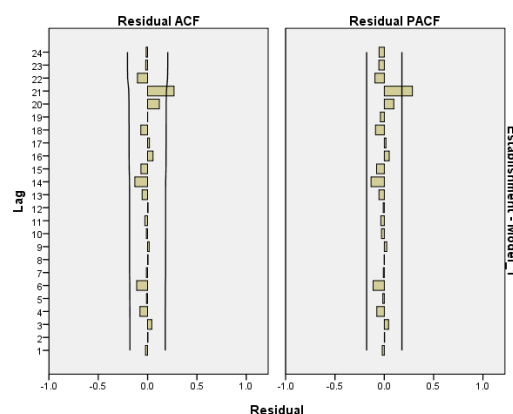
ภาพที่ 3 กราฟ ACF และ PACF ของ e_t จากวิธีบ็อกซ์-เจนกินส์



ภาพที่ 4 กราฟ ACF และ PACF ของ e_t จากวิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์



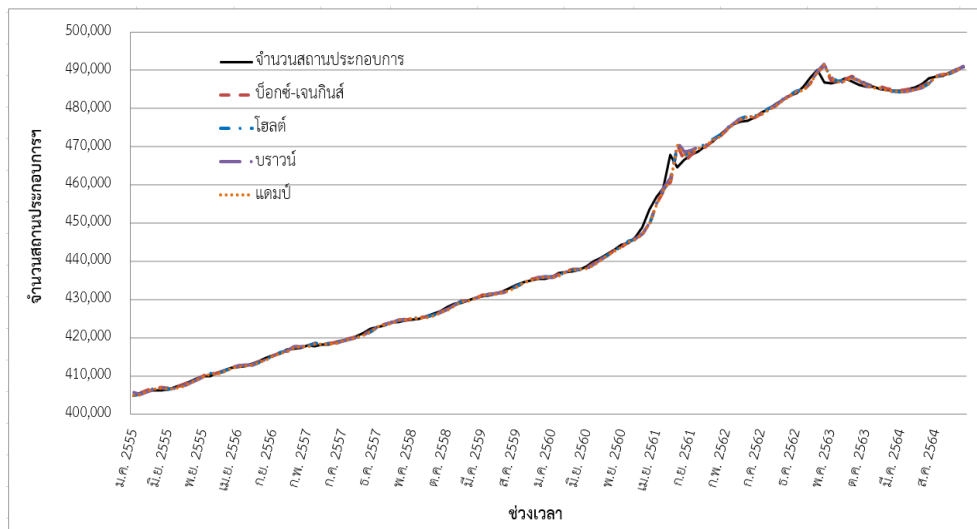
ภาพที่ 5 กราฟ ACF และ PACF ของ e_t จากวิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์



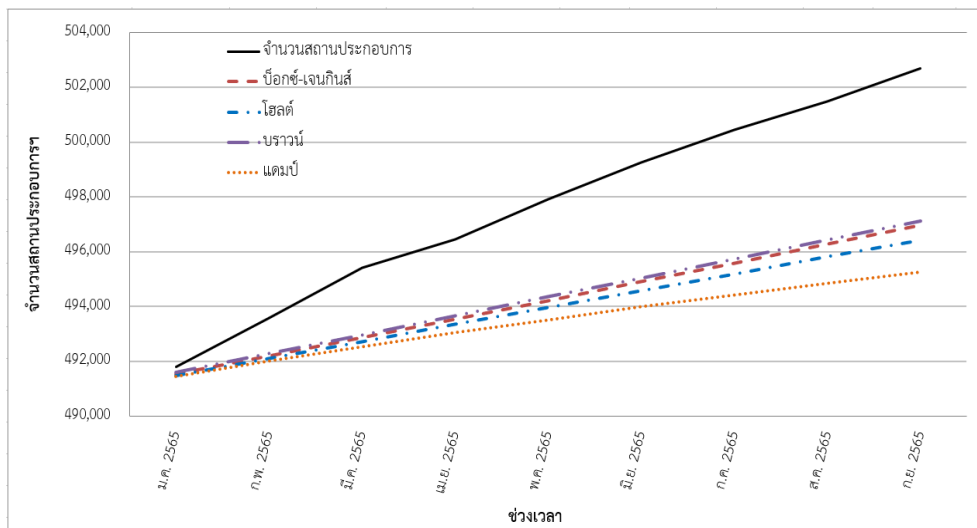
ภาพที่ 6 กราฟ ACF และ PACF ของ e_t จากวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแฉก

