PARAMETER
JURNAL MATEMATIKA, STATISTIKA DAN TERAPANNYA
e-ISSN. 2829-6303
Vol. 05 No. 1, April 2026
Pages 111-124

Forecasting Revenue Passenger Miles (RPMs) Using Unemployment Rate: A SARIMAX Modeling Approach

Ihsan Fathoni Amri[1]*, Suci Izzati[2], Rendi Andika Putra[3], Iva Aurellia Khalif[4], Febryana Dilla Setyaningrum[5], Isnaini Maulida[6], M. Al Haris[7]

[1]Department of Data Science, Faculty of Science and Agricultural Technology, Universitas Muhammadiyah Semarang
St. Kedungmundu Raya No.18, Semarang 50273, Central Java, Indonesia
[2,3,4,5,6,7]Department of Statistics, Faculty of Science and Agricultural Technology, Universitas Muhammadiyah Semarang
St. Kedungmundu Raya No.18, Semarang 50273, Central Java, Indonesia

E-mail Correspondence Author: *ihsanfathoni@unimus.ac.id*

Abstract
Understanding mobility-related economic dynamics in the United States requires forecasting methods capable of capturing seasonal patterns and external economic shocks. This study aims to forecast Revenue Passenger Miles (RPMs), representing passenger air travel activity, using the unemployment rate as an exogenous variable through the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model. Monthly data from 2015–2024 were obtained from the Federal Reserve Economic Data (FRED) database. RPMs were treated as the endogenous variable, while the unemployment rate served as the exogenous regressor. The analysis involved stationarity testing using the Augmented Dickey–Fuller (ADF) test, model selection based on the Akaike Information Criterion (AIC), and residual diagnostics through the Box–Ljung and Shapiro–Wilk tests. The SARIMAX(0,1,0)(0,1,1)[12] + X model was identified as the optimal specification, with statistically significant parameters and a Mean Absolute Percentage Error (MAPE) of 3.68%, indicating excellent forecasting accuracy.The results show a significant negative relationship between unemployment and passenger air travel activity, suggesting that worsening labor market conditions reduce mobility demand. These findings demonstrate the effectiveness of SARIMAX in forecasting transportation-related economic indicators.

Keywords: *Revenue Passenger Miles, SARIMAX, Unemployment Rate, Time Series Forecasting, Macroeconomic Indicators.*

: https://doi.org/10.30598/parameterv5i1pp111-124

Email: jurnalparameter@gmail.com
Homepage: https://ojs3.unpatti.ac.id/index.php/parameter

Submitted: August 2025
Accepted: April 2026

1. INTRODUCTION

The changing dynamics of the global economy over recent decades have highlighted the importance of transportation-related indicators in reflecting economic activity and consumer mobility. Among these indicators, Revenue Passenger Miles (RPMs), which measure passenger air travel activity, have gained increasing attention as proxies for mobility demand, regional connectivity, and aggregate economic activity. In the United States, RPMs and the unemployment rate are closely associated, as labor market conditions influence purchasing power, travel behavior, and overall mobility patterns [1]. Changes in unemployment levels may affect consumer spending capacity, leading to fluctuations in air travel demand. Therefore, integrating unemployment as an exogenous variable in forecasting RPMs can provide a more comprehensive understanding of transportation-related economic dynamics.

The key challenge lies in developing a predictive model that captures the complex, seasonal relationship between RPMs and unemployment. Inaccurate modeling may lead to policy errors in areas like interest rates and subsidies, as seasonal effects can distort long-term economic trends. Several previous studies have highlighted the importance of using time series forecasting methods, such as ARIMA and SARIMA, to predict economic indicators [2]. However, the limitations of these models in capturing the effects of exogenous variables have prompted the need for more advanced approaches. In this regard, the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model emerges as an appropriate solution, as it can simultaneously integrate seasonal components and external variables [3] [4].

To capture the dynamics between the two variables accurately, this study uses the SARIMAX model, which integrates exogenous factors and seasonal components within a time series framework. This model enables a more comprehensive analysis by incorporating both exogenous variables and seasonal components in the time series data [5]. Compared to traditional time series models, SARIMAX is proven to be more flexible in handling the complex and non-stationary characteristics of economic data [6]. This study analyzes the short- and long-term relationships between RPMs and U.S. unemployment using the SARIMAX model, which captures seasonal and exogenous influences. Grounded in monetarist theory and the Phillips Curve, the analysis highlights unemployment's significant impact on consumer prices [7]. Additionally, the quantitative approach through SARIMAX modeling contributes to strengthening applied econometrics in evidence-based policy making [8].

