The Development of Forecasting Models for Life Insurance Data by Employing Time-series Analysis and Machine Learning Technique

SUPIKA HUADSRI[1], WIKANDA PHAPHAN[1,2]
[1]Department of Applied Statistics, Faculty of Applied Science,
King Mongkut's University of Technology North Bangkok,
Bangkok 10800,
THAILAND

[2]Research Group in Statistical Learning and Inference
King Mongkut's University of Technology North Bangkok
Bangkok 10800
THAILAND

Abstract: - This article is conducted with the primary objective of investigating and comparing various forecasting models, aiming to identify the optimal model for life insurance data. For this investigation, we have employed a comprehensive dataset containing monthly direct premium data from the Thai life insurance sector, spanning from January 2003 to December 2022. Our approach involves the development of time-series models to forecast direct premiums, initially employing the SARIMAX framework. Subsequently, we have introduced an additional time-series forecasting model that incorporates SVR, collectively referred to as the SVR-SARIMAX model. The evaluation criteria used for model comparison encompass the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). The results of our analysis demonstrate that the SARIMAX model outperforms both the SVR and SVR-SARIMAX models, primarily due to the linear pattern in the relationship between the independent and dependent variables. Nevertheless, it is noteworthy that the proposed SVR-SARIMAX model exhibits an improvement in prediction accuracy compared to the standalone non-linear model (SVR), even though the linear model (SARIMAX) still demonstrates superior accuracy.

Key-Words: - Combined Model, Hybrid Model, Support Vector Regression, SARIMAX, Time Series Forecasting, Life Insurance Business Growth.

Received: October 8, 2023. Revised: December 15, 2023. Accepted: February 17, 2024. Published: March 22, 2024.

1 Introduction

In the contemporary context, heightened uncertainties in people's lives underscore the significance of life insurance, offering a means to mitigate emerging risks. Life insurance enterprises play a pivotal role by collecting premiums from individuals seeking coverage, subsequently protecting returns. Presently, life insurance manifests in three primary types: ordinary life insurance catering to middle to high-income earners, industrial life insurance tailored for middle to low-income earners, and group life insurance primarily covering a company's employees with good premium rates. The intricacies of life insurance extend across five categories: term life insurance, whole life insurance, endowment life insurance, annuity life insurance, and investment-unit link life insurance.

The financial sustenance of life insurance businesses heavily relies on premiums, constituting a substantial portion of their revenue. Consequently, premium trends serve as key indicators reflecting the growth of life insurance businesses. However, the utilization of statistical data from the Office of Insurance Commission (OIC) website, [1], presented challenges, stemming from incomplete information and inaccuracies in premium data over specific years. These challenges may result from errors in data collection processes, the data analysis program, or other potential factors.

The growth of life insurance is mainly dependent on the risk of insured people, [2], they analyze the hidden correlations among variables and use them for the risk calculation of an individual customer in the life insurance business. Widely utilized data mining techniques are employed, [3], to identify fraudulent claims in auto insurance, and, [4], analyze the factors affecting the demand for life insurance using descriptive statistics and a panel data model. In addition, [5], investigate the socio-demographic determinants of household expenditures on life insurance in Malaysia employing Cragg's two-part regression model, [6], predict policyholders' lapse decision of life insurance contracts using the random forest and the logistic model, [7], predicted risk in the life insurance business with supervised learning algorithms. In light of these reviews, statistical methods and machine learning are interesting methods for analyzing insurance data.

Our research methodology involves forecasting direct premiums using time series data on monthly life insurance business premiums in Thailand spanning January 2003 to December 2022. Recognizing the limitations of this data, characterized by its chronological nature and scarcity of data, thus this article aims to exploration of suitable time series forecasting models and the machine learning technique for predicting future direct premiums in the life insurance businesses with the ultimate goal is forecasting the growth of the life insurance business in Thailand through direct premiums.

2 Materials and Methods

2.1 Related Research

The ensemble machine learning models and SARIMAX to anticipate particulate matter (PM2.5) in Bangkok, Thailand published by [8]. Support vector regression (SVR), XGBoost, K-nearest neighbors, random forests, artificial neural networks, and many more machine learning models are included in the methodologies. The results indicate that the random forest model obtains the highest Pearson correlation coefficient (PCC) the lowest mean absolute error (MAE), and the root mean squared error (RMSE) in the training data. Nonetheless, the prophet and gradient boosting models outperform the other candidate models in the test data.