2. ผลการตรวจสอบความแม่นยำของตัวแบบพยากรณ์

เมื่อใช้ตัวแบบพยากรณ์ที่สร้างขึ้น (ตารางที่ 3) ในการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมชุดที่ 1 ตั้งแต่เดือนมกราคม 2555 ถึงเดือนธันวาคม 2564 จำนวน 120 เดือน พบว่า ค่าพยากรณ์จากทุกวิธีการพยากรณ์ใกล้เคียงกับค่าจริงมาก (ภาพที่ 7) แต่เมื่อใช้ในการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมชุดที่ 2 ตั้งแต่เดือนมกราคมถึงเดือนกันยายน 2565 จำนวน 9 เดือน พบว่า ค่าพยากรณ์ต่ำกว่าค่าจริงเล็กน้อย โดยกราฟเส้นของค่าจริงจะมีความชันมากกว่ากราฟเส้นของค่าพยากรณ์ จึงทำให้เมื่อเวลาเพิ่มขึ้น ความแตกต่างระหว่างค่าจริงและค่าพยากรณ์เพิ่มมากขึ้น (ภาพที่ 8) ผลการเปรียบเทียบความแม่นยำของตัวแบบพยากรณ์เมื่อใช้ในการพยากรณ์ค่าจริงชุดที่ 2 แล้วคำนวณค่า MAPE และ RMSE พบว่า ตัวแบบพยากรณ์ของวิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์มีความแม่นยำมากที่สุด เนื่องจากมีค่า MAPE และ RMSE ต่ำที่สุด โดยมีความผิดพลาดในการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมร้อยละ 0.66 (MAPE = 0.6639) หรือมีความผิดพลาดในการพยากรณ์ 3,720 แห่ง (RMSE = 3,720) และวิธีการพยากรณ์ที่มีความแม่นยำรองลงมา คือ วิธีบ็อกซ์-เจนกินส์ มีความผิดพลาดในการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมร้อยละ 0.69 (MAPE = 0.6885) หรือมีความผิดพลาดในการพยากรณ์ 3,845 แห่ง (RMSE = 3,845) (ตารางที่ 5)



ภาพที่ 7 การเปรียบเทียบจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมชุดที่ 1 กับค่าพยากรณ์



ภาพที่ 8 การเปรียบเทียบจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมชุดที่ 2 กับค่าพยากรณ์

ตารางที่ 5 ค่า MAPE และ RMSE ของข้อมูลชุดที่ 2

วิธีพยากรณ์	MAPE	RMSE
บ็อกซ์-เจนกินส์	0.6885	3,845
โฮลต์	0.7422	4,154
บราวน์	<u>0.6639</u>	<u>3.720</u>
แดมป์	0.8451	4,776

อภิปรายและสรุปผลการวิจัย

การศึกษารังนี้ได้นำเสนอวิธีการสร้างและคัดเลือกตัวแบบพยากรณ์ที่เหมาะสมกับอนุกรมเวลาจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมของประเทศไทยด้วยวิธีการทางสถิติ 4 วิธี ได้แก่ วิธีบ็อกซ์-เจนกินส์

