IRSTI 28.23.01 https://doi.org/10.26577/jpcsit2024-02i03-01

Dashqin Nazarli

National Aviation Academy, Baku, Azerbaijan
e-mail: dnazarli.32073@naa.edu.az

TIME SERIES FORECASTING MODELS OF NON-SCHEDULED PASSENGER AIR TRANSPORTATION

Abstract. The change in the time series of non-scheduled passenger air transportation is random and variable, which creates a number of problems in forecasting the demand for this type of transportation. In calculations based on trend models, it is usually not possible to take into account all extraneous factors affecting non-scheduled passenger air transportation. For this reason, the accuracy and practical significance of the forecast are low. Considering the mentioned facts, this paper investigates the application of combined autoregressive integrated moving average (ARIMA) and support vector machine (SVM) methods to improve the accuracy of charter air transportation demand forecasting. ARIMA and SVM models usually complement each other in forecasting due to their inherent characteristics. These features include detecting temporal dependencies and trends, as well as handling non-linear relationships within historical data. The integration of these methods aims to obtain optimal forecast results using the time series analysis of the ARIMA model and the non-linear relationship detection feature of the SVM model. The obtained results emphasize the ability of ARIMA-SVM models to adapt to the dynamic demand patterns of non-scheduled air transportation and also offer a number of efficient ideas for the optimization of operational strategies and resource allocation in this field. The theoretical-practical results of this study, conducted with ARIMA and SVM methods, will be effective in the field of non-scheduled passenger air transportation.

Key words: non-scheduled air transportation, support vector machine, non-linear models, time-series analysis, demand model.

1. Introduction

Non-scheduled air transportation, which includes charter flights and air taxis, experiences fluctuating demand influenced by a myriad of factors such as economic trends, seasonal variations, and sudden market shifts. Accurate forecasting in this context is crucial for optimizing resource allocation, improving operational efficiency, and enhancing strategic planning. Among the various methods employed for time series forecasting, autoregressive integrated moving average (ARIMA) and support vector machine (SVM) stand out for their distinct advantages. ARIMA is well-regarded for its capability to model and forecast linear patterns in time series data, making it a valuable tool for understanding underlying trends and seasonality. On the other hand, SVM is a powerful machine learning technique that excels at capturing complex, non-linear relationships within data [1], [2], [3], [4], [5]. In this paper, the individual characteristics and applications of ARIMA and SVM methods are examined in the context of forecasting non-scheduled passenger air transportation. By analyzing historical data using these two methodologies separately, we aim to provide a comprehensive evaluation of their respective strengths, limitations, and suitability for this particular forecasting challenge. Through empirical analysis, we seek to determine how each method performs under varying conditions and to what extent they contribute to enhancing the accuracy of demand forecasts in the non-scheduled air transportation field [5], [6], [7], [8], [9], [10], [11], [12].

2. Related works

Non-scheduled air transportation time series forecasting aims to forecast future data based on past data. When only one variable changes over time, a univariate prediction model is used. Multivariate time series forecasting is used when multiple variables and their values change over time.

Statistical models are used to forecast the demand for air transportation. Data stationarity should be checked before statistical modeling is performed. The main reason is that statistical variables (mean, variance, and autocorrelation) should not change over time. Examples of these models are statistical models such as Auto-Regressive Integrated Moving

Average (ARIMA) and SARIMA [13], [14], [15]. In the mentioned models, trend and seasonal factors are included in the time series data.

The application of time-series-based forecasting models poses certain difficulties due to the increased uncertainty and irregularity in the movement of air passengers. Sun et al. [4]-[5] proposed a nonlinear vector auto-regression neural network (NVARNN) method to predict passenger capacity in air transport. First, input characteristics were identified and extracted using the mean impact value (MIV) method. In the next step, the NVARNN method is used to obtain forecast results through data modeling. According to this study, multivariate forecasting methods consistently outperformed univariate forecasting methods. According to the results of the study, neural networks were superior to ARIMA and SARIMA models due to their complexity. Also, at the end of the study, six factors affecting the flow of passengers using air transport were identified [15], [16], [17], [18].

