JOURNAL OF STATISTICS & APPLIED SCIENCES

JOURNAL OF STATİSTİCS & APPLİED SCİENCES, VOLUME - 8
İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ, CİLT - 8
E-ISSN 2718-0999

ISSUE 8
JSAS, 2023

-JOURNAL OF STATİSTİCS & APPLİED SCİENCE-
-İSTATİSTİK VE UYGULAMALI BİLİMLER DERGİSİ-

Aralık 2023

Issue: 8

Sayı: 8

İmtiyaz Sahibi / Owner

Abdulkadir KESKİN

Baş Editör / Editor in- Chief

Dr. Abdulkadir KESKİN

Yayın Türü: 6 Aylık, Uluslararası, Hakemli

Publication Type: 6 Monthly, International, Refereed

e-ISSN: 2718-0999

Dizinleme Bilgileri/ Abstracting and Indexing Services

ASOS indeks	ACARINDEX	Google Scholar
doi Crossref	Academic Resource Index ResearchBib	CiteFactor Academic Scientific Journals

İletişim / Contact

E-posta: jsas.journal@gmail.com

https://dergipark.org.tr/tr/pub/jsas

Journal of Statistics & Applied Sciences, Issue - 8
ISSN 2718-0999

İletişim E-posta: jsas.journal@gmail.com : https://dergipark.org.tr/tr/pub/jsas

Journal of Statistics & Applied Sciences, Issue - 8
ISSN 2718-0999

İÇİNDEKİLER (CONTENT)

ARAŞTIRMA MAKALELERİ (RESEARCH ARTİCLES)

Mathematical Modeling of the Spread of Sars-Cov-2 at the Onset of Vaccination Against Covid-19 with CoronaVac in Türkiye **1-14**

Türkiye'de CoronaVac ile Covid-19'a Karşı Aşılamanın Başlangıcında Sars-Cov-2 Yayılımının Matematiksel Modellemesi

Ersin Sener, Ummu Sahin Sener

Assessing The Impact of Airlines' Strategic Decisions in Fleet Planning on Profitability by Implementing Panel Data Analysis **15-21**

Havayollarının Filo Planlamasındaki Stratejik Kararlarının Karlılık Üzerindeki Etkisinin Panel Veri Analizi ile Değerlendirilmesi

Cem Ersöz, Hüseyin Önder Aldemir

KİTAP İNCELEMELERİ (BOOK REVİEWS)

Araştırma Sanatı Wayne C. Booth, Gregory G. Colomb ve Joseph M. Williams (Çev. Ayşenur Aslan Fidan ve Deniz Aydın) Ankara: Nobel Yayınları, 2018 **22-24**

İbrahim Hakan Göver

Journal of Statistics & Applied Sciences, Issue - 8
ISSN 2718-0999

ÖNSÖZ

Sayın Okurlar,

Yeni bir yayın dönemine başlamanın heyecanını sizlerle paylaşmanın mutluluğunu yaşıyoruz. İstatistik ve Uygulamalı Bilimler Dergisinin bu sayısında, istatistik ve araştırma alanındaki güncel gelişmelere odaklanarak, birbirinden değerli iki makale ve bir kitap eleştirisi ile karşınızdayız.

İlk makalemiz, "Türkiye'de CoronaVac ile Covid-19'a Karşı Aşılamanın Başlangıcında Sars-Cov-2 Yayılımının Matematiksel Modellemesi" başlığını taşıyor. Ersin Sener ve Ummu Sahin Sener tarafından kaleme alınan bu makale, Covid-19'a karşı gerçekleştirilen aşılama sürecinin başlangıcındaki Sars-Cov-2 yayılımını matematiksel bir bakış açısıyla ele alıyor. Bu çalışma, pandemi yönetiminde alınacak stratejik kararlar konusunda ışık tutan önemli bir analiz sunmaktadır.

İkinci makalemiz ise Cem Ersöz ve Hüseyin Önder Aldemir tarafından kaleme alınan "Havayollarının Filo Planlamasındaki Stratejik Kararlarının Karlılık Üzerindeki Etkisinin Panel Veri Analizi ile Değerlendirilmesi" başlığını taşıyor. Bu çalışma, havayollarının filo planlamasında aldıkları stratejik kararların karlılığa olan etkisini panel veri analizi yöntemiyle değerlendirmekte olup, sektörde stratejik planlama üzerine yapılan önemli bir katkı sunmaktadır.

Ayrıca, bu sayımızda İbrahim Hakan Göver tarafından kaleme alınan "Araştırma Sanatı" isimli kitap incelemesi ile, Wayne C. Booth, Gregory G. Colomb ve Joseph M. Williams'ın kaleminden çıkan bu eser üzerinden araştırma sanatına dair önemli bir değerlendirme yapılmıştır.

Bu özel sayımızın hazırlanmasında emeği geçen tüm yazarlara ve hakemlere içten teşekkürlerimizi sunarız. Siz değerli okuyucularımızın, bu makale ve incelemelerle ilgili görüş ve düşüncelerini paylaşmalarını bekler, İstatistik ve Uygulamalı Bilimler Dergisinin bilimsel katkılarına katılma çağrısında bulunuruz.

Keyifli okumalar dileriz!

Saygılarımızla,

Dr. Abdulkadir Keskin

ISSN 2718-0999

PREFACE

Dear Readers,

We are thrilled to embark on a new publication cycle and share the excitement with you. In this issue of Journal of Statistics and Applied Sciences, we present two valuable articles and a book review, focusing on the latest developments in statistics and research.

The first article, titled "Mathematical Modeling of the Spread of Sars-Cov-2 at the Onset of Vaccination Against Covid-19 with CoronaVac in Türkiye," authored by Ersin Sener and Ummu Sahin Sener, delves into the mathematical modeling of Sars-Cov-2 transmission at the beginning of the Covid-19 vaccination period in Turkey. This study sheds light on strategic decisions for pandemic management.

The second article, written by Cem Ersöz and Hüseyin Önder Aldemir, is titled "Assessing The Impact of Airlines' Strategic Decisions in Fleet Planning on Profitability by Implementing Panel Data Analysis." This work evaluates the impact of strategic decisions in airline fleet planning on profitability using panel data analysis, contributing significantly to strategic planning in the sector.

Furthermore, in this issue, Ibrahim Hakan Göver provides a critical review of the book "Research Art" authored by Wayne C. Booth, Gregory G. Colomb, and Joseph M. Williams. This review offers a valuable assessment of research art based on this work.

We extend our sincere thanks to all the authors and reviewers who contributed to the preparation of this special issue. We invite our esteemed readers to share their thoughts and opinions on these articles and reviews, encouraging active participation in the scientific contributions of the Journal of Statistics and Applied Sciences.

We wish you pleasant reading!