A decomposition method with SARIMA and the decomposition method with SARIMAX models developed by [9], indicate that the decomposition method with the SARIMAX model outperforms the decomposition method with SARIMA with the lowest mean absolute percentage error (MAPE).

A decomposition method with the SARIMAX model and Artificial Neural Network (ANN), named DEC-SARIMAX-ANN introduced by [10]. A comparative analysis was conducted with ANN, SARIMA, SARIMAX, DEC-SARIMA, DEC-SARIMAX, DEC-SARIMA-ANN, and DEC-SARIMAX-ANN. The results indicate that the DEC-SARIMAX-ANN performs effectively and exhibits the lowest MAPE. They conclude that the combined model will achieve more accurate forecasting compared to a single forecasting method.

The daily electricity consumption in Thailand using a multiple regression model, an ANN, and an SVR introduced by [11], they suggested that the SVR is the best model to forecast the daily electricity consumption in Thailand.

2.2 Data Description and Preparation

All data were collected from online sources such as the Office of Insurance Commission (OIC), [1] and the Office of the Economic Development Council and National Society. The information includes monthly data on Thailand's life insurance business from January 2003 to December 2022, see all factors in Table 1. Tables 2 provide their basic statistical measurements.

Since the dataset is time series data with relatively limited information. Therefore, the data will be divided into two parts for the modelling process. The first part is the training set data, which includes information on direct insurance premiums in the life insurance business from January 2003 to December 2021. The second part is the test set data, which includes information from January 2022 to December 2022.

The training set data will be used for creating forecasting models, while the test set data will be employed to evaluate the performance of these models to identify the most suitable model for predicting the growth of the life insurance business in Thailand via direct premiums.

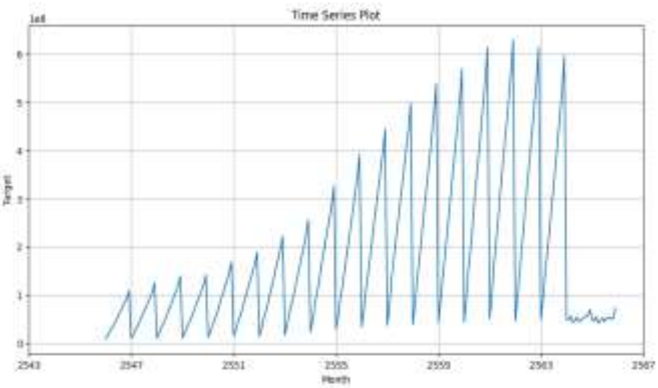

Fig. 1: Time series plots for the life insurance (January 2003 to December 2022)

Figure 1 depicts a time series plot of direct premiums for life insurance businesses in Thailand, demonstrating multiplicative seasonal variation.

2.3 SARIMAX

The SARIMAX model, [12], [13], which stands for Autoregressive Integrated Moving Average with Exogenous Variable Model, is used to forecast time series data with seasonal variation by selecting a suitable forecasting model. This model treats the correlation function as a subset of time series data and handles non-stationary data by introducing variance to convert it to a stationary state. In addition, the SARIMAX model considers external influences when correcting for anomalous data or outliers. External factors, also known as exogenous variables, are analyzed using Multiple Linear Regression (MLR) with the following equations:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + ... + \beta_k x_{k,t} + w_t , \quad (1)$$

here w_t is the stochastic residual, which is a proxy for the variables that may affect the Y variable but is not included in the regression model the time series of the error (ε_t) can be written in terms of the ARIMA model as follows:

$$w_t = \frac{\theta_q(B)\Theta_Q(B^s)}{\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D}\varepsilon_t . \quad (2)$$

The model of SARIMAX (p, d, q)(P,D,Q)s has the following equations :

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + ... + \beta_k x_{k,t}$$

$$+ \frac{\theta_q(B)\Theta_Q(B^s)}{\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D}\varepsilon_t . \quad (3)$$

2.4 SVR

The Support Vector Regression (SVR), [14], is a technique that uses support methods. Support Vector Machine (SVM), [15], is used to analyze the regression between input vectors and output variables, which can be used for time series forecasting, [16], by changing the class classification. SVR is a method of predicting values using SVM. The goal is to find a linear relationship between the n-dimensional input vector $n(x \in R^n)$ and the output variable $Y \in R$ and because SVR is modified from SVM, therefore, the SVR regression equation is similar to the hyperplane equation of SVM as follows:

$$f(x) = w^T \varphi(x) + b . \quad (4)$$

The coefficients w and b are estimated by minimizing the regularized risk function.