Looking at the research conducted in recent years, it appears that mixed forecasting models provide more effective results than individual models. In this study, the strengths of both unique model architectures are combined. This is a very important factor for forecasting statistics. A new forecasting model combining SARIMA and Singular Vector Regression (SVR) models was presented by researchers. The SARIMA model removed the non-stationarity of the series by analyzing the time series and applying the seasonal differences in the correct order. SVM was used to capture linear and non-linear patterns in time series data [18], [19], [20], [21], [22].

It should be noted that the forecasting model has only been tested in the short term. It was determined that there are factors that affect the demand for air transportation during the construction of the model and are not taken into account in the model. Therefore, sudden deviations in the model can be combined to increase the accuracy of the forecast.

3. Problem statement

Construction of forecasting models of non-scheduled passenger air transportation based on ARIMA-SVM methods and comparative analysis of the obtained forecasting results.

4. Method and Methodology

The differential autoregressive moving average model (ARIMA) is an important method for studying time series. In ARIMA (p, d, q), p is the number of autoregressive items, q is the moving average item number, and d is the number of differences made to make it a stationary sequence. The ARIMA (p, d, q) model is an extension of the ARMA (p, q) model. [5]

The ARIMA model in the following form:

$$Y_t = c + y_t + z_t \tag{1}$$

where,

$$y_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots. + \theta_q\varepsilon_{t-q} \tag{2}$$

$$z_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots. + \theta_q\varepsilon_{t-q} \tag{3}$$

We apply the method of least squares to find the unknown coefficients. For this, the following issue should be resolved:

$$\sum_{t=1}^{N}[(\bar{Y}_t - Y_t)]^2 \dot{\rightarrow} min \tag{4}$$

The solution to problem (4) is reduced to the following matrix equation:

$$A\varphi = B \tag{5}$$

is a dimensional symmetric matrix, the elements of which are as follows:

$$A = \begin{pmatrix} n & \sum_{t=1}^{N}\bar{Y}_{t-1} & \sum_{t=1}^{N}\bar{Y}_{t-2} \cdots \sum_{t=1}^{N}\bar{Y}_{t-p} \\ \sum_{t=1}^{N}\bar{Y}_{t-1} & \sum_{t=1}^{N}Y_{t-1}{}^{2} & \sum_{t=1}^{N}\bar{Y}_{t-2}\,\bar{Y}_{t-1} \ldots\ldots \sum_{t=1}^{N}\bar{Y}_{t-p}\,\bar{Y}_{t-1} \\ \sum_{t=1}^{N}\bar{Y}_{t-p} & \sum_{t=1}^{N}\bar{Y}_{t-1}\,\bar{Y}_{t-p} & \sum_{t=1}^{N}\bar{Y}_{t-2}\,\bar{Y}_{t-p} \ldots\ldots \sum_{t=1}^{N}\bar{Y}^{2}{}_{t-p} \end{pmatrix} \tag{6}$$

$$B = \begin{pmatrix} \sum_{t=1}^{N} \bar{Y}_t \\ \sum_{t=1}^{N} \bar{Y}_t \, \bar{Y}_{t-1} \\ \sum_{t=1}^{N} \bar{Y}_t \, \bar{Y}_{t-p} \end{pmatrix} \quad (7)$$

Considering expressions (6) and (7) in formula (5), unknown coefficients are founded.