Regards,

Dr. Abdulkadir Keskin

Journal of Statistics & Applied Sciences, Issue - 8
ISSN 2718-0999

Research Article

Received: date: 05.09.2023
Accepted: date: 28.10.2023
Published: date: 31.12.2023

Mathematical Modeling of the Spread of Sars-Cov-2 at the Onset of Vaccination Against Covid-19 with CoronaVac in Türkiye

Ersin Sener [1]* Ummu Sahin Sener [2]

[1]Department of Mathematics, Faculty of Science and Arts, Kırklareli University, Kırklareli, Türkiye; ersinsener@klu.edu.tr
[2]Department of Mathematics, Faculty of Science and Arts, Kırklareli University, Kırklareli, Türkiye; ummusahin@klu.edu.tr
Orcid: 0000-0002-5934-3652[1] Orcid: 0000-0001-9055-8734[2]
*Correspondence: ersinsener@klu.edu.tr

Abstract: The Sars-CoV-2 virus, first detected in Wuhan, China, became a global crisis that affected the entire world and was declared a pandemic by the World Health Organization (WHO) in March 2020. The most basic protective measure in the fight against pandemics facing humanity is vaccination. From this point of view, data is collected between January 13 and February 11, 2021 by taking the number of daily cases, deaths and recovered patients in Türkiye. During this period, vaccination against Covid-19 with Sinovac's CoronaVac vaccine is started in Türkiye. Mathematical predictive models of the observed values are constructed and compared using polynomial regression (up to the 3rd degree) and nonlinear regression, i.e., curve fitting methods, and **SIR** (**S**usceptible-**I**nfected-**R**emoved), which is a system of ordinary differential equations (ODEs). The efficiencies of these prediction models are tested, validated, and the most effective mathematical prediction models are proposed. The values of root mean square error (**RMSE**) and mean absolute percentage error (**MAPE**) are used as performance measures to compare the methods. The proposed prediction models are also used for forecasting. The number of new cases occurring each day is predicted using the time-dependent equations of the SIR method, which are solved using the Euler method. It is found that the SIR method is quite successful in predicting the observed values compared to the other methods, but the QR method are given more successful results in predicting the total number of deaths.

Keywords: Curve fitting, mathematical modeling, polynomial regression, SARS-CoV-2, SIR

1. Introduction

In early December 2019, a case of pneumonia of unknown etiology is discovered in Wuhan city, Hubei province, China, and the disease is reported to the World Health Organization (WHO) in late December 2019 [1]. In January 2020, a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is isolated from patients with infected pneumonia, and the disease caused by this virus is designated as coronavirus disease 2019 (COVID -19) in February 2020 [2, 3]. Since this virus spreads very rapidly within a few months and infects patients around the world, it is recognized as a pandemic by WHO in March 2020 [4]. Pandemic is the general term for epidemic diseases that spread rapidly over a large area in more than one country or continent in the world and cause deaths.

Building mathematical models using computational methods in pandemics is important to determine the rate of spread of the disease and the actions that need to be taken to prevent the spread of the disease. Mathematical modeling of the spread of epidemics has a long history and is initiated in the 1700s by Daniel Bernoulli, who developed a mathematical model to analyze the mortality caused by smallpox [5, 6]. Not much work had been done on this topic until the publication of Ross, which is developed in the

Citation: E. Şener and Ü.Ş. Şener , "Mathematical Modeling of the Spread of Sars-Cov-2 at the Onset of Vaccination Against Covid-19 with CoronaVac in Türkiye", Journal of Statistics and Applied Sciences, no. 8, pp. 1-14, Dec. 2023, doi:10.52693/ jsas.1355520

early 1900s and used a mechanistic a priori modeling approach with a set of equations to approximate discrete time dynamics that can be considered the foundation of mathematical epidemiology.

Mathematical and computational methods used in epidemiology can make important contributions to the spread, incidence, analysis, and control of disease [7]. The use of mathematical models in epidemiology has made it possible to define complex data, determine general rules for epidemic dynamics, predict parameters that cannot be directly measured, identify problems that might threaten public health, and select an optimal experimental design [8, 9]. The models are generally used to predict and explain trends in disease recurrence, spread, morbidity, or mortality [10]. With the correct interpretation of systems of dynamic equations, the development of analytical solutions, and the advancement of numerical methods, the methods used in epidemic modeling have evolved considerably [11]. By implementing infection prevention and control measures, it may be possible to reduce the spread of infection in the community and thus reduce the number of people who become infected in the early stages of the pandemic.

In the last 2 years, modeling the spread of the virus has been the main problem for researchers in the Covid-19 trial, which has negatively affected our lives and even stalled. A number of studies have used mathematical epidemiological models to analyze the transmission dynamics of COVID -19. SIR is one of the most commonly used mathematical epidemiological models [12]. Covid-19 data from seven countries-China, South Korea, Italy, Spain, Brazil, Germany, and France during the period from February to July 2020 are modeled using machine learning methods, SIR, and a time window SIR (TW-SIR) prediction model [13]. For Türkiye, Covid-19 data between March 11, 2020 and February 22, 2021 are used to examine monthly case counts with time series. In this study, a hybrid model of seasonal autoregressive integrated moving average (SARIMA) and neural network nonlinear autoregressive (NNAR) hybrid model is implemented [14]. The spread of Covid-19, many methods of time series analysis, and mathematical modeling methods of epidemiology are discussed very extensively [15].

In accordance with the information obtained from the literature review, it appears that mathematical epidemiological models are used for the early periods of Covid-19, but curve fitting methods are not used in these studies. The studies conducted to date have attempted to predict when the virus will peak. This study focuses on the period between January 13 and February 11, 2021, when vaccination against COVID -19 started and the first vaccine dose is administered. Our primary objective is to create mathematical prediction models for the spread of SARS-CoV-2 virus during the first dose of vaccination against SARS-CoV-2 coronavirus with Sinovac's CoronaVac vaccine, which is licensed for emergency use in Türkiye. The secondary objective is to propose an optimal model by comparing the success of the SIR model [13–16], which is a system of ordinary differential equations used in the construction of predictive models, linear regression (LR), polynomial regression (quadratic regression (QR) and cubic regression (CR)), and nonlinear regression (NLR) in estimating the general parameters of the spread of the pandemic [17]. In general, one of the objectives of this study is to determine which mathematical models can predict and create a priori the spread of the disease in the event of a possible pandemic in our globalized world. In addition, predictions of the daily number of cases and the daily number of deaths for the period February 12-15, 2021, are made using optimal mathematical predictive models.

2. Materials and Methods

2.1. Data of SARS-Cov-2

Our dataset consists of daily values of infected individuals (column 1), removed individuals (column 2), and deceased individuals (column 3) between January 13 and February 11, 2021 for Türkiye [18]. The dataset is split into two terms for modeling, the training term, and the test term. The first term is the training term between January 13 and February 5, 2021, and the training term is the percentage of 80 % of the total data. The second term is the test term between February 6 and February 11, 2021. The test term is used for model validation. The mathematical modeling processes of the data in the study are given in Figure 1 as a flow-chart.

In addition to the data set, the initial values needed to solve the model SIR using Euler's method are determined as follows: At $t = 0$ (January 13, 2021), there are susceptible individuals $(S_0 = 81992782)$, infected individuals $(I_0 = 104669)$, removed individuals $(R_0 = 2241616)$, and the population is $(N = 84339067)$.