The use of unemployment rate as a predictor of air travel demand is supported by consumer demand theory and macroeconomic demand dynamics. Rising unemployment reduces household disposable income and suppresses spending on discretionary services, including air travel [9]. Air transportation demand—particularly leisure travel—is highly income elastic, causing passenger demand to decline when employment security weakens . In addition, business travel activity tends to contract during economic downturns as firms reduce operational expenditures and travel budgets, contributing to declines in Revenue Passenger Miles (RPMs) [10]. The sharp collapse in airline traffic during the COVID-19 pandemic further illustrated the close relationship between labor market conditions and aviation demand [11]. These considerations justify the inclusion of unemployment rate as an exogenous variable in the SARIMAX model.

Given the importance of understanding mobility-related economic patterns, the findings of this study are expected to contribute to the literature on transportation

demand forecasting and support data-driven decision making related to air travel activity, economic mobility, and post-pandemic recovery.

The novelty of this study lies in the application of the SARIMAX model with carefully selected variables and rigorous assumption testing on U.S. economic data covering the period from 2015 to 2024. This model is expected to make a significant contribution to the development of big data–driven economic forecasting systems and to support more targeted fiscal policymaking.

2. METHOD

2.1. Data Source and Population

The data used in this study consist of monthly time series observations from January 2015 to December 2024. The dependent variable is Revenue Passenger Miles (RPMs), obtained from the Federal Reserve Economic Data (FRED) database via https://fred.stlouisfed.org/series/RPMD, while the exogenous variable is the Unemployment Rate, retrieved from the U.S. Bureau of Labor Statistics through FRED via https://fred.stlouisfed.org/series/UNRATE. The study period was selected to capture seasonal fluctuations, long-term economic trends, and the economic disruptions caused by the COVID-19 pandemic, which significantly affected labor market conditions and mobility patterns. For forecasting evaluation, the actual unemployment rate values for 2024 were used as known exogenous inputs in the SARIMAX model. Therefore, the analysis represents a test-set prediction scenario rather than a fully out-of-sample forecast.

2.2. Research Variables

Based on Table 1, RPMs serve as the dependent variable, while the unemployment rate acts as the independent variable in the SARIMAX model. The model assumes that changes in the unemployment rate influence variations in RPMs over time.

Table 1. Research Variables

Notation	Variable
Y_t	RPMs
X_t	unemployment rate

Based on Table 1, Revenue Passenger Miles (RPMs) is the endogenous (dependent) variable being forecasted, while the Unemployment Rate serves as the exogenous (independent) predictor in the SARIMAX model. This specification reflects the study's core hypothesis: that labor market conditions, as measured by the unemployment rate, contain predictive information relevant to air travel demand.

2.3. Research Paradigm

This study uses a quantitative-positivistic approach, modeling RPMs and unemployment as time series phenomena. SARIMAX is applied for its ability to capture both seasonality and exogenous effects simultaneously[12]. Previous research supports the effectiveness of exogenous-based models, such as ARIMAX and SARIMAX, in predicting economic variables. Qadrini et al. [13] demonstrated that the ARIMAX model performed best in projecting monetary cash flows based on the lowest RMSE value. Similarly, Amri et al. [14] ARIMAX was applied to predict hotel occupancy in Bali using tourist numbers as an exogenous variable, yielding strong accuracy (AIC: 1920.553; MAPE: 27%) and highlighting its relevance for macroeconomic analysis. Similarly, hybrid

models like ARIMA–GARCH have proven effective in capturing economic data volatility, as shown in forecasting the Jasa Marga stock index in Indonesia [15].

2.4. SARIMA

The ARIMA model can be extended to handle seasonal patterns by adding a seasonal component and is known as the SARIMA (Seasonal ARIMA) model [16]. This model is written in the following notation:

$$SARIMA\ (p,d,q)\ (P,D,Q)s \tag{1}$$

where:

- p, d, q are non-seasonal ARIMA parameters (Autoregressive, Differentiating, and Moving Average),
- P, D, Q are seasonal parameters,
- s is the length of the seasonal period.

In general, the equation form of the SARIMA model can be explained as follows:

$$\varphi_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_q(B^s)\alpha_t \tag{2}$$

With:

$$\Phi_p(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_p B^{Ps} \tag{3}$$

$$\Theta_Q(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_Q B^{Qs} \tag{4}$$

The SARIMA model includes autoregressive (AR), moving average (MA), and differencing components to handle data non-stationarity in both seasonal and non-seasonal patterns. This allows the model to capture complex data structures, such as trends and seasonal fluctuations [17].

2.5. SARIMAX

The Seasonal ARIMA with Exogenous Variables (SARIMAX) model is an extension of the SARIMA model by adding external variables (exogenous variables) as additional inputs believed to influence the dependent variable. This model is particularly useful when the modeled phenomenon is influenced not only by its own historical data and seasonal patterns, but also by other external factors [18]. The SARIMAX model is generally denoted as:

$$\text{SARIMAX (p, d, q) (P, D, Q) + X} \tag{5}$$

With the general form of the equation:

$$\varphi_p B \times \Phi_p(B^s) \times (1-B)^d \times (1-B^s)^D Z_t = \theta_q B \times \Theta_q(B^s) \times \alpha_t + \beta X_t \tag{6}$$

Information:

By including exogenous variables, this model provides more accurate predictions because it captures more information from the surrounding environment.