The SVM kernel is considered a function that takes low-dimensional input space and transforms it into higher-dimensional space, usually it converts non-separable problems to separable problems. It is mostly useful in non-linear separation problems. Consider the following formulas:

$$\text{Linear}: K(w,b) = w^T x + b \quad (8)$$

$$\text{Polynomial}: K(w,x) = (\gamma w^T x + b)^N \quad (9)$$

$$\text{Gaussian RBF}: K(w,x) = exp\left(-\gamma \left\| x_i - x_j \right\|^n\right) \quad (10)$$

$$\text{Sigmoid}: K\left(x_i, x_j = tanh\left(\alpha x_i^T \, x_j + b\right)\right) \quad (11)$$

5. Experimental results

First of all, statistical data for non-scheduled passenger air transportation was collected to build the calculation model. These data are presented in Figure 1. Statistics data covering the years 2020–2023 (total 48 months) were used to build a forecast model based on ARIMA-SVM models (Figure 1). Based on the given statistical indicators, a forecast for 2023 will be made based on the years 2020–2022, and the results will be compared with the actual indicators of 2023.

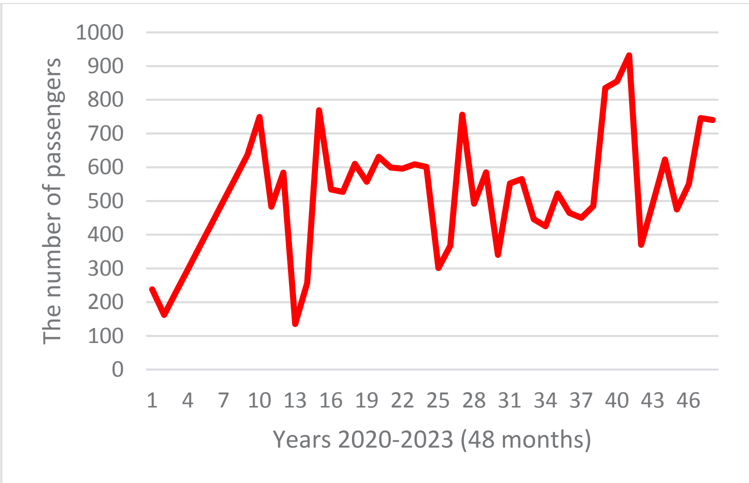

Figure 1 – Monthly statistics of non-scheduled passenger air transportation for 2020–2023

Figure 2 shows the autocorrelation function for statistical data on non-scheduled passenger air transportation. It is clear from here that calculations will be made according to formula (2), taking into account (p = 3) in the ARIMA model.

Here, **UCL** is the upper confident level and **LCL** is the lower confident level.

After the autocorrelation function is established, the ARIMA model is reported based on the preliminary results obtained. By substituting these values in formula (2), () and the values of c are obtained (reports were made in the MATLAB 2023a software package). Preliminary calculation results are shown in Figure.3. As can be seen from Figure. 3, If we compare the results obtained during the calculations based on the formula (2) with the actual indicators, we will see that there are serious differences in some points from the observations made. This indicates that those actual results are anomalous in the general results. In general, anomalous deviations in the general trend are observed in non-scheduled air transportation.

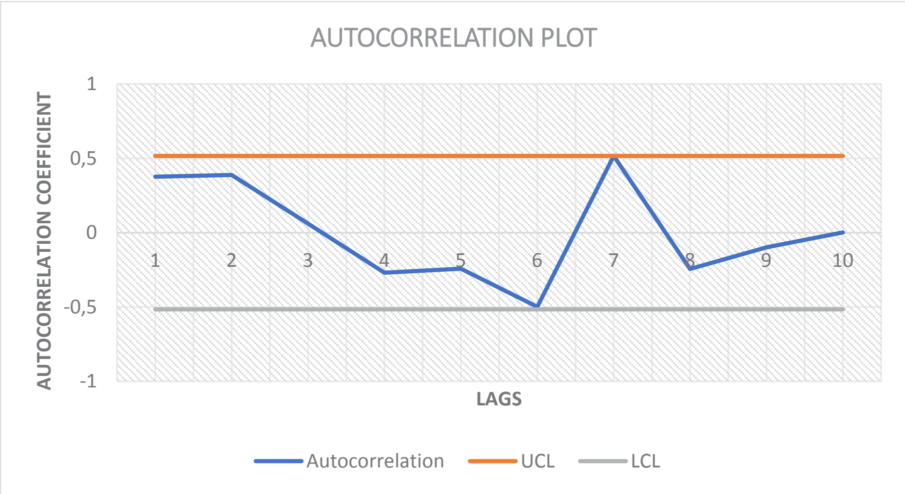

Figure 2 – Autocorrelation function for non-scheduled passenger air transportation

