Journal of Innovative Transportation, 2(1), 2103

Research Article

JOURNAL OF INNOVATIVE TRANSPORTATION

e-ISSN: 2717-8889

Airport passenger forecast with time series: case of study Samsun Çarşamba Airport

Tahsin Baykal[a,*], Fatih Ergezer[b], Serdal Terzi[c]

[a]Graduate School of Natural and Applied Sciences, Suleyman Demirel University 100/2000 Council of Higher Education PhD Scholarship Student, Turkey

[b]Department of Civil Engineering, Engineering Faculty, Suleyman Demirel University, Turkey

[c]Department of Civil Engineering, Engineering Faculty, Suleyman Demirel University, Turkey

Highlights

- The number of passengers at the airport was estimated by time series
- Correlation between the actual and estimated numbers of passengers is 0.628
- The number of passengers expected to occur in the future has decreased

Abstract

This study, it is aimed to estimate the number of airport passengers expected to be in the future with time series analysis. For this purpose, the total number of domestic and international passengers between January 2007 and May 2021 belonging to Samsun Çarşamba Airport was used. First, the normality test was applied to the data and it was observed that the data complied with the normal distribution. Using data from January 2007 to June 2018, the total number of passengers between July 2018 and May 2021 was estimated by time series. The estimated results were compared with the actual results and the R^2 value was 0.628. Then, with all data between January 2007 and May 2021, the total number of passengers expected to occur between June 2021 and March 2025 was estimated.

Information

Received: 28.06.2021

Received in revised: 05.07.2021

Accepted: 06.07.2021

Keywords: *Time Series, Transportation, Airline Transportation, Airline Transportation Planning*

1. Introduction

The service of transporting persons and cargoes from one place to another for the benefit of time and place is called "Transportation". Transport has social, economic, cultural, and political impacts on the development of society [1]. The transportation system, on the other hand, is defined as the set of all relevant physical, social, economic, and institutional components, whose functions and interactions are organized, brought together to transport people and goods from one place to another [2-4]. In recent years, the airline transportation sector has gained a more commercial dimension in our country and has become a large market where consumer demand is increasing. The number of private airlines serving domestic flights has increased. With the competitive environment, the travel prices in air transportation have decreased significantly and have become almost competitive with road transportation in intercity transportation. As a result of the increase in air transportation demand, new flight lines have been formed and the number of flights has increased [1].

Atay et al. [5] aimed to estimate the passenger, freight demand, and domestic fleet of the 3rd Istanbul airport, which was built as a substitute for Istanbul Ataturk Airport, in the coming years. They made predictions using artificial neural networks (ANN) and Adaptive Network-Based Fuzzy Logic Inference System (ANFIS) methods with historical data of Istanbul Ataturk Airport. As a result, they have obtained important information about the performance characteristics of the airport and whether the 3rd Airport will be able to respond to the possible demand as well as the estimated passenger and cargo values in the coming years. Efendigil and Eminler [6] have compared artificial intelligence techniques for passenger demand forecasting in the aviation sector with regression technique, which is an econometric model. In this study

*Corresponding author: tahsinbaykal@gmail.com (Tahsin Baykal, +90 246 2110835)

https://doi.org/10.53635/jit.958682

using ANFIS, ANN, and regression analysis, they found that the ANN technique gave the most successful and reliable results in passenger demand estimation in the aviation sector. Another aspect they aim at in the study is to find the factors that affect airline passenger demand. For this reason, they compiled the factors found in the literature, then subjected them to correlation analysis and determined the related ones. Önder and Kuzu [7] examined four air traffic density parameters: total passenger traffic, total cargo traffic, total air traffic, and total commercial air traffic. In addition, they have detected seasonal effects of these parameters. Classical time series methods such as moving averages as statistical analysis tools, exponential smoothing, Brown's single-parameter linear exponential smoothing method, Brown's circumstantial exponential smoothing method, Holt's two-parameter linear exponential smoothing method, and methods of components allocation of time series were implemented in air traffic in January 2007-May 2013. In the research, they focused on the applicability of classical time series methods (Smoothing and Parsing). In practice, they used four parameters of Turkish air traffic density.