2.6. Model Selection and Estimation Criteria

The selection of the best model in time series analysis is based on the **Akaike Information Criterion (AIC)**. AIC is used to balance the trade-off between the goodness of fit and the complexity of the model by introducing a penalty term for the number of estimated parameters [19].

$$AIC = -2\ln(L) + 2k \tag{7}$$

2.7. Model Evaluation

To assess the prediction performance of the SARIMAX model, several common quantitative evaluation metrics in time series analysis were used Mean Absolute Percentage Error (MAPE). Furthermore, residual analysis was performed to ensure that prediction errors were not systematic. MAPE measures the error as a percentage, relative to the actual value [20].

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \tag{8}$$

3. RESULTS AND DISCUSSION

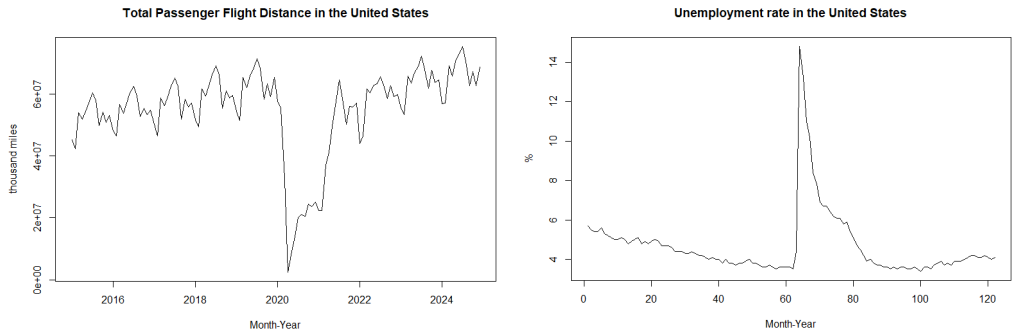
3.1. Initial Data Exploration and Pre-Modeling

The following table presents descriptive statistics of the endogenous variable, Revenue Passenger Miles (RPMs), and exogenous variable, Unemployment Rate, used in the modeling:

Table 2. Descriptive Summary of Variables

Statistics	Revenue Passenger Miles (RPMs)	Unemployment Rate
Average	55.024.167	4,67
Median	58.235.850	4,10
Maximum value	75.257.458	14,80
Minimum value	2.551.127	3,40
Standard deviation	13.935.903	1,72

Based on the table above, the average total flight distance is 55,024,167, with a median of 58,235,850. The maximum and minimum values are 75,257,458 and 2,551,127, respectively, while the standard deviation is 13,935,903, indicating high variability in the monthly RPMs data. Meanwhile, for the exogenous variables, the average unemployment rate is 4.67, with a median of 4.10. The maximum value is 14.80 and the minimum value is 3.40, with a standard deviation of 1.72. The relatively small standard deviation indicates a more stable distribution of exogenous values.



(a) (b)

Figure 1. (a) Revenue Passenger Miles (RPMs) and (b) Unemployment Rate

The actual data plot shows a consistent seasonal pattern each year, indicating a seasonal component in the Revenue Passenger Miles (RPMs) time series. Furthermore, there was a particularly sharp decline in 2020, which coincided with the onset of the COVID-19 pandemic. This event was also accompanied by a significant increase in the unemployment rate, as reflected in the exogenous variables. This causes the data pattern to become more irregular and indicates possible non-stationarity. To test for data stationarity, the Augmented Dickey-Fuller (ADF) test was used. The test results are shown in the following table:

Based on the ADF test results, the p-value was 0.3799, which is greater than $\alpha = 0.05$. Therefore, the null hypothesis that the data contains a unit root (non-stationary) cannot be rejected. This means the data is non-stationary and requires differencing before building a SARIMAX model. For SARIMAX modeling purposes, the time series data is divided into two parts: training data and test data. The training data is used to build and estimate model parameters, while the test data is used to evaluate the model's performance on data not used during the training process. Details of the data split are shown in the following Table 3.

Table 3. Data Partition Based on Time Series Chronology

Data Types	Period	Number of Observations
Training Data	January 2015 – December 2023	108
Test Data	January 2024 – December 2024	12

This separation aims to test the generalization ability of the model in predicting future data, which is an important aspect in time series modeling.

3.2. Exogenous Variable Analysis

To examine the relationship between exogenous and endogenous variables, Pearson correlation was used to assess linear strength, while the Granger Causality test evaluated predictive causality within the time series context. The results are summarized in the following Table 4.