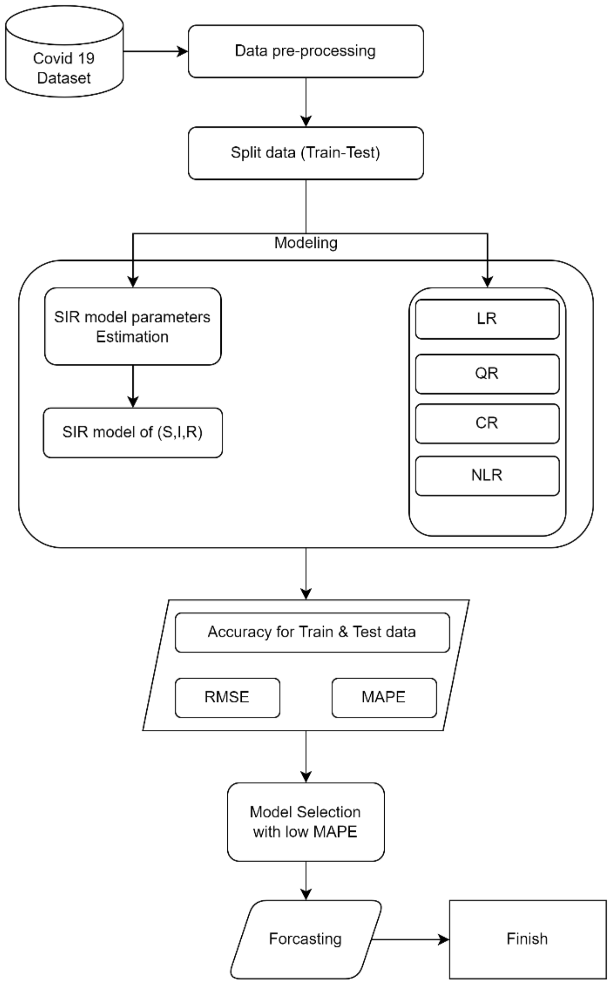

Figure 1. Mathematical modeling flow-chart

2.2. Methodology

2.2.1. SIR Model

In mathematical modeling of epidemics, the ordinary differential equation model, the so-called SIR (**S**usceptible-**I**nfected-**R**emoved (recovered and dead)) model, is one of the basic models. Individuals in a population of N are assumed to belong to one of three groups at time t [10].

$S(t)$: The class of individuals that are not infected now but will be infected later (Susceptible-S).

$I(t)$: The class of individuals who have contracted the disease and are now ill, i.e., individuals who infect others or are associated with an infection and infect others.

$R(t)$: The class of individuals who are removed (recovered and died).

The number of individuals in each of these groups changes with time, i.e. $S(t)$, $I(t)$, and $R(t)$ are functions of time t. The sum of the individuals in these three groups gives the total population size N in other words $N(t) = S(t) + I(t) + R(t) = constant, t \geq 0$. As long as the pandemic continues, people from the " S " group can progress to the " I " group, and people from the " I " group can progress to the " R " group.

$$\frac{dS}{dt} = -\beta S(t)I(t), S(0) = S_0 \tag{1}$$

$$\frac{dI}{dt} = \beta S(t)I(t) - \gamma I(t), I(0) = I_0 \tag{2}$$

$$\frac{dR}{dt} = \gamma I(t), R(0) = R_0 \tag{3}$$

$\beta > 0$ is the contact rate or infection rate of the disease, $\gamma > 0$ is the recovery rate from infected persons to recovered persons. In the Eq. (1), the infection rate of healthy people at a given time is proportional to the ratio of healthy to infected people. That is, it is proportional to the product of $S(t)$, and $I(t)$. The rate of change is negative because the healthy population is infected, and the number of healthy people always decreases. In Eq. (2) the rate of change of infected people is given by the difference between the rate at which healthy people become infected and the rate at which infected people move into the group of those who removed according to Eq. (3), the rate at which people who recover or die leave the infected group is directly proportional to the number of infected people. The relationship between $S(t)$, $I(t)$, and $R(t)$ can be seen in the flowchart of the model SIR in Figure 2.

$$S \xrightarrow{\quad\beta\quad} I \xrightarrow{\quad\gamma\quad} R$$

Figure 2. Flowchart of the SIR pandemic model

Assuming that the time value t goes to infinity $\lim_{t\to\infty} S(t) = S_\infty$ and $\lim_{t\to\infty} R(t) = R_\infty$. The number of infected individuals can fall to zero or behave nonmonotonically by first increasing to a maximum and then falling to zero. When $I'(0) = \beta S(0)I(0) - \gamma I(0) > 0$, prevalence begins to increase. The necessary and sufficient condition for the first increase in the number of infected is $\beta S(0) - \gamma > 0$ or $\frac{\beta S(0)}{\gamma} > 1$.

Dividing Eq. (1) by Eq. (3);

$$\frac{dS}{dR} = \frac{-\beta S(t)I(t)}{\gamma I(t)} = -\frac{\beta S(t)}{\gamma} \tag{4}$$

Solving this equation for t, Eq. (5) can be obtained as follows:

$$S = S(0)e^{-\beta R/\gamma} \geq S(0)e^{-\beta N/\gamma} > 0 \tag{5}$$

From this it follows that $S_\infty > 0$ and S_∞ is the final size of the pandemic, the pandemic is extinguished when $\lim_{t\to\infty} I(t) = I_\infty = 0$ and R_∞ is bounded by N. The equations given in Eq. (1-3) for the model SIR are ODEs that can be solved using Euler's formula.

$$S_{n+1}(t) = S_n(t) - \beta S_n(t)I_n(t)\Delta t = S_n(t)[1 - \beta I_n(t)\Delta t] \tag{6}$$

$$I_{n+1}(t) = I_n(t) + (\beta S_n(t)I_n(t))\Delta t = I_n(t)[1 + (\beta S_n(t) - \gamma)\Delta t] \tag{7}$$

$$R_{n+1}(t) = R_n(t) + (\gamma I_n(t))\Delta t \tag{8}$$

where $\Delta t = t_{n+1} - t_n$ is a small time change, $S_{n+1}(t)$, $I_{n+1}(t)$, and $R_{n+1}(t)$ are susceptible, infected and recovered individuals, respectively, calculated from the previous step.

2.2.2. Polynomial Regression

Regression analysis is one of the most common statistical methods to study and model the relationship between variables. It identifies the relationship between a dependent variable and one or more independent variables. Assuming a model of the relationship between variables and estimates of parameter values, a predictive regression equation is developed.

The linear regression model, the simplest regression model given by the equation $\hat{y} = X\beta + \epsilon$, is a general model used to construct any linear relationship in the unknown parameters β. $\hat{y} = \beta_0 + \beta_1 x + \epsilon$

is called the linear regression model [19]. x is the independent variable, that is predictor or regressor variable, and y is called the dependent variable or, in other words, the response variable. Since this equation involves only one regressor variable, this model is called a simple linear regression model. The linear regression model is a general model that can be used in cases where the relationship between explanatory variables and response variable is linear.
Even in complex nonlinear relationships, polynomials can be used extensively in situations where the response is curvilinear, since modeling can be done by fitting polynomials over small intervals of x. Polynomial regression is a special case of the general linear regression model and includes the quadratic and higher order values of the independent variable(s) to make the regression function curvilinear [20]. In general, a k^{th} order polynomial model for one variable is given in Eq. (9).

$$\hat{y} = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_k x^k + \epsilon \tag{9}$$

The model given in Eq. (9) can be solved by the least squares method. Taking $k = 1$, one obtains a linear function, taking $k = 2$, a quadratic function, and taking $k = 3$, one obtains a cubic function.
Unless required by the nature of the data or for other reasons, the degree of the polynomial used should be kept as low as possible. In our study, by using polynomials up to the 3^{th} degree, an attempt is made to find out at which polynomial of what degree the best result is obtained, and it is found that the most suitable curve was in the third-degree polynomial regression.

2.2.3. Non-linear Regression

Note that the shape and parameters of the curve can be determined by a nonlinear regression approach as in Eq. (10) if the data set can be represented by a nonlinear regression curve when the observed values are plotted [21].

$$\hat{y} = ae^{bx} + \epsilon \tag{10}$$

2.2.4. Model Goodness of Fit

The evaluation of the fitted prediction models can be done with the Root Mean Squared Error $(RMSE)$ and the Mean Absolute Percentage Error $(MAPE)$. $RMSE$ is the square root of the variance of the residuals and is given in Eq. (11).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)} \tag{11}$$

where n is the total number of observations, y_i is the i^{th} observed values, and y_i is the i^{th} predicted values. The fact that the *RMSE* value is very close to 0 indicates the absolute fit of the model to our data set (the lower *RMSE*, the better model fit). The second criterion in the examination of model fit, *MAPE*, is a measure of the estimation accuracy of an estimation method in curve fitting [22]. Usually, accuracy is expressed as a ratio defined by the following formula Eq. (12).