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(f(x), y\right) + \frac{1}{2}\|w\|^2 \quad (5)$$

Or

$$R(C) = C\frac{1}{N}\sum_{i=1}^{N} L_\varepsilon\left(f(x), y\right) + \frac{1}{2}w^T w \quad (6)$$

In this article, ε is the insensitive loss function is used to find the Hyperplane High dimensional feature space equation by estimating the maximum distance of data, [17].

$$L_\varepsilon\left(f(x), y\right) = \begin{cases} |f(x) - y| - \varepsilon, & |f(x) - y| \geq \varepsilon \\ 0, & otherwise \end{cases} . \quad (7)$$

SVR aims to find the lowest value.

$$R(C) = C\sum_{i=1}^{N}(\xi + \xi^*) + \frac{1}{2}\|w\|^2 \quad (8)$$

Using the Lagrange multipliers (ξ, ξ^*) to solve the problem of Equation (8) with conditions according to Equation (7). Therefore, the estimated function from the training data set will be able to create SVR equations to predict values. Output from the input vector from Equation (4), where the weight vector (w) is as in Equation (9).

$$w = \sum_{t=1}^{N}(\beta_i^* - \beta_i)\varphi(x_i) , \quad (9)$$

for $x_i, i = 1, 2, ..., n$.

The parameter can be considered a threshold and plays a role in the trade-off between the empirical risk and the flatness. Moreover, C and ε are both predefined and have a significant impact on the predicting performance. Using the Lagrange multiplier and the Karush-Kuhn-Tucker criteria will yield the SVR function's general form:

$$R(\beta_i^* - \beta_i) = \sum_{i=1}^{N}(\beta_i^* - \beta_i) \times K(x_i, x_j) + b . \quad (10)$$

The value of $K(x_i, x_j)$ equals the inner product of vectors x_i and x_j the feature space, $\varphi(x_i)$ and $\varphi(x_j)$. The Gaussian radial basis function (RBF) is the most common kernel function which is written as:

$$K(x_i, x_j) = \exp\left(\frac{-\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \quad (11)$$

SVR has been more widely used in research due to its high accuracy and applicability to data with linear and non-linear relationships. Additionally, the SVR method has a fast processing speed and is suitable for small data sets. Therefore, this article selects the SVR for predicting life insurance data.

2.5 Performance Criterion

This article uses the mean absolute percentage error (MAPE), [18], coefficient of determination (R^2), and root-mean-squared error (RMSE) to evaluate the performance of the forecasting models and they are as follows:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}{n}}, \quad (12)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100, \quad (13)$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_i - \hat{y}_t)^2}{\sum_{t=1}^{n}(y_t - \overline{y})^2}. \quad (14)$$

Here y_t represents actual data at time t.
$\hat{y}_t$ represents the predicted value at time t.
$\overline{y}$ represents the average of actual data.
n represents the number of actual data.

The better model will have the highest R^2 value, lowest RMSE, and lowest MAPE.

2.6 Software Utilized

The results presented in this article were acquired through the utilization of Python in Google Colab. The principal functions employed in this research are enumerated as follows:

- The pmdarima library was used for auto_arima.
- The statsmodels.tsa.statespace.sarimax library was utilized for forecasting data with the SARIMAX model.
- The statsmodels.api library was employed for the estimation of various statistical models.
- sklearn.svm was used for data forecasting with the SVR model.
- sklearn.metrics were applied for calculating RMSE, MAPE, and R^2.

3 Combined Forecast Model

A classic statistical model that can capture a linear trend is the time-series model such as the SARIMAX model which is used when the data has many independent variables. However, nonlinear patterns in time series data are captured using machine learning algorithms like SVR. To capture both linear and nonlinear patterns, this article then proposes the idea of combined predictions, which involves using a machine learning technique in conjunction with a time-series model. To create a combined forecast model, this combination is achieved by choosing appropriate weights for each forecasting technique. The combined forecast model can be expressed using the following formula:

$$SVM - SARIMAX = \sum_{j=1}^{m} w_j \hat{y}_{jt} \quad (15)$$

This article uses the simple average method for weighting two forecasting techniques. This method gives each forecasting technique equal weight when combined, hence the combined forecast model formula becomes:

$$SVM - SARIMAX = \frac{\sum_{j=1}^{m}\hat{y}_{jt}}{2} \quad (16)$$

Consequently, the accuracy of the combined forecast is contingent on the average of two forecasting values from two forecasting techniques. For instance, at time t=1, the resultant model obtained by combining SARIMAX and SVR is:

$$y_{SVR_1 - SARIMAX_1} = \frac{45,883,424 + 47,992,221}{2} \quad (17)$$

Then the combined forecast model procedure is shown in Figure 2.