Table 4. Statistical Test Results Between Exogenous and Endogenous Variables

Data Types	Value	n	lag
Pearson Correlation	$r = -0.8477$, $p < 0.0001$	108	-
Granger Causality	$F(1,91) = 6.7683$, $p = 0.01083$	95 (post-differencing)	1

The Pearson correlation yielded r = −0.8477 ($p < 0.0001$, n = 108), indicating a strong, statistically significant, and negative linear relationship: as the unemployment rate increases, Revenue Passenger Miles (RPMs) tend to decrease. The Granger Causality test was applied to the doubly-differenced series (one non-seasonal and one seasonal difference), consistent with the stationarity transformation used in SARIMAX modeling, with lag order set to 1. The test yielded $F(1, 91) = 6.7683$, $p = 0.01083$ (< 0.05), indicating that past values of the unemployment rate contain statistically significant predictive information for RPMs, thereby justifying its inclusion as an exogenous variable in the model.

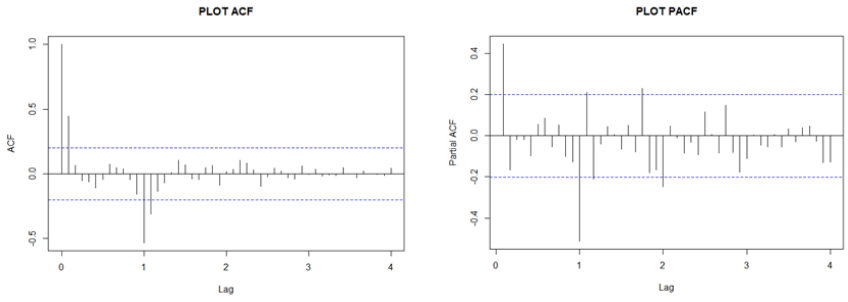
3.3. Data stationarity test

The stationarity test aims to determine whether time series data has a unit root, which is essential before modeling. Initial test results using the Augmented Dickey-Fuller (ADF) showed the data was non-stationary (p-value > 0.05), prompting both seasonal and non-seasonal differencing.

After differencing, a second ADF test yielded a p-value of 0.01, which is below the 0.05 threshold. This result rejects the null hypothesis of a unit root, indicating that the data is now stationary and ready for SARIMAX modeling without further differencing.

3.4. Model identification

In time series analysis, the model identification stage is a crucial step in determining the model structure that best fits the data that has undergone stationarity. This identification is performed by observing patterns in the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots formed after differencing the data, both in the non-seasonal and seasonal components. Through visual interpretation of these patterns, the initial values of the SARIMA model parameters can be estimated to optimally describe the data dynamics. This stage provides a crucial foundation for building accurate predictive models in subsequent processes.

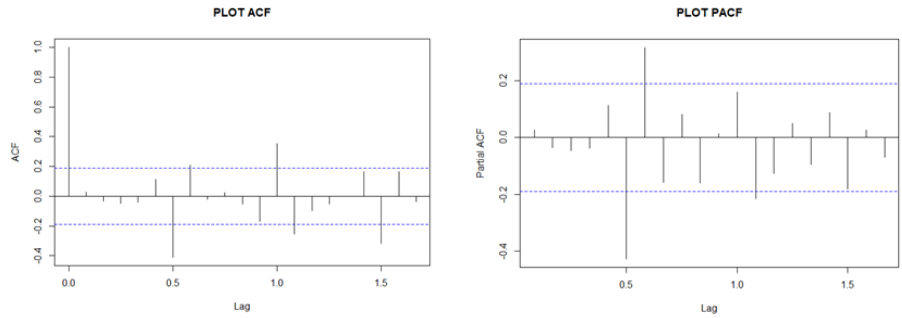


(a) (b)

Figure 2. ACF and PACF Plots After Seasonal and Non-Seasonal Differencing: (a) ACF Plot, and (b) PACF Plot

The ACF and PACF plots above are the results of one non-seasonal and one seasonal differencing process each. Based on the resulting pattern, a non-seasonal component is detected, suggesting an AR (1) or possibly AR (2) model, indicated by a clear cutoff at lag 3 in the PACF plot. The ACF plot shows a decay pattern that begins to appear at lag 2. For the seasonal component, a MA (1) model is indicated by a significant value at lag 12 and a cutoff at lag 24 in the ACF plot. The PACF plot also shows significance at multiples of lag 12, but exhibits a decay pattern. Based on this identification, an initial

estimate of the SARIMA model with parameters (p,d,q) (P, D, Q)[12] is obtained, namely (1,1,0) (0,1,1)[12] or (2,1,0)(0,1,1)[12].



(a) (b)

Figure 3. ACF and PACF Plots After Non-Seasonal Differencing: (a) ACF Plot, and (b) PACF Plot

To expand the model exploration, further analysis was conducted on the ACF and PACF plots from the first-stage non-seasonal differencing results. Both plots showed significance at lag 6, so the model search scope was expanded to SARIMA(6,1,6)(0,1,1)[12].