This study aims to be able to estimate the number of passengers who will use air transportation in the future and create one of the evaluation criteria for strategic planning and investment forecasting of airlines.

2. Methodology

2.1. Data

The total number of domestic and international passengers of Samsun Çarşamba Airport between January 2007 and May 2021 was estimated using the time series method.

2.2. Time series

A time series is a set of data that consists of a chronological sequence of numerical data that occurs at certain regular periods related to a variable. Data on time series are stochastic, in other words, they receive random values at certain moments, and it is not possible to predict these values they receive [8]. The peculiarity of time series is that data is sorted by time, and sequential observations are often interdependent. Thanks to this inter-period dependency, reliable predictions can be created [9].

Time series (X_1, X_2, ..., X_i) are obtained by tracking the values received by a random variable in the range of Δt over time. In the event of statistical dependence between Xi and Xi+1 values in consecutive moments, it creates a stochastic process. It is not enough to determine the probability density function and parameters of variable X in statistical analysis of an example taken from a stochastic process time series. In addition, the internal dependence of the process should be examined and internal dependence can be measured by autocorrelation coefficients.

If the x_i and x_{i+k} values at the two moments with a time interval of kt are considered as two different random variables, the correlation coefficient between them is called $\rho_k = Cov^{X_i, X_{i+k}}/_{\sigma_x^2}$ as the k-range autocorrelation coefficient. This coefficient is estimated by Equation 1 from the n-element sample on hand.

$$r_k = \frac{\sum_{i=1}^{n-k}(x_i-\bar{x})(x_{i+k}-\bar{x})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad k = 1,2,\ldots \tag{1}$$

Here; xi is the time series data, n is the number of data, and k is the interval. However, the prediction obtained by Equation 1 is not an impartial estimate. The expression given in Equation 2 can be used to obtain an impartial estimate.

$$\bar{r}_k = r_k + \frac{1}{n-k} \tag{2}$$

Here, r_k is the value calculated by Equation 1 and the $\bar{r}_k$ is the neutral estimate. The standard deviation of the sampling distribution of the r_k is calculated as in Equation 3.

$$\sigma_{r_k} = \frac{(n-k-1)^2}{n-k} = \frac{1}{\sqrt{n}} \tag{3}$$

σ_{r_k}= The standard deviation of the sampling distribution. It is appropriate to perform a statistical analysis of a time series by dividing it into the following components.

Trend (trend), jump: The trend indicates that the parameters of the process (average, standard deviation) gradually increase or decrease over time. If the jump is sudden, it means an increase or decrease. In hydrological processes, trends and jumps occur at the end of natural influences such as landslides, forest fires, or deterioration of homogeneity due to human influence such as reforestation, dam construction, etc. In case of a trend in the process, this trend should be corrected by playing.

Periodic component: It is understood that there is a periodic component if the parameters of the process are changed with a certain T period. In this case, parameters can be expressed in Fourier openings.

Internal dependent component: X_{i-1}, X_{i-2}, ... values and statistical dependence.

Random (standalone) component: The other components above are the remaining components after extracting from the time series [10].

3. Result and Discussion

In the study, it was first determined whether the data matched the normal distribution. For this purpose, a normality test was performed in the SPSS package program. The statistical values of the data set are given in Table 1.

Table 1. Statistical values for the data set

Parameters	Values
Mean	645790
Median	568090
Standart devaition	433454
Minimum	40077
Maximum	1783839
Skewness	0.634
Kurtosis	-0.429

Hair et al. [11] Skewness measure of the symmetry of distribution; in most instances, the comparison is made to a normal distribution. A positively skewed distribution has relatively few large values and tails off to the right, and a negatively skewed distribution has relatively few small values and tails off to the left. Skewness values falling outside the range of -1 to +1 indicate a substantially skewed distribution.