4 Results and Discussion

This article seeks to assess the efficacy of forecasting insurance data by comparing the Combined Model, Support Vector Regression, and SARIMAX models. The analysis of the data was conducted using Python in Google Colab.

Consequently, the research findings were divided according to the following research methodologies.

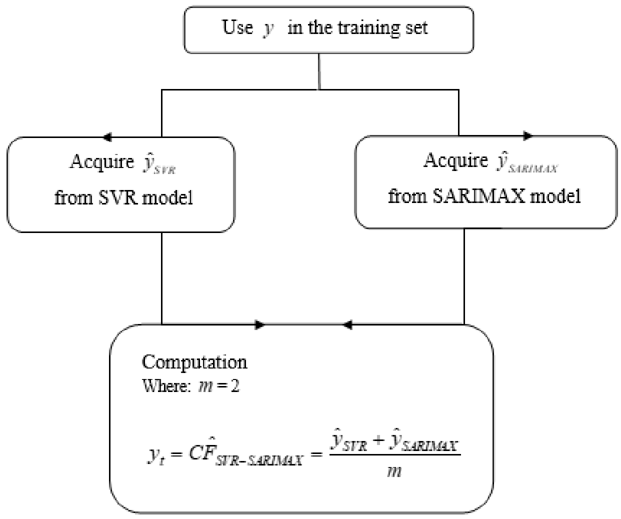

Fig. 2: Diagram for forecasting in the combined model

Figure 3 illustrates the Pearson correlation of each feature. Pearson correlation indicates the strength of linear association, ranging from -1 to 1, and illustrates that X2, X4, X6, X7, X9, X10, X11, and X12 exhibit a high association with Y. Meanwhile, X3 and X5 demonstrate a moderate association with Y, and X1 and X8 show a low association.

Table 3 displays the evaluation of forecast accuracy for SARIMAX models applied to life insurance data. The best-performing SARIMAX model was SARIMAX(2, 1, 2)x(1, 0, 2, 3), with RMSE and MAPE values of 3,508,870.19 and 5.62, respectively, representing the lowest error in the insurance data. Furthermore, this model achieved the highest R^2 value of 0.78 compared to others. Notably, the differences in R^2 values were minimal among the models, indicating their competitiveness. Thus, the SARIMAX(2, 1, 2)x(1, 0, 2, 3) model stands out as the best choice for forecasting in the context of insurance data using the SARIMAX approach.

Figure 4 checks the assumption of error of the dependent variable for time series data. It is clear that the variance of errors is constant and the average variance is equal to zero. Additionally, the errors also have a normal distribution.

Table 4 reveals that the R^2 values for SVR, SARIMAX, and SVR-SARIMAX models are 0.1820, 0.7802, and 0.6007, respectively. Similarly, the RMSE values distinctly indicate that SARIMAX models outperform both SVR and SVR-SARIMAX models. Table 5 and Figure 5 reported that the predicted values of SARIMAX are closer to the actual value than both SVR and SVR-SARIMAX models.

Table 6 presents forecasts for actual values 12 months ahead in 2023 utilizing SVR, SARIMAX, and SVM-SARIMAX models by each independent variable are average over the past three months.

5 Conclusion

In conclusion, the results presented in this article suggest that the linear forecasting model (time series model) may be more suitable than non-linear forecasting models (machine learning) for life insurance data, as the independent variables exhibit a high association with the dependent variable. Moreover, a proposed model, combining a linear forecasting model with non-linear forecasting models, demonstrates an enhancement in prediction accuracy compared to standalone non-linear models, even though linear models still display higher accuracy. It is conceivable that the non-linear forecasting model may not be optimal, and there could exist other non-linear forecasting models that might be more suitable for life insurance data than SVR.

For future research, it is recommended to explore other machine learning techniques, such as the Multilayer Perceptron and Recurrent Neural Network, to further enhance the accuracy of predictive models. Additionally, interval forecasting like [19] and [20] cloud be extended.