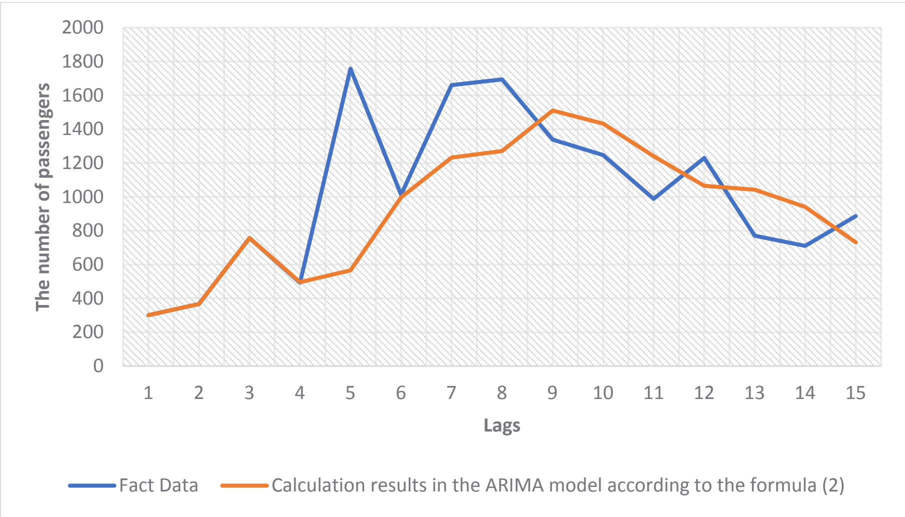

Figure 3 – Calculation results in the ARIMA model based on the formula (2)

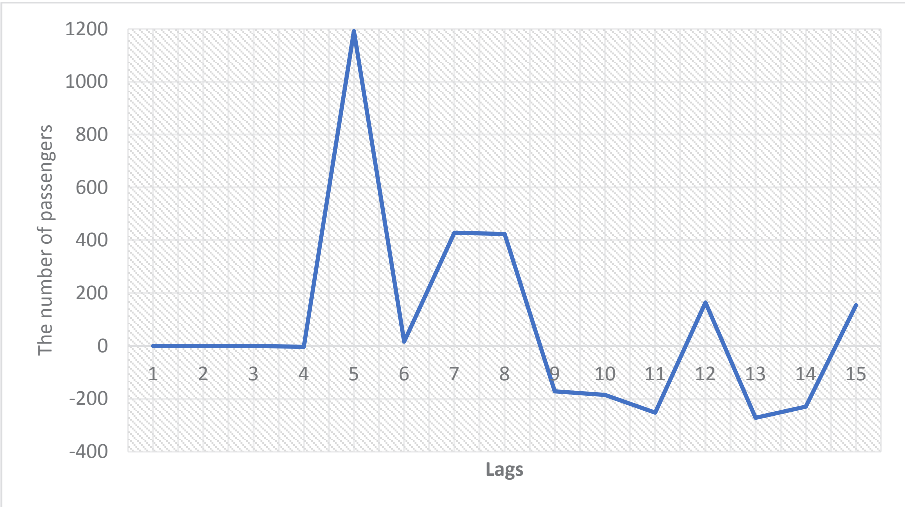

Figure 4 – The differences between the real results and the calculation results during the calculation in the ARIMA model is based on the formula (2)

In the next step, in the ARIMA model, the difference between the initial calculation results and the real data is calculated (Figure 4), and the autocorrelation function (Figure 5) is constructed for this difference, and reports are continued. It is clear from Figure 5 that the next calculations will be made according to formula (3), taking into account (q = 3) in the ARIMA model. After solving the system equation obtained by writing the corresponding values in the formula (3), the values of () and are obtained (reports were made in the MATLAB 2023a software package). Calculation results are obtained by substituting these values into formula (3) (Figure. 6).