When Table 1 is examined, the distortion coefficient is 0.634. This is proof that the data set complies with the normal distribution.

After the normalization test, the total number of passengers between July 2018 and May 2021 was estimated using the time series analysis and the total number of passengers between January 2007 and June 2018. The estimated total number of passengers for the years July 2018-May 2021 is given in Figure 1.

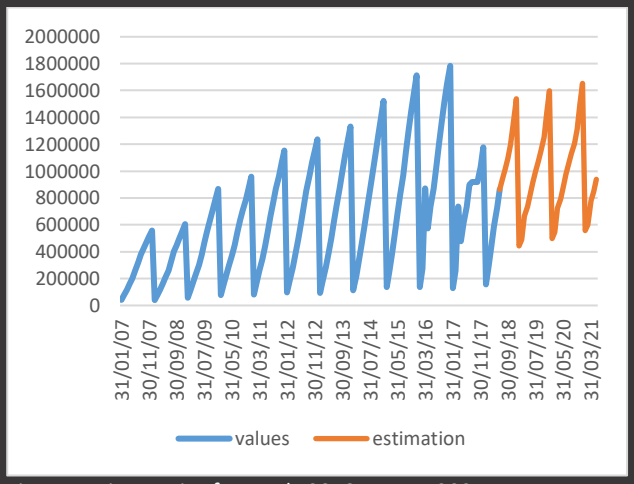

Figure 1. Time series from July 2018 to May 2021

In addition, the estimated values were compared with the actual total number of passengers between July 2018 and May 2021. As a result of the comparison, it was concluded that the time series analysis was successful in estimating the total number of passengers. Correlation coefficient (R2), Mean Absolute Scaled Error (MASE), Symmetric Mean Absolute Percentage Error (SMAPE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) values for the time series analysis result are given in Table 2 and the scatter diagram is given in Figure 2.

Table 2. Error-values for estimating the total number of passengers between July 2018 and May 2021

R^2	MASE	SMAPE	MAE	RMSE
0.628	0.96393	0.331675	163298.3	207711.4

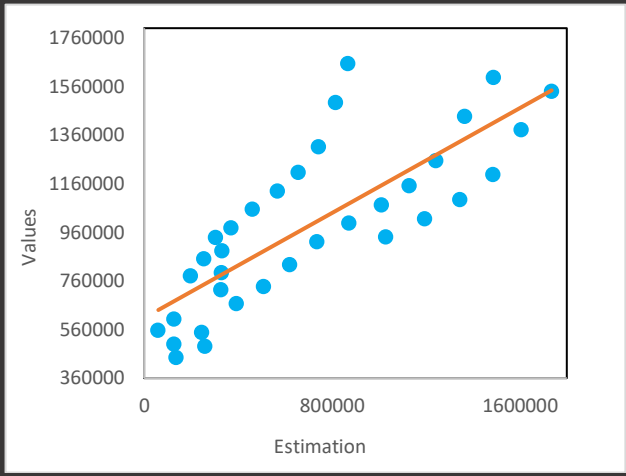

Figure 2. Estimated number of passengers between July 2018 and May 2021

Using data from January 2007 to May 2021, the total number of passengers for the 45 months between June 2021 and March 2025 was estimated by time series. Mase, SMAPE, MAE, and RMSE values were calculated as a result of the estimate. The calculated error values are given in Table 3.