In total, 196 SARIMA model combinations were studied, ranging from the baseline SARIMA (0,1,0) (0,1,0)[12] to the maximum SARIMA (6,1,6) (1,1,1)[12]. The details are as follows:

- Model 1: SARIMA (0,1,0) (0,1,0)[12]
- Model 2: SARIMA (1,1,0) (0,1,0)[12]
- Model 3: SARIMA (2,1,0) (0,1,0)[12]

. . .

- Model 52: SARIMA (1,1,0) (1,1,0)[12]
- Model 53: SARIMA (2,1,0) (1,1,0)[12]

. . .

- Model 196: SARIMA (6,1,6) (1,1,1)[12]

All models were evaluated based on the Akaike Information Criterion (AIC) and parameter significance. The five models with the lowest AIC values are presented in the following table SARIMA and SARIMAX table:

Table 5. Comparison of SARIMA Model Candidates

Model	AIC	Parameter Significance
SARIMA(0,1,1)(0,1,1)[12]	3206.65	Significant
SARIMA(2,1,0)(0,1,1)[12]	3207.248	Significant
SARIMA(0,1,1)(1,1,1)[12]	3207.495	Not Significant
SARIMA(0,1,2)(0,1,1)[12]	3207.898	Not Significant
SARIMA(2,1,0)(1,1,1)[12]	3208.076	Not Significant

Table 6. Comparison of SARIMAX Model Candidates

Model	AIC	Parameter Significance
SARIMAX(0,1,0)(0,1,1)[12]	3178.433	Significant
SARIMAX(0,1,0)(1,1,1)[12]	3179.285	Not Significant
SARIMAX(1,1,0)(0,1,1)[12]	3180.297	Not Significant
SARIMAX(0,1,1)(0,1,1)[12]	3180.355	Not Significant
SARIMAX(1,1,0)(1,1,1)[12]	3181.129	Not Significant

Although the SARIMA (1,1,0) (0,1,1)[12] model was initially considered as a candidate, the estimation results indicated that its parameters were not statistically significant. In contrast, the SARIMA (0,1,1) (0,1,1)[12] and SARIMAX (0,1,0) (0,1,1)[12] models yielded the lowest AIC values and had all parameters statistically significant, making them the most suitable models.

Table 7. Estimated Parameters of the SARIMA Model

Coefficient	Estimated Value	Significance
ma1	0,4642	< 0,0001
sma1	-0,9997	<0,0001

Table 8. Estimated Parameters of the Best SARIMAX Model

Coefficient	Estimated Value	Significance
sma1	-1	< 0,0001
xreg	-2.890.296,9	< 0,0001

The SARIMA model can be expressed in the form of the following equation:

$$(1 - B)(1 - B^{12})Y_t = (1 + 0{,}4642B)(1 - 0{,}9997B^{12})\varepsilon_t \quad (9)$$

The best SARIMAX model can be expressed in the form of the following equation:

$$(1 - B)(1 - B^{12})Y_t = (1 - B^{12})\varepsilon_t - 2.890.296{,}9x_t \quad (10)$$

3.5. Residual assumption test

After the SARIMAX model was constructed, a residual assumption test was conducted to evaluate model adequacy. Two approaches were used: the White Noise test (Box–Ljung) to assess whether residuals contain autocorrelation, and the Shapiro–Wilk test to evaluate whether the residuals are normally distributed in accordance with classical statistical assumptions. The results of both tests are presented in Table 9 and 10.

Table 9. Residual diagnostic test results for SARIMA model

Test	**P-value**	**Conclusion**
Box-Ljung	0,3791	No autocorrelation detected
Shapiro-Wilk	< 0,0001	Residuals are not normally distributed

Table 10. Residual diagnostic test results for SARIMAX model

Test	**P-value**	**Conclusion**
Box-Ljung	0,4104	No autocorrelation detected
Shapiro-Wilk	< 0,0001	Residuals are not normally distributed

For both models, the Box–Ljung test yielded p-values greater than 0.05, indicating no significant autocorrelation in the residuals. This suggests that the residuals behave as white noise, implying that each model has effectively captured the essential patterns in the time series. However, the Shapiro–Wilk test results revealed that the residuals for both models are not normally distributed, as indicated by p-values below 0.0001. This violation

is likely the result of external disturbances, particularly the COVID-19 pandemic, which introduced structural shifts and outliers in the data.

Although the assumption of normality is not fully satisfied, this limitation is still considered acceptable for forecasting purposes. In time series forecasting, the absence of autocorrelation and the white-noise behavior of residuals are generally more important than strict residual normality. Since both models produce independent residuals and demonstrate satisfactory forecasting performance, the non-normality of residuals does not substantially reduce the reliability of the forecasts. Therefore, both the SARIMA and SARIMAX models remain appropriate for forecasting applications in this study.