Acknowledgement:
The authors would like to express their gratitude to King Mongkut's University of Technology North Bangkok for supporting funding. Contract no. KMUTNB-67-BASIC-16.

References:

[1] Office of Insurance Commission (OIC), March 30, 2023. [online]. Available https://www.oic.or.th/th/industry/statistic/.

[2] V. Hiwase and Avinash Agrawal, Review on application of data mining in life insurance, *International Journal of Engineering & Technology*, Vol. 7, 2018, pp. 159-162.

[3] Teerawat Simmachan, Weerapong Manopa, Pailin Neamhom, Achiraya Poothong and Wikanda Phaphan, Detecting Fraudulent Claims in Automobile Insurance Policies by Data Mining Techniques, *Thailand Statistician*, Vol. 21, No.3, 2023, pp. 552-568.

[4] Sara Emamgholipour, Mohammad Arab, Zahra Mohajerzadeh, Life insurance demand: Middle East and North Africa, *International Journal of Social Economics*, Vol. 44, 2017, pp. 521-529.

[5] Andrew Tan, Steven Yen, Abdul Hasan, and Kamarudin Muhamed, Demand for Life Insurance in Malaysia: An Ethnic Comparison Using Household Expenditure Survey Data, *Asia-Pacific Journal of Risk and Insurance*, Vol.8, No. 2, 2014, pp. 179-204.

[6] Michele Azzone, Emilio Barucci, Giancarlo Moncayo, and Daniele Marazzina, A machine learning model for lapse prediction in life insurance contracts, *Expert Systems with Applications*, Vol. 191, 2022, pp. 116261.

[7] Noorhannah Boodhun, and Manoj Jayabalan, Risk prediction in life insurance industry using supervised learning algorithms, *Complex & Intelligent Systems*, Vol.4, 2018, pp. 145–154.

[8] Patchanok Srisuradetchai, and Wararit Panichkitkosolkul, *Using Ensemble Machine Learning Methods to Forecast Particulate Matter (PM2.5) in Bangkok, Thailand*, In Surinta, O., Kam Fung Yuen, K. (eds) Multi-disciplinary Trends in Artificial Intelligence, Lecture Notes in Computer Science, Springer, 2022, pp. 204-215.

[9] Chalermrat Nontapa, Chainarong Kesamoon, Nicha Kaewhawong, and Peerasak Intrapaiboon, *A New Time Series Forecasting Using Decomposition Method with SARIMAX Model*, In Yang, H., Pasupa, K., Leung, A.CS., Kwok, J.T., Chan, J.H., King, I. (eds) Neural Information Processing, Communications in Computer and Information Science, Springer, 2020.

[10] Chalermrat Nontapa, Chainarong Kesamoon, Nicha Kaewhawong and Peerasak Intrapaiboon, A New Hybrid Forecasting Using Decomposition Method with SARIMAX Model and Artificial Neural Network, *International Journal of Mathematics and Computer Science*, Vol.16, No.4, 2021, pp.1341-1354.

[11] Warut Pannakkong, Thanyaporn Harncharnchai, Jirachai Buddhakulsomsiri, Forecasting Daily Electricity Consumption in Thailand Using Regression, Artificial Neural Network, Support Vector Machine, and Hybrid Models, *Energies*, Vol. 15, No. 9, 2022, pp. 3105.

[12] Yupaporn Areepong, and Rapin Sunthornwat, Forecasting modeling of the number of cumulative COVID-19 cases with deaths and recoveries removal in Thailand, *Science, Engineering and Health Studies*, Vol. 15, 2021, pp. 21020004.

[13] Ahmed Elshewey, Mahmoud Shams, Abdelghafar Elhady, Samaa Shohieb, Abdelaziz Abdelhamid, Abdelhameed Ibrahim and Zahraa Tarek, A Novel WD-SARIMAX Model for Temperature Forecasting Using Daily Delhi Climate Dataset, *Sustainability*, Vol. 15, No. 1, 2023, pp. 757.

[14] Haiying Huang, Wuyi Zhang, Gaochao Deng, James Chen, Predicting Stock Trend Using Fourier Transform and Support Vector Regression, In Proceedings of 2014 IEEE 17th International Conference on Computational Science and Engineering, Chengdu, China, 2014, pp. 213-216.