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{12}$$

where y_t is the observed value, and $\hat{y}_t$ is the predicted value at time t. The absolute value of the ratio in Eq. (12) is summed for each predicted time point and divided by the n number of observed values. The *MAPE* value is indeed very close to 1, indicating a relatively good fit to our data set. To investigate the optimal model fit for the datasets we have mathematically modeled, these two values are calculated. The proposed optimal model is constructed considering these criteria.

3. Results

This section is focused to the development of the models and the comparison of their performances. For this purpose, the data for the period February 6-11, 2021, corresponding to 20% of our dataset, are used as a test dataset to check the prediction performance of all the proposed models. The analyzes of the prediction models according to the explanatory variables are presented in the following subheadings.

3.1. SIR Model Parameters Estimation

When modeling the spread of Sars-Cov-2 virus using SIR, it is important to determine the model parameters. For this reason, the observed parameters γ-recovery rate and $\Re_0$-reproduction number are calculated from the data sets are given in Figure 3a-3b.

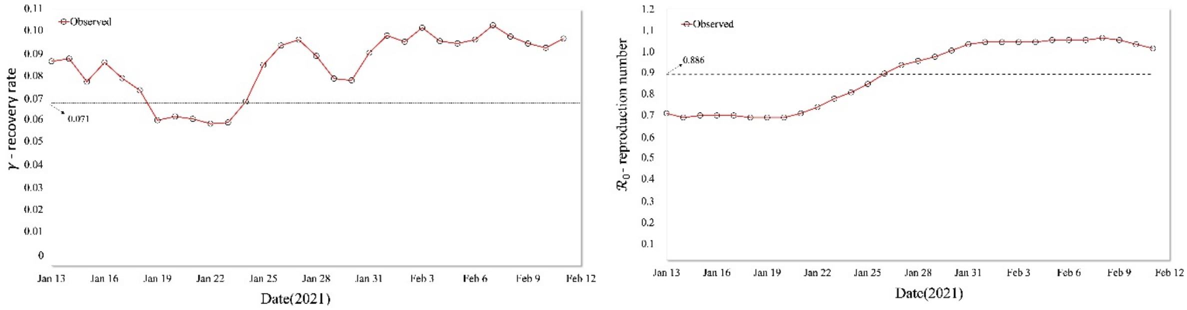

Figure 3. (a)- Daily Recovery rate, (b)-Daily Reproduction number

The daily recovery rate, which indicates that those infected with the virus are less likely to transmit the disease, i.e., that they have recovered, is calculated daily using Eq. (8) and shown in Figure 3a. The Ministry of Health of the Republic of Türkiye has set the average time for an individual infected with Sars-CoV-2 virus to recover at 14 days [23]. Therefore, the recovery rate for an infected individual is $\gamma = 1/14 = 0.071$, which is shown in Figure 3a. The daily recovery rate we calculated shows that the recovery period of infected individuals lasts more than 14 days [24]. However, considering the statements of the official authorities, the daily recovery rate was set as $\gamma = 0.071$ in the analyzes.

The average infectivity of transmission of the virus to another individual by an individual infected with Sars CoV-2 virus can be defined as reproduction.

The reproduction number $\Re_0$, which is an important value for determining actions to be taken depending on the course of the pandemic, is calculated daily [7]. $\Re_0$ is the average of the daily reproduction number $\Re_0 = 0.886$, as shown in Figure 3b.

At $\Re_0 > 1$ the risk of infection continues, $\Re_0 < 1$ the risk of infection decreases and may end, and $\Re_0 = 1$ the risk of disease remains constant [7]. The $\widehat{\Re}_0$ value calculated for the period in question is expected to end in the following days if current conditions are maintained.

3.2. Proposed Prediction Models of Susceptible (S)

The Susceptible (S) class of individuals that are not yet infected but may become infected later. Five different methods are used to model S in population N, namely SIR, linear regression (LR), quadratic regression (QR), cubic regression (CR), and non-linear regression (NLR), and the results of the predictive models are shown in Table 1.

The metrics *RMSE* and *MAPE* are used to determine the fitting success of curve fitting methods used to determine the curve that best predicts the available data. The $RMSE_S$ and $MAPE_S$ values calculated by each prediction method for the train and test terms are given in Table 1. The five different methods used to predict the S values are compared for the train and test terms into which we divided the data set to establish a model.

Table 1. Results of proposed prediction models of S

Model	Model Parameters	$RMSE_S$		$MAPE_S$	
		Train	Test	Train	Test
SIR	$\beta = 7.76e-10$ $\gamma = 0.071$	1604.66	2712.66	1.62E-03	3.23E-03
LR	$\beta_0 = 81988284$ $\beta_1 = -6777.42$	2217.73	7857.92	2.36E-03	9.22E-03
QR*	$\beta_0 = 81984576$ $\beta_1 = -5953.33$ $\beta_2 = -31.696$	1659.99	2185.15	1.74E-03	2.44E-03
CR	$\beta_0 = 81989358$ $\beta_1 = -7966.15$ $\beta_2 = 158.1$ $\beta_3 = -4.87$	858.21	8675.56	8.40E-04	9.84E-03
NLR	$a = 81988315$ $b = -8.27e-5$	2246.92	7355.40	2.42E-03	8.59E-03

When we examine the $RMSE_S$ and $MAPE_S$ values for the train term, the CR method yields the lowest $RMSE_S = 858.21$ and $MAPE_S = 8.40E-04$ values. However, the CR method has the largest error values with $RMSE_S = 8675.56$ and $MAPE_S = 9.84E-03$ in the test term. When estimating S, the $RMSE_S$ values are expected to be approximately the same for the train and test terms. The method QR, which has similar error values $RMSE_S = 1659.99, 2185.15$, for the model and test terms, respectively.
Instead of evaluating the metrics $RMSE_S$ and $MAPE_S$ separately for the train and test terms, the curve fitting method that provides the best fit is determined by evaluating both values together. From this point of view, it is decided that the QR, which gives the smallest change in both metrics for both periods, would be a useful curve fitting method for estimating the susceptible individuals. The model QR given in Eq. (13) is proposed as the optimal model for predicting S in the population. The QR-quadratic regression prediction model of S is given in Eq. (13).

$$\hat{y} = 81984576 - 5953.33 \times day - 31.696 \times day^2 + \epsilon \quad (13)$$

See Figure 4 for a tangible representation of the predictions of the methods and the observed values. The predictions of the models are given below in Figure 4 to see the extent to which the prediction methods we used to fit the curve predict the observed values.