3.6. Data forecasting

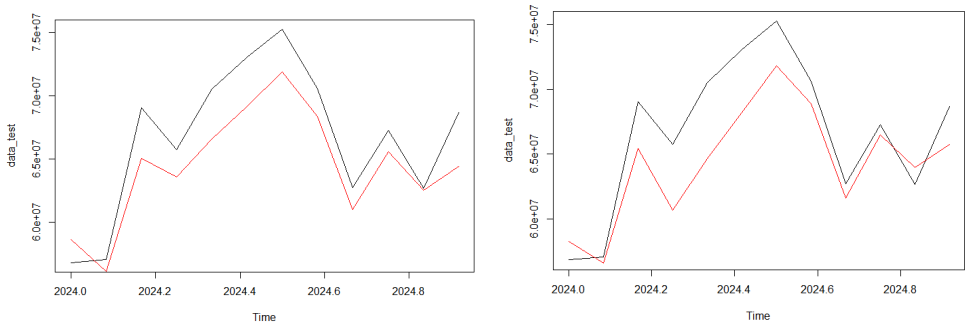
Forecasting is the process of predicting the value of a variable in future periods based on historical trends. In this study, the forecast was performed on the Revenue Passenger Miles (RPMs) by incorporating an exogenous variable—unemployment rate—as one of the influencing factors. The forecast covered a one-year period from January to December 2024.

To evaluate the model's performance, the predicted values from the SARIMAX and SARIMA models were compared with the actual observations throughout 2024. The comparison is presented in table below.

Table 11. Comparison between actual and predicted values in 2024

Time	Actual Value	Predicted SARIMAX	Predicted SARIMA
January	56824550	58634386	58289613
February	57070731	56133848	56591943
March	69049381	65074711	65436476
April	65739107	63605551	60659549
May	70516306	66575539	64625104
June	73080503	69146418	68191555
July	75257458	71901256	71813513
August	70612934	68397405	68887757
September	62722315	60991449	61610299
October	67289994	65564970	66472894
November	62674521	62536041	63989961
December	68684657	64444815	65738220

Therefore, both models are deemed effective, with SARIMAX offering marginally better performance for forecasting monthly Revenue Passenger Miles (RPMs) in 2024. Although the difference in MAPE values is relatively small, the results indicate that economic conditions represented by the unemployment rate contribute to explaining fluctuations in Revenue Passenger Miles (RPMs). This suggests that the SARIMAX model may provide more reliable forecasts for transportation planning during periods of economic change.



(a) **(b)**

Figure 4. Plot of actual and predicted value: (a) SARIMAX (b) SARIMA

Overall, the forecast values are relatively close to the actual values, indicating that the model was effective in capturing the movement patterns in the data. To quantitatively assess forecasting accuracy, the Mean Absolute Percentage Error (MAPE) was used, defined as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \tag{11}$$

SARIMA calculation:

$$\text{MAPE} = \frac{\left|\left(\frac{56824550 - 58289613}{56824550}\right)\right| + \cdots + \left|\left(\frac{68684657 - 65738220}{68684657}\right)\right|}{12} \times 100\%$$

$$\text{MAPE} = \frac{\left|\left(\frac{-1465063}{56824550}\right)\right| + \cdots + \left|\left(\frac{2946437}{68684657}\right)\right|}{12} \times 100\%$$

$$\text{MAPE} = \frac{0.0258 + \cdots + 0.0429}{12} \times 100\%$$

$$\text{MAPE} = 3.9846\%$$

SARIMAX calculation:

$$\text{MAPE} = \frac{\left|\left(\frac{56824550 - 58634386}{56824550}\right)\right| + \cdots + \left|\left(\frac{68684657 - 64444815}{68684657}\right)\right|}{12} \times 100\%$$

$$\text{MAPE} = \frac{\left|\left(\frac{-1809836}{56824550}\right)\right| + \cdots + \left|\left(\frac{4239842}{68684657}\right)\right|}{12} \times 100\%$$

$$\text{MAPE} = \frac{0.0318 + \cdots + 0.0617}{12} \times 100\%$$

$$\text{MAPE} = 3.6762\%$$

The SARIMAX model achieved a lower MAPE than the SARIMA model, suggesting that the inclusion of the Unemployment Rate as an exogenous variable slightly improved forecasting accuracy. A MAPE value below 10% generally indicates high predictive accuracy. Therefore, both models can be considered effective for forecasting monthly Revenue Passenger Miles (RPMs) in 2024, although the SARIMAX model demonstrated slightly better predictive performance.

To provide a clearer comparison between the final selected models, the AIC values, parameter significance, and forecasting accuracy are summarized in the following table.

Table 12. Summary of Final SARIMA and SARIMAX Models

Model	AIC	Parameter Significance	MAPE (%)	Conclusion
SARIMA(0,1,1)(0,1,1)[12]	3206.650	Significant	3.9846	Accepted
SARIMAX(0,1,0)(0,1,1)[12] + Unemployment Rate	3178.433	Significant	3.6762	Best Model

Based on the comparison results presented in Table 12, the SARIMAX(0,1,0)(0,1,1)[12] model with the Unemployment Rate as an exogenous variable was selected as the best forecasting model. This model produced the lowest AIC value and the lowest MAPE value, while all estimated parameters were statistically significant. These findings indicate that incorporating the Unemployment Rate improved forecasting performance compared to the standard SARIMA model.