[15] Cortes Corinna, and Vapnik, Vladimir, Support-vector networks, *Machine Learning*, Vol.20, No. 3, 1995, pp. 273–297.

[16] Hanifah Muthiah, Umu Sa'adah, and Achmad Efendi, Support Vector Regression (SVR) Model for Seasonal Time Series Data, In Proceedings of the Second Asia Pacific International Conference on Industrial Engineering and Operations Management, Surakarta, Indonesia, 2021, pp. 3191-3200.

[17] Esperanza Gonzalo1, Zulima Muñiz, Paulino Nieto , Antonio Sánchez and Marta Fernández, Hard-Rock Stability Analysis for Span Design in Entry-Type Excavations with Learning Classifiers, *Materials*, Vol. 9, No.7, 2016, pp. 531.

[18] Arnaud Myttenaere, Boris Golden, Bénédicte Grand, and Fabrice Rossi, Mean absolute percentage error for regression models, *Neurocomputing*, Vol.192, 2016, pp. 38-48.

[19] Patchanok Srisuradetchai, A Novel Interval Forecast for K-Nearest Neighbor Time Series: A Case Study of Durian Export in Thailand, *IEEE Access*, Vol. 12, 2024, pp. 2032-2044.

[20] Patchanok Srisuradetchai, Wikanda Phaphan, Using Monte-Carlo Dropout in Deep Neural Networks for Interval Forecasting of Durian Export, *WSEAS Transactions on Systems and Control*, Vol. 19, 2024, pp. 10-21.

Table 1. Description of Variables

Variable Name	Measurement Level	Description	Unit
Y	Ratio Scale	Number of direct premiums received from the life insurance business in Thailand at month t	baht
X1	Ratio Scale	The policy interest rate at month t	percentage
X2	Ratio Scale	The sum insured as of month t	thousand baht
X3	Ratio Scale	The number of life insurance companies as of month t	companies
X4	Ratio Scale	The number of life insurance policies as of month t	income
X5	Ratio Scale	The gross domestic product at month t	billions of baht
X6	Ratio Scale	The advertising expenditures of life insurance companies	thousand baht
X7	Ratio Scale	The number of first-year premiums of the policy Ordinary life insurance at month t	thousand baht
X8	Ratio Scale	The number of first-year premiums of the policy Industrial life insurance as of month t	thousand baht
X9	Ratio Scale	The number of first-year premiums of the policy Group life insurance at month t	thousand baht
X10	Ratio Scale	The number of insurance premiums for the next year of the policy Ordinary life insurance at month t	thousand baht
X11	Ratio Scale	The number of insurance premiums for the next year of the policy Industrial life insurance as of month t	thousand baht
X12	Ratio Scale	The number of insurance premiums for the next year of the policy Group life insurance at month t	thousand baht

Table 2. Descriptive statistics of life insurance data from 1 January 2003 to 31 December 2022

Variable Name	Range	Mean	S.D.	Min	Max
Y	619,864,190	175,497,200	154,368,300	8,764,067	628,628,300
X1	5	2	1	0	5
X2	3,428,145,960	959,894,500	791,486,800	47,732,210	3,475,878,000
X3	3	23	1	21	24
X4	4,196,430	1,403,044	989,971	109,870	4,306,300
X5	3,098	2,964	926	1,432	4,530
X6	6,660,806	1,111,564	1,033,270	18,390	6,679,196
X7	93,637,950	26,873,540	23,217,760	982,630	94,620,580
X8	1,527,540	465,044	361,715	14,967	1,542,507
X9	11,142,343	2,771,273	2,403,750	155,289	11,297,630
X10	394,847,817	103,999,500	94,079,360	5,492,579	400,340,400
X11	7,322,318	3,388,361	2,149,980	341,733	7,664,051
X12	20,195,549	5,603,258	4,558,032	340,678	20,536,230