วิธีการทำให้เรียบด้วยเลขชี้กำลังของโฮลต์ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ และวิธีการทำให้เรียบด้วยเลขชี้กำลังที่มีแนวโน้มแบบแคมป์ ผลการศึกษาพบว่า วิธีที่มีความแม่นยำมากที่สุด คือ วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ มีความผิดพลาดในการพยากรณ์ร้อยละ 0.66 หรือมีความผิดพลาดในการพยากรณ์ 3,720 แห่ง ตัวแบบพยากรณ์แสดงดังนี้

$$\hat{Y}_{t+m} = 490,283.3277 + 691.56405 \left[(m-1) + \frac{1}{0.53102} \right]$$

โดยที่ $m = 1$ แทนเดือนมกราคม 2565

วิธีที่มีความแม่นยำรองลงมา คือ วิธีบ็อกซ์-เจนกินส์ มีความผิดพลาดในการพยากรณ์ร้อยละ 0.69 หรือมีความผิดพลาดในการพยากรณ์ 3,845 แห่ง ตัวแบบพยากรณ์แสดงดังนี้

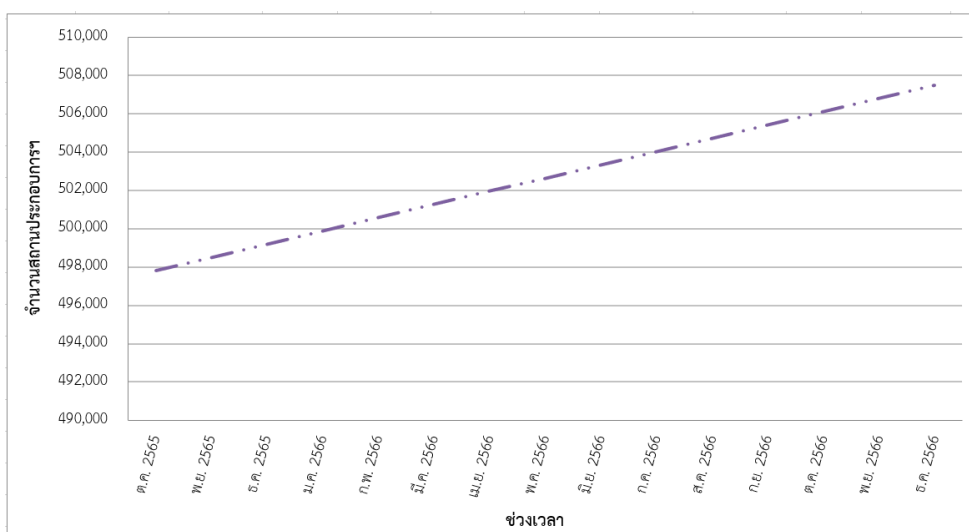
$$\hat{Y}_t = 76.38203 + 1.89251Y_{t-1} - 0.89251Y_{t-2} - 0.75596e_{t-1}$$

โดยที่ Y_{t-j} แทนอนุกรมเวลา ณ เวลา $t - j$ และ e_{t-1} แทนค่าคลาดเคลื่อนจากการพยากรณ์ ณ เวลา $t - 1$

เมื่อใช้วิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์ในการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมในอนาคต ตั้งแต่เดือนตุลาคม 2565 ถึงเดือนธันวาคม 2566 พบว่า จำนวนสถานประกอบการมีแนวโน้มเพิ่มขึ้นอย่างชัดเจน (ตารางที่ 6 และภาพที่ 9) อย่างไรก็ตาม การศึกษาครั้งนี้ได้พิจารณาเพียงปัจจัยเวลาเท่านั้นในการสร้างตัวแบบพยากรณ์ ซึ่งจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมมีการเปลี่ยนแปลงอยู่เสมอ และการเปลี่ยนแปลงอาจเกิดจากปัจจัยอื่น ๆ นอกเหนือจากปัจจัยเวลา เช่น สถานการณ์โควิด 19 ในช่วงเดือนมีนาคม 2563 ถึงเดือนเมษายน 2564 ทำให้จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมลดลง อาจเนื่องมาจากสถานประกอบการประสบกับภาวะขาดทุนหรือเลิกกิจการ ดังนั้นเมื่อมีปัจจัยที่มีผลกระทบต่อการเปลี่ยนแปลงของจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมหรือมีข้อมูลที่เป็นปัจจุบันมากขึ้น ผู้วิจัยควรนำมาปรับปรุงตัวแบบเพื่อให้ได้ตัวแบบพยากรณ์ที่มีความแม่นยำยิ่งขึ้น สำหรับใช้ในการพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมในอนาคต ซึ่งจะประโยชน์ต่อหน่วยงานต่าง ๆ ทั้งภาครัฐ ภาคธุรกิจเอกชน ภาคประชาชน และภาคส่วนต่าง ๆ ที่เกี่ยวข้อง ได้ทราบสถานการณ์และแนวโน้มการเพิ่มขึ้นหรือลดลงของจำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคมต่อไป