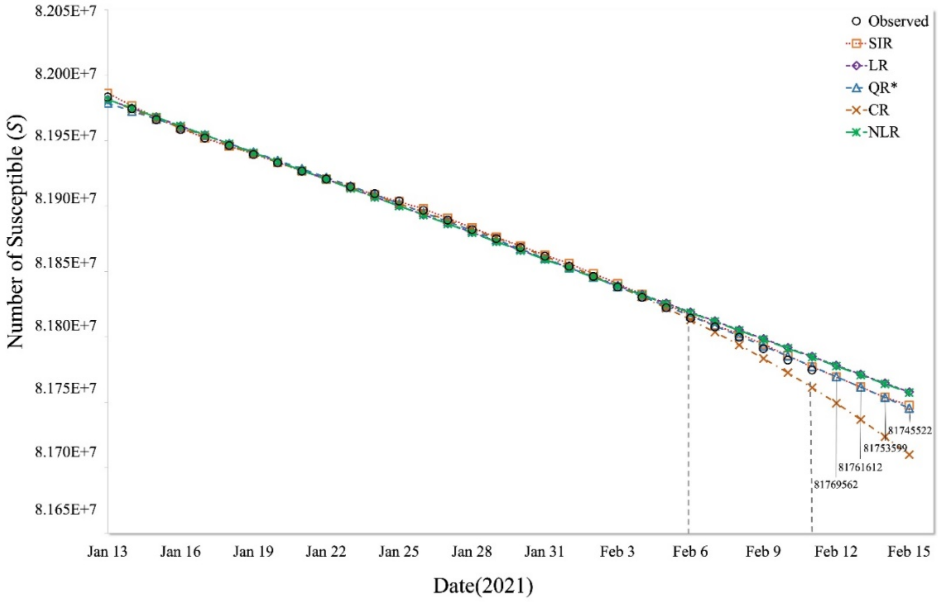

Figure 4. The graph of proposed prediction models of Susceptible (S)

The train and test term data are the observed values. Our goal is to build the predictive models over the train term and validate the built predictive models over the test term. To this end, in Figure 4, we see the predictions of S obtained using the method QR, and the predicted values of S are overlapped with the observed values. This overlap is quite successful. In addition, our forecasting is shown with the QR model between February 12-15, 2021, in Figure 4. In these projections, S is forecasted that will be approximately 81745522 via QR model by using Eq. (13) on February 15, 2021.

3.3. Proposed Prediction Models of Infected Individuals (I)

The class (I) of individuals who have the disease and are now infected, i.e., people who infect others or are associated with the infection and infect others. Five different methods are used to model I in population N SIR, LR, QR, CR, and NLR, respectively, and the results of the predictive models are shown in Table 2.

Table 2. Results of proposed prediction models of I

Model	Model Parameters	$RMSE_I$		$MAPE_I$	
		Train	Test	Train	Test
SIR*	$\beta = 7.76e-10$ $\gamma = 0.071$	681.19	902.40	0.579	0.927
LR	$\beta_0 = 105127$ $\beta_1 =$ -815.603	954.55	2166.10	0.817	1.97
QR	$\beta_0 = 105414$ $\beta_1 = -879.480$ $\beta_2 = 2.57$	947.72	1781.86	0.796	1.62
CR	$\beta_0 = 105378$ $\beta_1 = -864.11$ $\beta_2 = 1.008$ $\beta_3 = 0.037$	947.66	1707.44	0.794	1.57
NLR	$a = 105543.64$ $b = -0.009$	949.01	1639.94	0.792	1.54

The $RMSE_I$ and $MAPE_I$ values calculated for I from each prediction method for the train and test terms are given in Table 2. The five different methods used to predict the I values are compared for the train and test terms into which we divided the dataset to build a predictive model.

When we examine the $RMSE_I$ values for the train and the test terms, the method SIR, which is a system of ODEs, yields the lowest value with $RMSE_I = 681.19, 902.40$, respectively. When examining the $MAPE_I$ values for the train and the test terms, the method SIR provides the lowest value with $MAPE_I = 0.579, 0.927$. In predicting I, the values of $RMSE_I$, and $MAPE_I$ are expected to be approximately equal for the train and test terms. Thus, the model SIR is the most effective model for predicting I. The SIR prediction model of I is given in Eq. 14.

$$\hat{I} = 7.76e - 10 \times SI - 0.071 \times I \tag{14}$$

where $\hat{I}$ is the predictor of I. A projection of all the prediction methods and the observed values can be found in Figure 5. The predictions of the models are given below in Figure 4 to see the extent to which the prediction methods we used to fit the curve predict the observed values of I.

Predictions of I obtained with the SIR method overlap with the observed values of I, as can be seen in Figure 5. This overlap is quite successful. In addition, our forecasting with the SIR model during February 12-15, 2021, are shown in Figure 5. In these projections, I is forecasted that will be about 82693 for February 15, 2021, using the SIR model with Eq. 14.

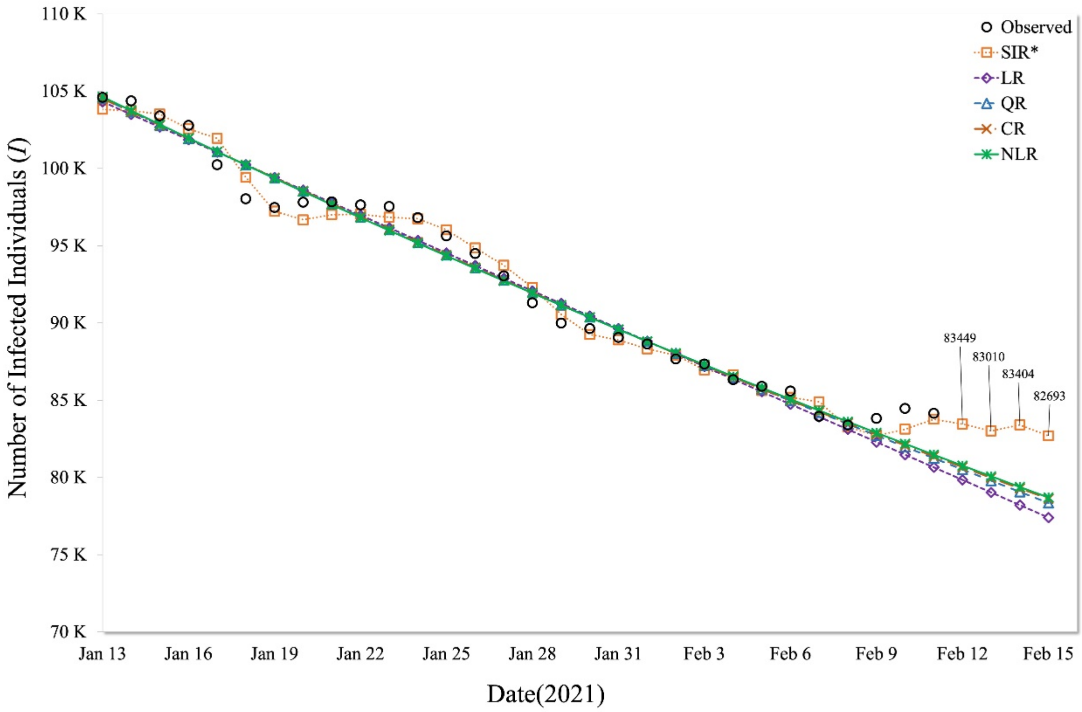

Figure 5. The graph of proposed prediction models of Infected (I)

3.4. Proposed Prediction Models of Removed Individuals (R)

The class of removed individuals (R) that are removed from the timely infected individuals (recovered and dead). Five different methods are used to model R in population N, namely SIR, LR, QR, CR, and NLR, and the results of the predictive models are shown in Table 3.

The $RMSE_R$ and $MAPE_R$ values calculated for R from each prediction method for the train and test terms are given in Table 3. In addition, Table 3 compares the five different methods used to predict R values for the train and test terms.

Table 3. Results of proposed prediction models of R.