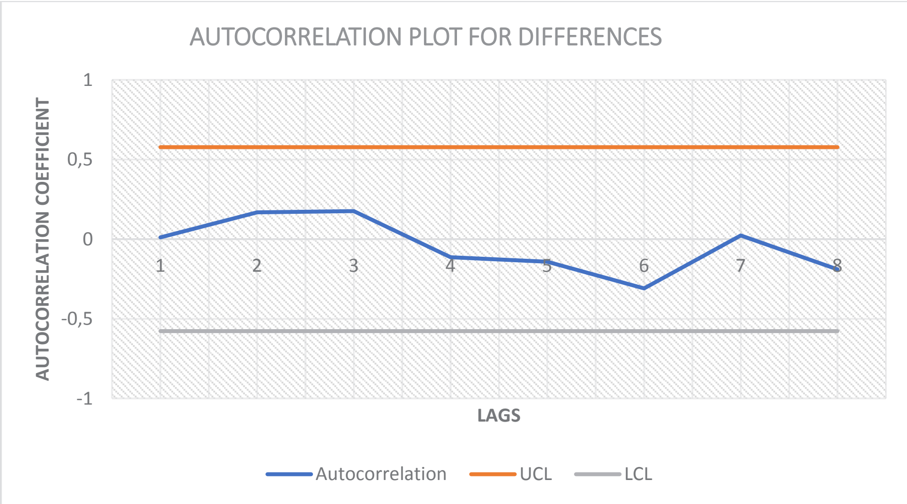

Figure 5 – The autocorrelation function of time series is calculated in the ARIMA model according to formula (2)

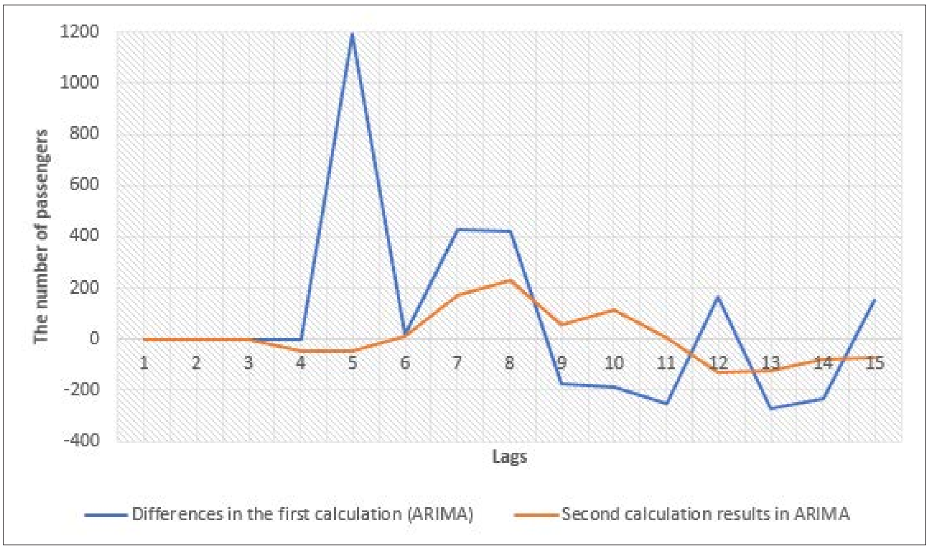

Figure 6 – Calculation results in the ARIMA model are based on the formula (3)

To determine the final calculation results in the ARIMA model (), (), () the values of c and variables are substituted in the formula (1), respectively, and the final calculation results are obtained. The mentioned calculation results are shown in Figure.7. It is clear from Figure.7 that the calculation results are quite optimal and close to the real data.