ตารางที่ 6 ค่าพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม (แห่ง) ตั้งแต่เดือนตุลาคม 2565 ถึงเดือนธันวาคม 2566 จากวิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์

ช่วงเวลา	ค่าพยากรณ์	ช่วงเวลา	ค่าพยากรณ์	ช่วงเวลา	ค่าพยากรณ์
ต.ค. 2565	497,810	มี.ค. 2566	501,268	ส.ค. 2566	504,725
พ.ย. 2565	498,501	เม.ย. 2566	501,959	ก.ย. 2566	505,417
ธ.ค. 2565	499,193	พ.ค. 2566	502,651	ต.ค. 2566	506,108
ม.ค. 2566	499,884	มิ.ย. 2566	503,342	พ.ย. 2566	506,800
ก.พ. 2566	500,576	ก.ค. 2566	504,034	ธ.ค. 2566	507,492



ภาพที่ 9 การพยากรณ์จำนวนสถานประกอบการที่ลงทะเบียนกับกองทุนประกันสังคม (แห่ง) ตั้งแต่เดือนตุลาคม 2565 ถึงเดือนธันวาคม 2566 จากวิธีการทำให้เรียบด้วยเลขชี้กำลังของบราวน์

เอกสารอ้างอิง

1. Office of Human Resource Management, Chulalongkorn University. Social security fund [Internet]. 2022 [update 2022 Nov 1; cited 2022 Nov 1]. Available from: <https://www.hrm.chula.ac.th/newhrm/กองทุนประกันสังคม/>
2. Social Security Fund. Number of establishments registered with the social security fund from 2012 to 2022 [Internet]. 2022 [update 2022 Nov 1; cited 2022 Nov 1]. Available from: https://www.sso.go.th/wpr/assets/upload/files_storage/sso_th/455f422c652dd17e813abab0542721f5.pdf
3. Sirijarupat P, Pimdee P. Variables affecting attitudes towards the social security system for employees in Samut Prakan province. JIE. 2014; 13(2): 145-152. Thai.
4. Keerativibool W. Forecasting model for the number of insured persons (article 39). RRJ.ST. 2015; 18(1): 21-35. Thai.
5. Thailand Development Research Institute. Social security system design project to support changes in work styles in the future [Internet]. 2022 [update 2022 Nov 1; cited 2022 Nov 1]. Available from: https://www.nesdc.go.th/ewt_w3c/ewt_dl_link.php?nid=11869
6. Taesombut S. Quantitative forecasting. Bangkok: Physics Center; 2006. Thai.
7. Ket-iam S. Forecasting technique. 2nd ed. Songkhla: Thaksin University Book Center; 2005. Thai.
8. Manmin M. Time series and forecasting. Bangkok: Foreprinting; 2006. Thai.
9. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: forecasting and control. 3rd ed. New Jersey: Prentice Hall; 1994.
10. Bowerman BL, O'Connell RT. Forecasting and time series: an applied approach. 3rd ed. California: Duxbury Press; 1993.