Model	Model Parameters	$RMSE_R$		$MAPE_R$	
		Train	Test	Train	Test
SIR*	$\beta = 7.76e-10$ $\gamma = 0.071$	1646.72	902.40	0.0628	0.0908
LR	$\beta_0 = 2245656$ $\beta_1 = 7593.027$	2499.15	5943.90	0.0912	0.0239
QR	$\beta_0 = 2249077$ $\beta_1 = -6832.811$ $\beta_2 = 29.239$	2098.88	909.61	0.0702	0.0327
CR	$\beta_0 = 2244331$ $\beta_1 = 8830.256$ $\beta_2 = -159.108$ $\beta_3 = 4.829$	1554.66	10223.73	0.0561	0.373
NLR	$a = 2247053$ $b = 0.003$	2238.77	3698.38	0.0806	0.149

When examine the $RMSE_R$ values for the model term, the CR method yields the lowest $RMSE_R = 1554.66$ value. For the test term, the CR method has the largest error value with $RMSE_R = 10223.73$. In predicting R, the values of $RMSE_R$ and $MAPE_R$ are expected to be approximately the same for the training and test terms. The method SIR, which has similar error values with $RMSE_R = 1646.72, 902.40$, $MAPE_R = 9.08E-02$ for the train and test terms, respectively, is the most effective method in predicting R. The SIR prediction model of R is given in Eq. 15.

$$\hat{R} = 0.071 \times I \tag{15}$$

where $\hat{R}$ is the predictor of R. For a projection of all prediction methods and observed values, see Figure 6 for R. Figure 6 is shown below to visually illustrate the extent to which the prediction methods predict the observed values of R and to see the forecasted values for the 4-day period February 12-15, 2021.

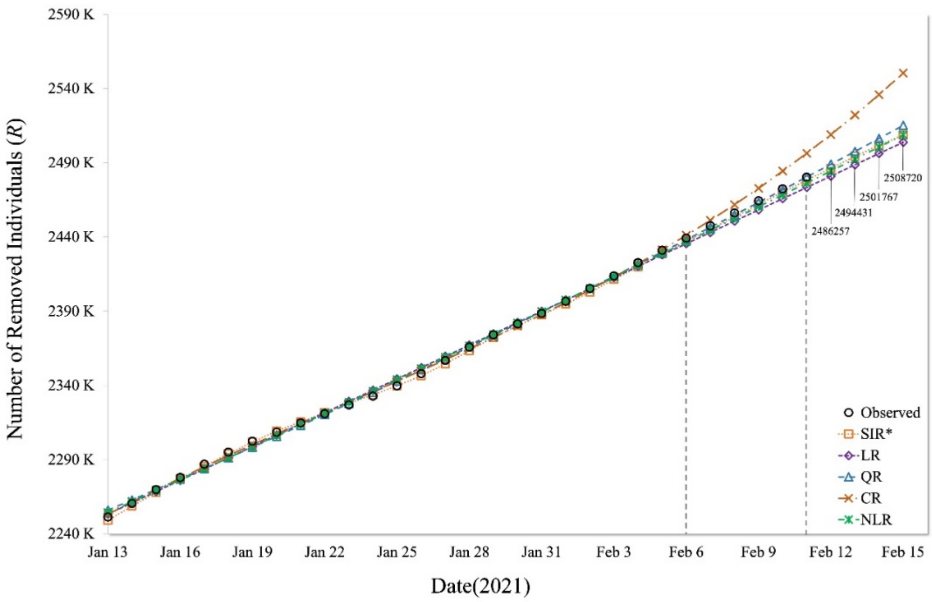

Figure 6. The graph of proposed prediction models of Removed individuals (R)

Predicting changes in the number of daily new cases (NC_{daily}) is of paramount importance to decision makers in the spread of the virus. Precisely for this purpose, we sought an answer to the question: **How to estimate the number of daily new cases?** By simple mathematical operations with the values S, I, and R that we calculated before this section, it is possible to predict the number of (NC_{daily}) by the equation in Eq. 16 to predict. The number of the NC_{daily} can be calculated in time t via SIR method in Eq. 16.

$$I_t^{SIR} + R_t^{SIR} = TC_t^{SIR} \tag{16}$$

where, I_t^{SIR} is the predicted value of I, R_t^{SIR} is the predicted value of R, and TC_t^{SIR} is the predicted value of Total Case (TC).The number of NC_{daily} can be calculated by subtracting the TC at time $t=1$ from $t=0$ (*time* as *day*), in other words, the increase in TC C gives the number of NC_{daily} in Eq. 17.

$$NC_{daily} = TC_{t+1}^{SIR} - TC_t^{SIR} \tag{17}$$

The observed and predicted values of NC_{daily} are shown in Figure 7. As can be seen in Figure 7, the predictions of NC_{daily} obtained by the method SIR via Eq. 17 obtained, overlap with the observed value. This overlap is quite successful. In addition to this overlap, our forecasting with the SIR during February 12-15, 2021 are shown in Figure 7. In these projections, the NC_{daily} is forecasted to add approximately 7865 patients to the infected individuals on February 15, 2021.

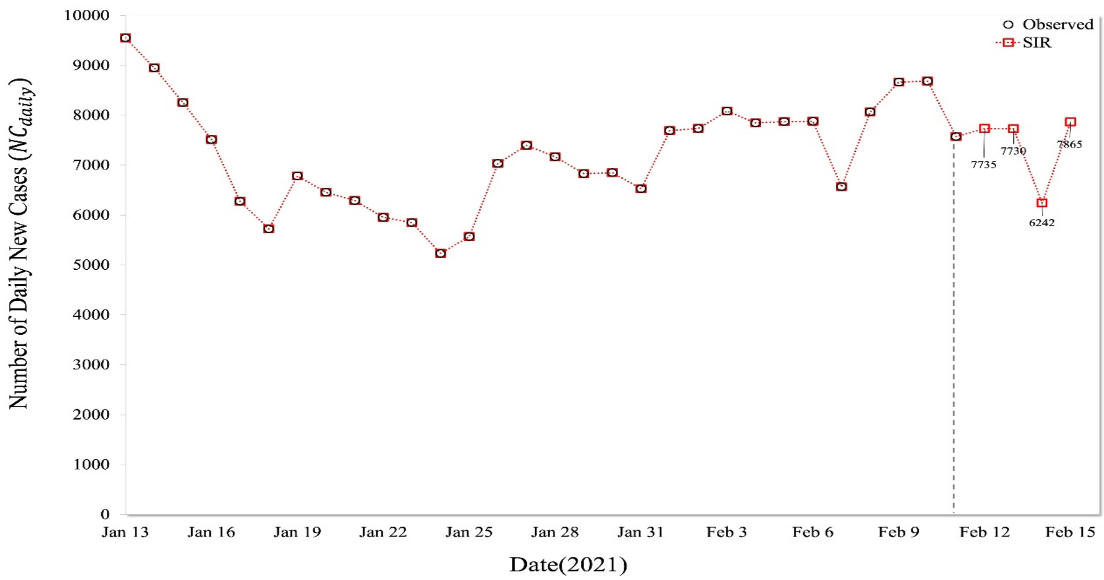

Figure 7. The graph of prediction models of daily new cases (NC_{daily}) via SIR method

3.5. Prediction Models of Total Death (D_{total})

Three different methods are used to model total mortality (D_{total}) in population N, namely LR, QR, and NLR, and the results of the predictive models are shown in Table 4.

The *RMSE* and *MAPE* values calculated for D_{total} from each prediction method for the train and test terms are given in Table 4. In addition, three different methods for predicting D_{total} values for the train and test terms are compared in Table 4.