Figure 7 shows forecasting results for 2023 based on ARIMA-SVM models based on data for 2020–2022. The obtained forecasting results were compared with the actual indicators for 2023. As observed from Figure 7, the ARIMA model results are closer to the actual results, but there are sudden deviations at several points. The main reason for the occurrence of this situation is related to the characteristics of non-scheduled passenger air transportation. Typically, when building statistical models of this type, a number of smoothing methods are used to account for sudden deviations in the forecast results. Forecasting results obtained in SVM models based on kernel functions are also close to the actual indicators.

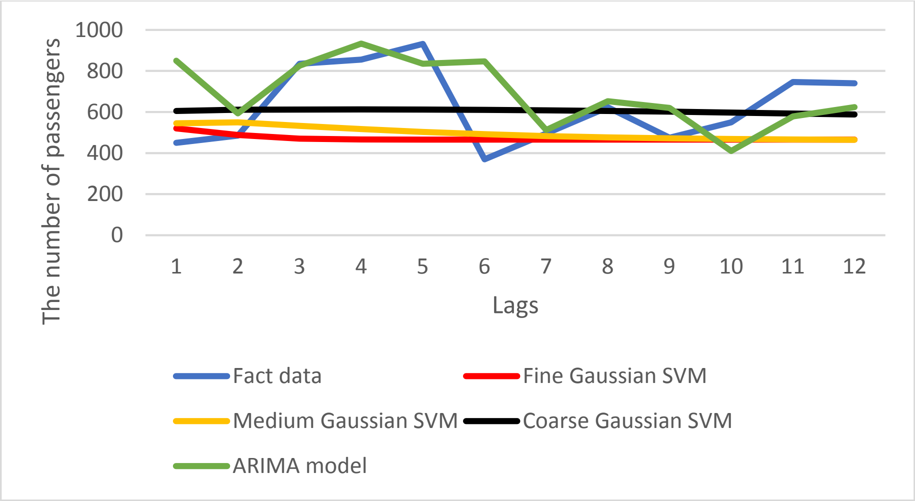

Figure 7 – Forecasting indicators of non-scheduled passenger air transportation based on ARIMA-SVM models

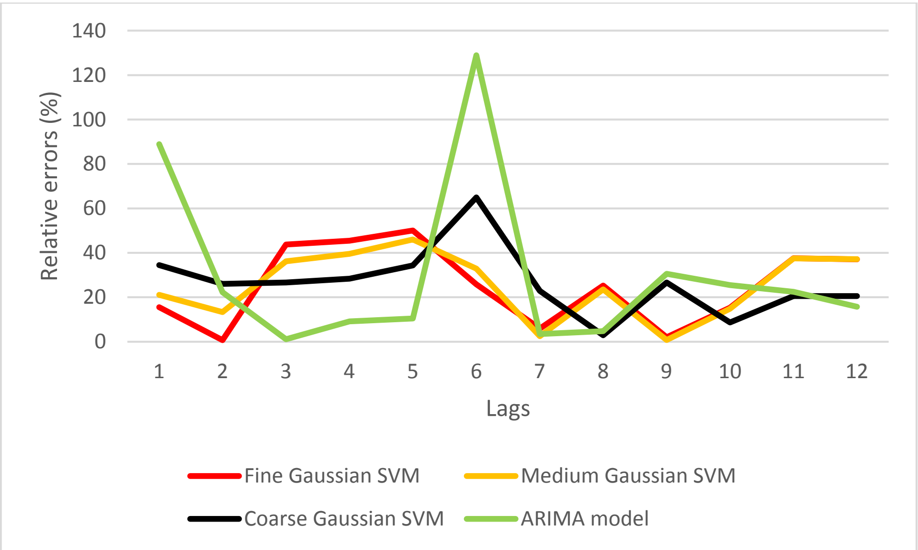

Figure 8 – Relative error of ARIMA-SVM models forecasting results based on actual indicators

Figure 8 shows the variation of the relative errors of the forecasting results obtained based on the ARIMA-SVM models with respect to the actual indicators. As can be seen from Figure 3, the relative errors of the models are, respectively, ARIMA (30.2%, Fine Gaussian SVM (25.2%), Medium Gaussian SVM (24.7%), and Coarse Gaussian SVM (28.1%).