4. CONCLUSION

This study aims to analyze and predict the value of Revenue Passenger Miles (RPMs) by considering the influence of external factors, such as the unemployment rate in the United States. To this end, the SARIMAX method is used, which is capable of handling seasonal time series data and accounting for exogenous variables. The analysis process begins with a stationarity test, model identification using the ACF and PACF, and determining the order of the seasonal ARIMA model. The ARIMA(1,1,0)(0,1,1)[12] model was initially considered because it supported the autocorrelation pattern. However, estimation results showed that the AR(1) parameter was not statistically significant. As an alternative, the ARIMA(0,1,0)(0,1,1)[12] + X model was used, with X representing the unemployment rate. This model produced statistically significant estimates for all parameters ($p < 0.05$), including a statistically significant predictive contribution from the unemployment rate as an exogenous variable. Although this model is not immediately apparent from the ACF and PACF, it is simpler and better at explaining data variation. Model validation was conducted through residual diagnostic tests (including the Ljung–Box test, normality test, and white noise check), and further assessed using the Akaike Information Criterion (AIC) and Mean Absolute Percentage Error (MAPE). The SARIMAX(0,1,0)(0,1,1)[12] + Unemployment Rate model demonstrated better performance than its non-exogenous counterpart, with a lower AIC and MAPE of 3.6762%, compared to 3.9846% for the SARIMA model. In conclusion, this study demonstrates that the unemployment rate is significantly associated with Revenue Passenger Miles (RPMs) and that its inclusion as an exogenous predictor substantially improves SARIMAX forecasting accuracy. The integration of macroeconomic variables into SARIMAX modeling enhances forecasting precision, offering valuable insights for strategic planning in the air transportation industry, economic policy-making, and public sector decision-making.

Acknowledgments

The authors would like to thank all individuals who contributed to this research through technical assistance, data processing support, coding assistance, and constructive discussions during the development of the manuscript.

Funding Information

Authors state no funding involved.

Author Contributions Statement

Author 1: Conceptualization, methodology, data analysis, and writing-original draft preparation.
Author 2: Data collection and data curation.
Author 3: Data preprocessing and software implementation.
Author 4: Statistical analysis and data visualization.
Author 5: Data cleaning and validation.
Author 6: Result interpretation and manuscript review.
Author 7: Writing-review & editing and formatting.
All authors contributed to discussing the results and approved the final manuscript.

Conflict of Interest Statement

Authors state no conflict of interest.

Data Availability

The data used in this study are publicly available and described in the Method section of this manuscript.

REFERENCES

[1] W. Lin, J. Z. Huang, and T. McElroy, "Time Series Seasonal Adjustment Using Regularized Singular Value Decomposition," Journal of Business and Economic Statistics, vol. 38, no. 3, pp. 487–501, Jul. 2020, doi: 10.1080/07350015.2018.1515081.

[2] G. T. Wilson, "Time Series Analysis: Forecasting and Control, 5th Edition, by George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel and Greta M. Ljung, 2015. Published by John Wiley and Sons Inc., Hoboken, New Jersey, pp. 712. ISBN: 978-1-118-67502-1," J. Time Ser. Anal., vol. 37, no. 5, pp. 709–711, Sep. 2016, doi: 10.1111/jtsa.12194.

[3] I. F. Amri, W. N. Ramadhan, S. Ainurrofiah, and M. Al Haris, "Pemodelan ARIMA dan ARIMAX untuk Memprediksi Jumlah Produksi Padi di Kota Magelang," Square : Journal of Mathematics and Mathematics Education, vol. 5, no. 2, pp. 93–105, Oct. 2023, doi: 10.21580/square.2023.5.2.17059.

[4] L. Putri and Zilrahmi, "Forecasting Inflation Rate in Indonesia Using Autoregressive Integrated Moving Average Method," UNP Journal of Statistics and Data Science, vol. 3, no. 3, pp. 288–295, Aug. 2025, doi: 10.24036/ujsds/vol3-iss3/377.

[5] R. J. . Hyndman and George. Athanasopoulos, Forecasting : principles and practice. OTexts, 2014.

[6] M. I. Rizki and T. A. Taqiyyuddin, "Penerapan Model SARIMA untuk Memprediksi Tingkat Inflasi di Indonesia," Jurnal Sains Matematika dan Statistika, vol. 7, no. 2, Aug. 2021, doi: 10.24014/jsms.v7i2.13168.

[7] J. Hossain, "Comparative Analysis of ARIMA, SARIMAX, and Random Forest Models for Forecasting Future GDP of the UK in Relation to Unemployment Rate," International Journal of Management, Accounting and Economics, vol. 10, no. 11, pp. 2383–2126, 2023, doi: 10.5281/zenodo.10473611.