Table 4. Results of proposed prediction models of D_{total}

Model	Model Parameters	$RMSE_R$		$MAPE_R$	
		Train	Test	Train	Test
LR	$\beta_0 = 23319$ $\beta_1 = 139.716$	67.04	243.85	0.231	0.874
QR*	$\beta_0 = 23151$ $\beta_1 = 177.104$ $\beta_2 = 1.438$	6.5	15.35	0.0218	0.0567
NLR	$a = 23369$ $b = 0.006$	84.93	309.08	0.293	1.11

When examining the $RMSE_{D_{total}}$ values for train and test terms, the QR method yields the lowest value with $RMSE_{D_{total}} = 6.5,\ 15.35$, respectively. When examining the $MAPE_{D_{total}}$ values for the train and test terms, the QR method provides the lowest value with $MAPE_{D_{total}} = 0.0218, 0.0567$.

In the predictor of D_{total}, it is expected that there will be approximately similar $RMSE_{D_{total}}$, and $MAPE_{D_{total}}$ values closest to 0 for train and test terms. Thus, the QR model is the most effective model for predicting D_{total}. The QR model of D_{total} is given below in Eq. 18.

$$\hat{y} = 23151 - 177.104 \times day - 1.438 \times day^2 + \epsilon \qquad (17)$$

The extent to which prediction methods predict the observed values of D_{total} and the forecasting values over the 4-day period between February 12-15, 2021 is given in Figure 8.

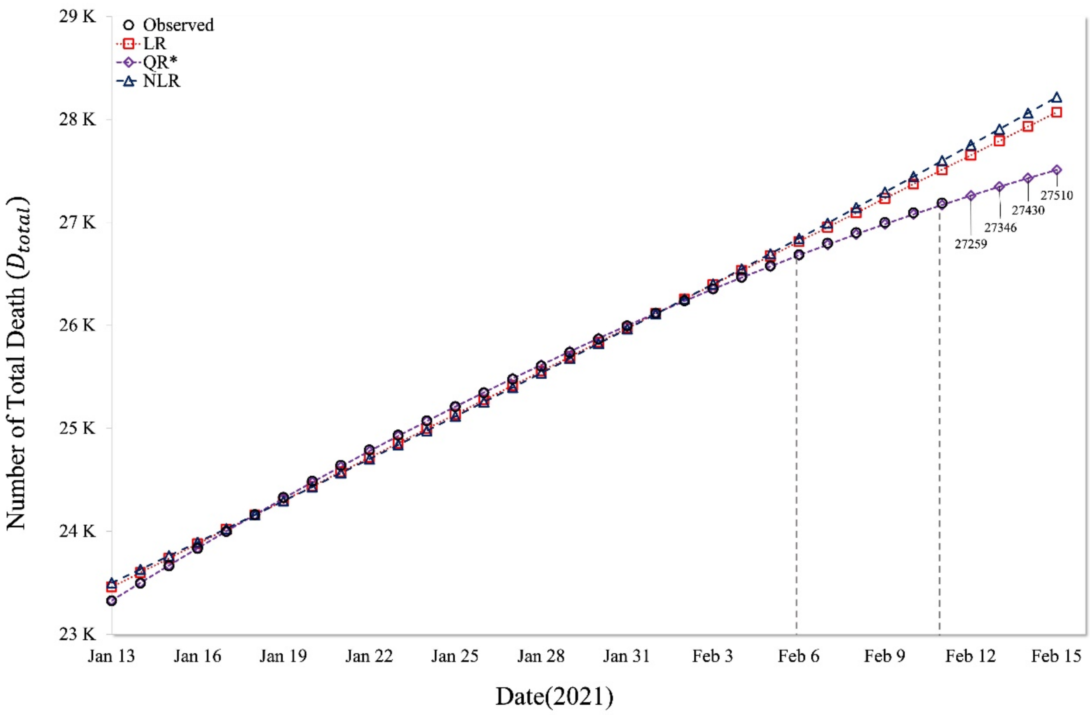

Figure 8. The graph of prediction models of daily new cases (D_{total})

The predictions of D_{total} obtained by the method QR are overlapped with the observations. This overlap is quite successful. In addition to this overlap, our forecasting is shown with the QR method between

February 12 and February 15, 2021 in Figure 8. In these results, it is forecasted that the number of D_{total} will be about 27510 individuals on February 15, 2021. The sum of those who recovered and those who died is the number of people who recovered from the Sars-CoV-2 virus. In this case, we need to look at what percentage of people who died from Sars-CoV-2 make up the total number of people infected. Considering this situation, the following Figure 9 shows what percentage of infected persons die every day in Türkiye.

In the period we studied, the average mortality rate of infected persons is 1.04 % percent. In other words, about one patient in 100 infected persons is dead.

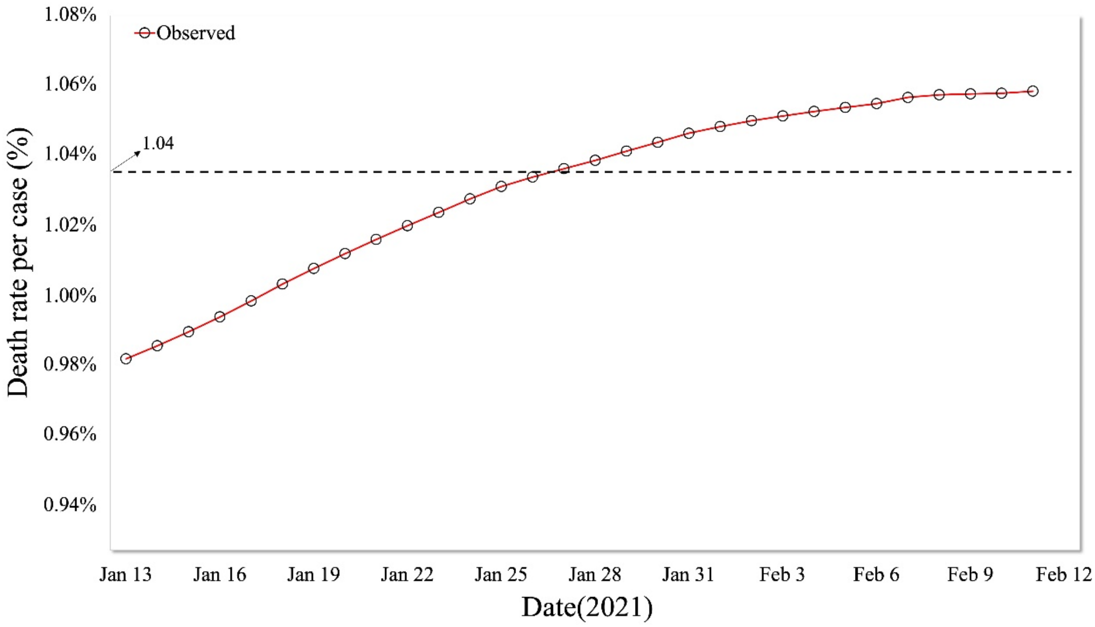

Figure 9. The graph of Death data per case (%)

4. Conclusions

In this study, which we conducted to investigate the situation of Covid-19 pandemic in Turkey, modeling, and further estimations for the daily number of recovered and deceased cases are made. The method SIR, the commonly used polynomial regression method (up to 3rd order) and nonlinear regression methods that are commonly used in modeling the spread of a disease are applied and these methods are compared to determine the most successful method.