6. Conclusion

In this paper, ARIMA and SVM (fine, medium, and coarse) methods are proposed for forecasting non-scheduled passenger air transportation. The results show that the forecasting results obtained based on the Medium Gaussian SVM model are more effective compared to the actual indicators, and the relative error is smaller than other models. When comparing the results of the ARIMA model, it is observed that the model expresses the general trend of the actual indicators, but there are sudden deviations in the forecasting values (this can be explained by the characteristics of non-scheduled passenger air transportation or the influence of extraneous variables not taken into account in the model). It is possible to overcome this non-linear problem by applying SVM models based on kernel functions. For this reason, the application of the ARIMA-SVM model in a hybrid form can be more effective in order to obtain more optimal forecast results. In conclusion, it can be noted that the results obtained in the article can be used in the application of neuro models in future studies.

Acknowledgments

We express our gratitude to Nadir Aghayev, professor of the Computer Systems and Programming department of the National Aviation Academy, for his profound contributions and close participation in the preparation of statistical data and the processing of mathematical formulas during the research.

Funding

This research received no external funding

Conflicts of Interest

The authors declare no conflict of interest

References

1. Aghayev, N., Nazarli, D. (2024). Support vector machines for forecasting non-scheduled passenger air transportation. Problems of Information Technology, 15(1), 3–9. https://doi.org/10.25045/jpit.v15.i1.01
2. Milenkovic M, Švadlenka L, Melichar V, Bojovic N, Avramovic Z (2016) SARIMA modelling approach for railway passenger fow forecasting. Transport 33:1113–1120. https://doi.org/10.3846/16484142.2016.1139623
3. Sun S, Sun S, Lu H, Tsui K, Wang S (2019) Nonlinear vector auto-regression neural network for forecasting air passenger fow. J Air Transp Manag 78:54–62. https://doi.org/10.1016/J.JAIRT RAMAN.2019.04.005
4. Madhavan M, Ali Sharafuddin M, Piboonrungroj P, Yang CC (2020) "Short-term forecasting for airline industry: the case of Indian air passenger and air cargo," Glob Bus Rev, https://doi.org/10.1177/0972150920923316
5. Lwesya F, Kibambila V (2017) "A comparative analysis of the application of seasonal ARIMA and exponential smoothing methods in short run forecasting tourist arrivals in Tanzania," Online. [Online]. Available: www.iiste.org
6. Mohd Lip N, Jumery NS, Ahmad Termizi FA, Mulyadi NA, Anuar N, Ithnin H (2020) Forecasting International Tourist Arrivals in Malaysia Using Sarima and Holt-Winters Model. J Tour Hosp Environ Manag 5(18):41–51. https://doi.org/10.35631/jthem.518004
7. Li C (2019) Combined forecasting of civil aviation passenger volume based on ARIMA-Regression. Int J Syst Assur Eng Manag 10(5):945–952. https://doi.org/10.1007/s13198-019-00825-6
8. Xu S, Chan HK, Zhang T (2019) Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. Transp Res Part E Logist Transp Rev 122(August 2018):169–180. https://doi.org/10.1016/j.tre.2018.12.005
9. Tang X., Deng G. (2016) Prediction of Civil Aviation Passenger Transportation Based on ARIMA Model. Open Journal of Statistics, 6, 824-834. https://doi.org/10.4236/ojs.2016.65068.
10. Y. Xie, P. Zhang, and Y. Chen, (2021) "A fuzzy ARIMA correction model for transport volume forecast," Mathematical Problems in Engineering, Article ID 6655102, 10 pages, 2021. https://doi.org/10.1155/2021/6655102
11. J. Wu, Z. B. Li, L. Zhu, and C. Li, (2017) "Hybrid model of ARIMA model and GAWNN for dissolved oxygen content prediction," Transactions of the Chinese Society for Agricultural Machinery, vol. 48, pp. 205–210. https://doi.org/j.issn.1000-1298.2017.S0.033
12. Tang X., Deng G. Prediction of Civil Aviation Passenger Transportation Based on ARIMA Model. Open Journal of Statistics, 6, 824-834. 2016. DOI: https://doi.org/10.4236/ojs.2016.65068.
13. Samagaio, A., Wolters, M., 2010. Comparative analysis of government forecasts for the Lisbon Airport. Journal of Air Transport Management 16, 213-217. DOI:10.1016/j.jairtraman.2009.09.002