[8] M. I. Rizki and T. A. Taqiyyuddin, "Penerapan Model SARIMA untuk Memprediksi Tingkat Inflasi di Indonesia," Jurnal Sains Matematika dan Statistika, vol. 7, no. 2, Aug. 2021, doi: 10.24014/jsms.v7i2.13168.
[9] L. J. Santos, A. V. M. Oliveira, and D. M. Aldrighi, "Testing the differentiated impact of the COVID-19 pandemic on air travel demand considering social inclusion," J. Air Transp. Manag., vol. 94, Jul. 2021, doi: 10.1016/j.jairtraman.2021.102082.
[10] X. Sun, S. Wandelt, and A. Zhang, "COVID-19 pandemic and air transportation: Summary of Recent Research, Policy Consideration and Future Research Directions," Transp. Res. Interdiscip. Perspect., vol. 16, Dec. 2022, doi: 10.1016/j.trip.2022.100718.
[11] J. B. Sobieralski, "COVID-19 and airline employment: Insights from historical uncertainty shocks to the industry," Transp. Res. Interdiscip. Perspect., vol. 5, May 2020, doi: 10.1016/j.trip.2020.100123.
[12] J. Hossain, "Comparative Analysis of ARIMA, SARIMAX, and Random Forest Models for Forecasting Future GDP of the UK in Relation to Unemployment Rate," International Journal of Management, Accounting and Economics, vol. 10, no. 11, pp. 2383–2126, 2023, doi: 10.5281/zenodo.10473611.
[13] L. Junaedi, N. Damastuti, and A. Widodo, "Penerapan Metode Seasonal ARIMA (SARIMA) untuk Peramalan Penjualan Barang dengan Pola Musiman Tahunan," JISEM Jurnal Program Studi Informatika Universitas Katolik Widya Mandala Surabaya, vol. 01, pp. 38–48, 2025, doi: 10.33508/jisem.v1i01.7403.
[14] A. L. M. Serrano et al., "Statistical Comparison of Time Series Models for Forecasting Brazilian Monthly Energy Demand Using Economic, Industrial, and Climatic Exogenous Variables," Applied Sciences (Switzerland), vol. 14, no. 13, Jul. 2024, doi: 10.3390/app14135846.
[15] Ihsan Fathoni Amri, A. Arya, Yolan Triky, Kaia Raissa Akmalia, Abdul Ghufron, and M. Al Haris, "Forecasting Hotel Occupancy Rates in Bali Province using the SARIMAX Method with Tourist Data as an Exogenous Variable," EKSAKTA: Journal of Sciences and Data Analysis, pp. 120–131, Oct. 2024, doi: 10.20885/eksakta.vol5.iss2.art2.
[16] I. F. Amri, W. Sari, V. A. Widyasari, N. Nurohmah, and M. Al Haris, "The ARIMA-GARCH Method in Case Study Forecasting the Daily Stock Price Index of PT. Jasa Marga (Persero)," EIGEN MATHEMATICS JOURNAL, vol. 7, no. 1, pp. 25–33, Apr. 2024, doi: 10.29303/emj.v7i1.174.
[17] L. Qadrini, A. Asrirawan, N. Mahmudah, M. Fahmuddin, and I. F. Amri, "Forecasting Bank Indonesia Currency Inflow and Outflow Using ARIMA, Time Series Regression (TSR), ARIMAX, and NN Approaches in Lampung," Jurnal Matematika, Statistika dan Komputasi, vol. 17, no. 2, pp. 166–177, Dec. 2020, doi: 10.20956/jmsk.v17i2.11803.
[18] J. Kasali and A. A. Adeyemi, "Model-Data Fit using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and The Sample-Size-Adjusted BIC," Square : Journal of Mathematics and Mathematics Education, vol. 4, no. 1, pp. 43–51, Apr. 2022, doi: 10.21580/square.2022.4.1.11297.
[19] I. Nabillah and I. Ranggadara, "Mean Absolute Percentage Error untuk Evaluasi Hasil Prediksi Komoditas Laut," JOINS (Journal of Information System), vol. 5, no. 2, pp. 250–255, Nov. 2020, doi: 10.33633/joins.v5i2.3900.
[20] R. Yulianti et al., "BAREKENG: Journal of Mathematics and Its Applications Comparison Of Sarima And Sarimax Methods For Forecasting Harvested Dry Grain Prices In Indonesia "Comparison Of Sarima And Sarimax Methods For Forecasting Harvested Dry Grain Prices In Indonesia 320 Yulianti, et al. Comparisson Of Sarima Dan Sarimax Methods For Forcasting Harvested…," BAREKENG: J. Math. & App, vol. 19, no. 1, pp. 319–0330, 2025, doi: 10.30598/barekengvol19iss1pp0319-0330.