Table 3. Error-values for time series analysis

MASE	SMAPE	MAE	RMSE
1.273823	0.535092	237269.895	261217.01

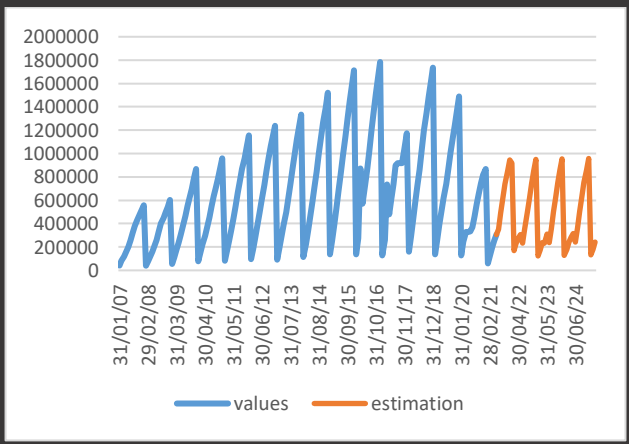

Figure 3. Estimated 45-month total number of passengers time series

When Figure 3 is examined, the most passengers are expected to arrive in December 2024 with 957982 people

and the least passengers in January 2023 with 123489 people.

4. Conclusion

Thanks to global developments and investments, air transportation has started to be in demand. Due to its advantages such as air transportation, speed, safety, and comfort, it is more preferred than land, rail, and sea transportation. For these reasons, it is of great importance to plan airline transportation correctly and reliably.

In this study, the total number of domestic and international passengers of Samsun Çarşamba Airport between January 2007 and May 2021 and the number of passengers expected to occur in the future were determined by the time series method. The total number of passengers between July 2018 and May 2021 was estimated using data from January 2007 to June 2018. The prediction results were compared with the actual results and it was observed that they produced good results. Then, using the total passenger numbers between January 2007 and May 2021, the total number of passengers expected to occur between June 2021 and March 2025 was estimated. When the forecast results are examined, the most passengers are expected to be in December 2024 with 957982 people.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contribution Statement

T. Baykal: Formal analysis, Methodology, Writing – Original Draft – **F. Ergezer:** Formal analysis, Methodology, Writing – Original Draft – **S. Terzi:** Methodology, Supervision, Writing – Review & Editing.

References

[1] Tortum, A., Gözcü, O., & Çodur, M. Y. (2014). Türkiye'de Hava Ulaşım Talebinin Arıma Modelleri ile Tahmin Edilmesi. *Journal of the Institute of Science and Technology,* 4(2), 39-54.

[2] Erel, A., (2001). Transportation Planning I and II Unpublished Lecture Notes. İstanbul.

[3] Yayla , N., (2002). Highway Engineering. İstanbul, Birsen Press Inc.

[4] Çodur, M.Y., (2012). Traffic accident prediction models: applications for surrounding highways of Erzurum. PhD Thesis, Atatürk University, Graduate School of Natural and Applied Sciences.

[5] Atay, M., Eroğlu, Y. & Seçkiner, S. U. (2019). Demand Forecasting of 3. Istanbul Grand Airport via Artificial Neural Networks and Adaptive Neuro Fuzzy Inference Systems for Optimization of Domestic Aircraft Fleet of Turkish Airlines. *Endüstri Mühendisliği*, 30(2), 141-156.

[6] Efendigil, T. & Eminler, Ö. E. (2017). The Importance of Demand Estimation in the Aviation Sector: A Model to Estimate Airline Passenger Demand. *Journal of Yaşar University*, 12, 14-30.

[7] Önder, E., & Kuzu, S. (2014). Forecasting air traffic volumes using smoothing techniques. *Journal of Aeronautics and space Technologies*, 7(1), 65-85.

[8] Gujarati, D., (2004). Basic Econometrics. USA, The McGraw Hill Companies.

[9] Vandaele, W., (1983). Applied Time Series and Box-Jenkins Models. Florida Academic Press Inc.

[10] Bayazıt, M. (1996). Probability Methods in Civil Engineering (1. Edition). İstanbul: Istanbul Technical University Civil Engineering Press.

[11] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2013). Multivariate Data Analysis. Essex, England.