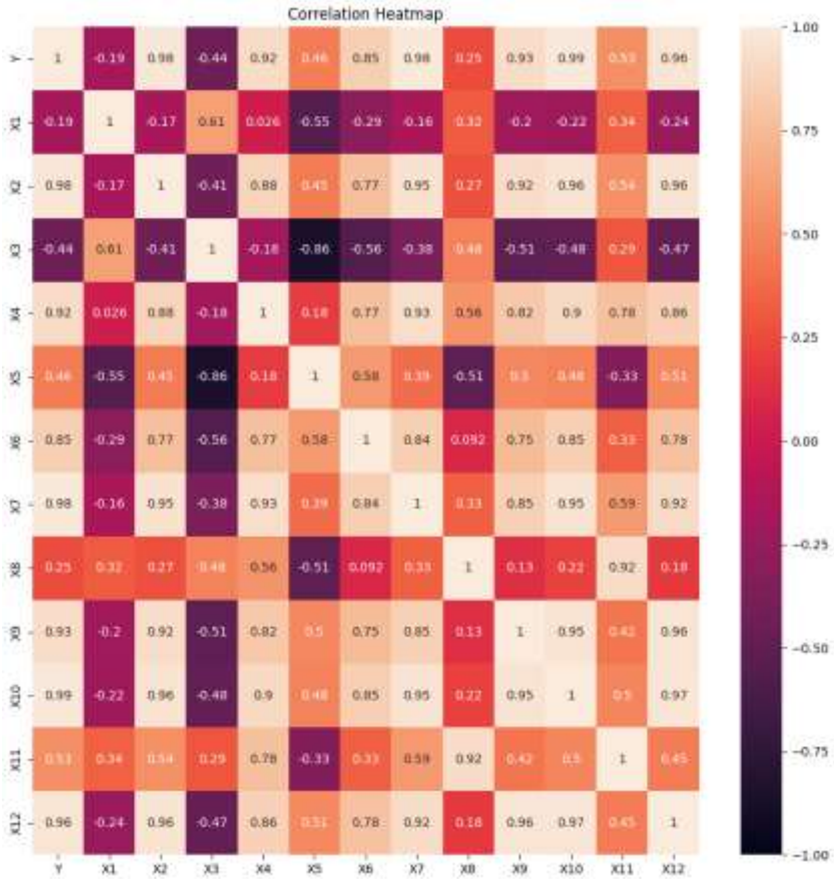

Fig. 3: Correlation heatmap of each features

Table 3. Evaluating forecast accuracy of SARIMAX models

Model	RMSE	MAPE	R^2
SARIMAX(2, 1, 2)x(1, 0, 2, 3)	**3,508,870.19**	**5.62**	**0.78**
SARIMAX(0, 1, 0)x(2, 0, 0, 6)	3,843,282.55	6.40	0.73
SARIMAX(3, 1, 2)x(2, 0, 2, 9)	3,599,253.03	5.80	0.76

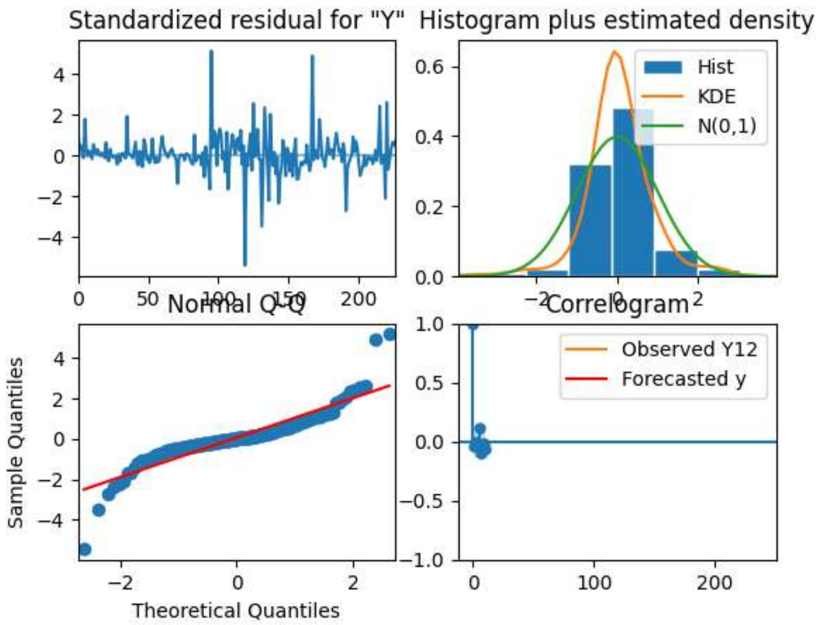

Fig. 4: Standardized residual for the train “Y”

Table 4. Comparison of five methods for the test dataset

Criteria	Models		
	SVR	SARIMAX	SVM-SARIMAX
RMSE	8,138,648.99	**3,508,870.19**	4,729,896.98
MAPE	12.2349	**5.6287**	7.3210
R^2	0.1820	**0.7802**	0.6007