14. Scarpel, R.A., 2013. Forecasting air passengers at São Paulo International Airport using a mixture of local experts model. Journal of Air Transport Management 26, 35-39. https://doi.org/10.1016/j.jairtraman.2012.10.001

15. Profillidis, V.A. (2012). An ex-post assessment of a passenger demand forecast of an airport. Journal of Air Transport Management, 25, 47–49. DOI: 10.1016/j.jairtraman.2012.08.002

16. K. Taneja, S. Ahmad, K. Ahmad and S. D. Attri, "Time series analysis of aerosol optical depth over New Delhi using Box-Jenkins ARIMA modeling approach," *Atmospheric Pollution Research,* vol. 7, no. 4, pp. 585-596, 2016. DOI: 10.1016/j.apr.2016.02.004

17. Suhartono, Lee, M.H., Prastyo, D.D. (2015). Two levels ARIMAX and regression models for forecasting time series data with calendar variation effects. *AIP Conference Proceedings*, 1691, no. 050026. http://scitation.aip.org/content/aip/proceeding/aipcp. DOI: 10.1063/1.4937108

18. Meng Ge , Zhang Junfeng , Wu Jinfei , Han Huiting , Shan Xinghua and Wang Hongye, Mathematical Problems in Engineering Volume 2021, Article ID 9961324, 5 pages https://doi.org/10.1155/2021/9961324

19. G. Ren and J. Gao, "Comparison of NARNN and ARIMA models for short-term metro passenger flow forecasting," in *Proceedings of the 19th COTA International Conference of Transportation Professionals g*, Nanjing, China, 2019. DOI: 10.1155/2021/9961324

20. Y. Xie, P. Zhang, and Y. Chen, "A fuzzy ARIMA correction model for transport volume forecast," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6655102, 10 pages, 2021. https://doi.org/10.1155/2021/6655102

21. M. M. H. Khan, N. S. Muhammad, and A. El-Shafie, "Wavelet based hybrid ANN-ARIMA models for meteorological Drought forecasting," *Journal of Hydrology*, vol. 590, Article ID 125380, 2020. https://doi.org/10.1016/j.jhydrol.2020.125380

22. J. Wu, Z. B. Li, L. Zhu, and C. Li, "Hybrid model of ARIMA model and GAWNN for dissolved oxygen content prediction," *Transactions of the Chinese Society for Agricultural Machinery*, vol. 48, pp. 205–210, 2017. DOI:10.6041/j.issn.1000-1298.2017.S0.033

Information About Author

Dashqin Nazarli is a PhD student of Air Transport Production department at National Aviation Academy (Baku, Azerbaijan, dnazarli.32073@naa.edu.az). His research interests include preparing forecasting methods for non-scheduled air transportation. ORCID iD: 0009-0001-8063-3670

Submission received: 03 August, 2024.
Revised: 16 August, 2024.
Accepted: 26 August, 2024.