Table 5. An example forecasting of 12 months in 2022

Month	Actual	Models		
		SVR	SARIMAX	SVR-SARIMAX
Jan	49,519,730.95	45,883,424	47,992,221	46,937,822
Feb	45,345,018.00	30,977,795	47,574,761	39,276,278
Mar	54,127,046.22	55,990,949	59,759,573	57,875,261
Apr	40,895,672.55	46,729,081	41,344,472	44,036,776
May	45,810,000.37	39,035,039	50,083,351	44,559,195
Jun	52,133,255.38	52,571,441	55,912,274	54,241,857
Jul	44,210,264.68	39,184,606	46,458,020	42,821,313
Aug	51,026,476.78	52,571,441	52,060,688	52,316,064
Sep	53,524,523.88	39,035,039	53,265,264	46,150,151
Oct	48,611,505.07	47,571,569	43,697,959	45,634,764
Nov	52,847,154.98	47,571,569	50,118,001	48,844,785
Dec	71,902,982.29	56,899,184	65,723,074	61,311,129

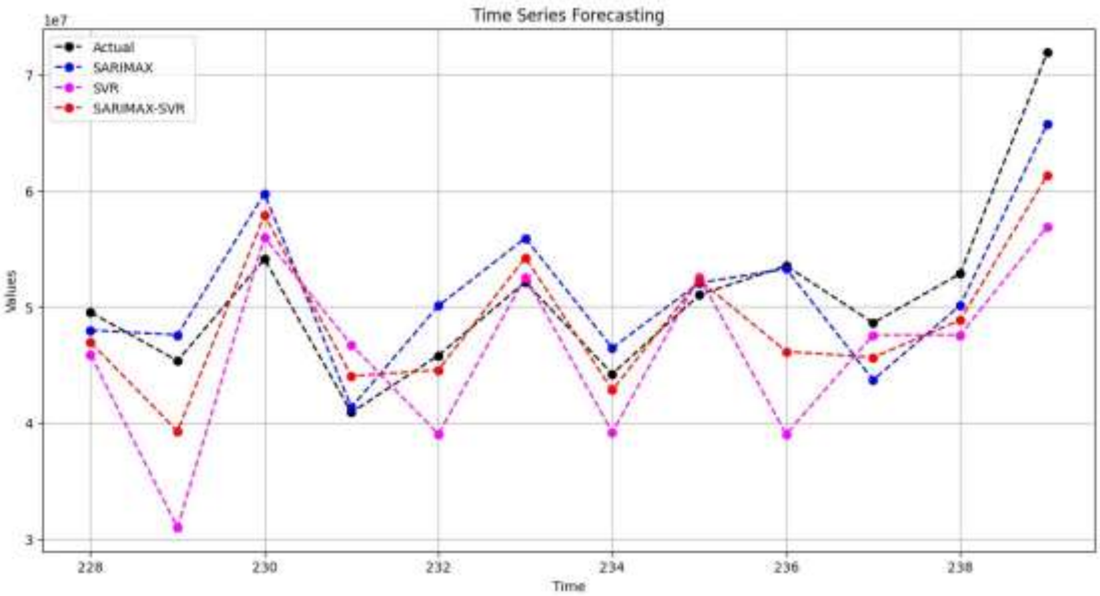

Fig. 5: Time Series Forecasting

Table 6. An example forecasting of 12 months in 2023

Month	Actual	Models		
		SVR	SARIMAX	SVR-SARIMAX
Jan	-	69,016,819	61,509,332	65,263,075
Feb	-	45,980,880	57,807,382	51,894,131
Mar	-	48,906,920	54,470,340	51,688,630
Apr	-	49,074,751	51,227,676	50,151,213
May	-	47,571,569	49,029,952	48,300,760
Jun	-	55,543,190	49,801,474	52,672,332
Jul	-	39,184,606	48,842,697	44,013,651
Aug	-	39,184,606	50,721,533	44,953,069
Sep	-	55,543,190	51,473,279	53,508,234
Oct	-	39,184,606	51,225,199	45,204,902
Nov	-	39,035,039	50,958,835	44,996,937
Dec	-	47,571,569	49,029,760	48,300,664

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Both authors made equal contributions to the current study.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This research was funded by King Mongkut's University of Technology North Bangkok. Contract no. KMUTNB-67-BASIC-16.

Conflict of Interest

The authors have no conflicts of interest to declare.