In our analysis, the QR method best succeeded in predicting susceptible individuals in the population, whereas the SIR method was best able to predict infected and removed individuals. Modeling of the estimate of the number of new cases transmitted each day, which has not previously occurred, was demonstrated in this study using the method SIR. The powerful overlap of the observed values of the daily number of cases and the values estimated by the method SIR shows that the method with the value $RMSE(NC_{daily}) = 61.718$ is a very suitable method for our sample. The number of individuals who died due to Sars-CoV-2 shows that the QR method is the best predictor of the dataset. In this study, where we made a forecasting, on February 15, 2021, in the population $S = 81745522$, $I = 82693$, $R = 2508720$, $NC_{daily} = 7865$, $D_{total} = 25710$, and the mortality rate of individuals exposed to the virus is calculated as 1.04%.

Evaluating the research and forecasting regarding the pandemic's course together, it is expected that the pandemic will persist, as it is calculated to be $R = 69531$ on February 15, 2021. It seems possible to get rid of the Sars-CoV-2 virus by taking precautions to reduce human contact with each other and by continuing vaccination effectively and rapidly.

Author Contributions: Conceptualization, E.S. and U.S.S.; methodology, E.S. and U.S.S; software, E.S.; validation, E.S. and U.S.S; formal analysis, E.S.; investigation, E.S. and U.S.S; resources, E.S. and U.S.S; data curation, E.S.;

writing—original draft preparation, E.S. and U.S.S; writing—review and editing, E.S. and U.S.S; visualization, E.S.; supervision, E.S.; project administration, E.S.; funding acquisition, E.S. and U.S.S.

Funding: This research received no external funding. Data is available in a public-open access repository, the Health Ministry of Türkiye Republic data archive in [21].

Acknowledgments: Some of the results of this study were presented at the 10th International Eurasian Conference on Mathematical Sciences and Applications (IECMSA-2021).

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1] F. Zhou *et al.*, “Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study,” *The Lancet*, vol. 395, no. 10229, pp. 1054–1062, Mar. 2020, doi: 10.1016/S0140-6736(20)30566-3.
[2] A. L. Phelan, R. Katz, and L. O. Gostin, “The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance,” *JAMA*, vol. 323, no. 8, pp. 709–710, Feb. 2020, doi: 10.1001/jama.2020.1097.
[3] Coronaviridae Study Group of the International Committee on Taxonomy of Viruses *et al.*, “The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2,” *Nat. Microbiol.*, vol. 5, no. 4, pp. 536–544, Mar. 2020, doi: 10.1038/s41564-020-0695-z.
[4] J. W. M. Chan *et al.*, “Short term outcome and risk factors for adverse clinical outcomes in adults with severe acute respiratory syndrome (SARS),” *Thorax*, vol. 58, no. 8, pp. 686–689, Aug. 2003, doi: 10.1136/thorax.58.8.686.
[5] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, 2011. doi: 10.1515/9781400841035.
[6] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. John Wiley & Sons, 2000.
[7] F. Arroyo-Marioli, F. Bullano, S. Kucinskas, and C. Rondón-Moreno, “Tracking R of COVID-19: A new real-time estimation using the Kalman filter,” *PLOS ONE*, vol. 16, no. 1, p. e0244474, Jan. 2021, doi: 10.1371/journal.pone.0244474.
[8] A.-J. Valleron, “Roles of mathematical modelling in epidemiology,” *COMPTES RENDUS-Acad. Sci. PARIS Ser. 3*, vol. 323, no. 5, pp. 429–434, 2000.
[9] F. Brauer, “Lecture Notes in Mathematical Epidemiology”.
[10] M. Martcheva, *An Introduction to Mathematical Epidemiology*, vol. 61. in Texts in Applied Mathematics, vol. 61. Boston, MA: Springer US, 2015. doi: 10.1007/978-1-4899-7612-3.
[11] N. M. Ferguson, “Mathematical prediction in infection,” *Medicine (Baltimore)*, vol. 37, no. 10, pp. 507–509, Oct. 2009, doi: 10.1016/j.mpmed.2009.07.004.
[12] M. Abotaleb *et al.*, “Modeling Covid-19 Infection Cases and Vaccine in 5 Countries Highly Vaccinations,” *Turk. J. Math. Comput. Sci.*, vol. 13, no. 2, pp. 403–417, Dec. 2021, doi: 10.47000/tjmcs.905508.
[13] Z. Liao, P. Lan, Z. Liao, Y. Zhang, and S. Liu, “TW-SIR: time-window based SIR for COVID-19 forecasts,” *Sci. Rep.*, vol. 10, no. 1, Art. no. 1, Dec. 2020, doi: 10.1038/s41598-020-80007-8.
[14] İ. Demi̇r and M. Ki̇ri̇sci̇, “Forecasting COVID-19 Disease Cases Using the SARIMA-NNAR Hybrid Model,” *Univers. J. Math. Appl.*, vol. 5, no. 1, Art. no. 1, Mar. 2022, doi: 10.32323/ujma.1010490.
[15] A. Adiga, D. Dubhashi, B. Lewis, M. Marathe, S. Venkatramanan, and A. Vullikanti, “Mathematical Models for COVID-19 Pandemic: A Comparative Analysis,” *J. Indian Inst. Sci.*, vol. 100, no. 4, pp. 793–807, Oct. 2020, doi: 10.1007/s41745-020-00200-6.
[16] E. Eroglu, A. A. Esenpinar, E. Bozkurt, and S. Tek, “Mathematical Modeling of Covid-19 Phenomenon; The Cases: Germany, Israel and Canada,” *Fresenius Environ. Bull.*, vol. 29, no. 10, pp. 9063–9074.
[17] B. M. Ndiaye, L. Tendeng, and D. Seck, “Analysis of the COVID-19 pandemic by SIR model and machine learning technics for forecasting.” arXiv, Apr. 03, 2020. doi: 10.48550/arXiv.2004.01574.
[18] A. A. Toda, “Susceptible-Infected-Recovered (SIR) Dynamics of COVID-19 and Economic Impact.” arXiv, Mar. 26, 2020. doi: 10.48550/arXiv.2003.11221.
[19] H. (Howie) Weiss, “The SIR model and the Foundations of Public Health,” *Mater. Matemàtics*, pp. 1–17, 2013.
[20] B. Barrett, “Regression Analysis: Concepts and Applications,” *Technometrics*, vol. 37, no. 2, pp. 229–229, 1995, doi: 10.1080/00401706.1995.10484308.
[21] “Genel Koronavirüs Tablosu.” Accessed: Sep. 05, 2023. [Online]. Available: https://covid19.saglik.gov.tr/TR-66935/genel-koronavirus-tablosu.html
[22] H. L. Seal, “Studies in the History of Probability and Statistics. XV: The Historical Development of the Gauss Linear Model,” *Biometrika*, vol. 54, no. 1/2, pp. 1–24, 1967, doi: 10.2307/2333849.
[23] J. D. Gergonne, “The application of the method of least squares to the interpolation of sequences,” *Hist. Math.*, vol. 1, no. 4, pp. 439–447, Nov. 1974, doi: 10.1016/0315-0860(74)90034-2.
[24] S. Bakanlığı, “Temaslı takibi, salgın yönetimi, evde hasta izlemi ve filyasyon,” *Bilimsel Danışma Kurulu Ank. Halk Sağlığı Genel Müdürlüğü*, 2020.

Journal of Statistics & Applied Sciences, Issue - 8
ISSN 2718-0999

Research Article

Received: date: 19.11.2023
Accepted: date: 31.12.2023
Published: date: 31.12.2023

Assessing The Impact of Airlines' Strategic Decisions in Fleet Planning on Profitability by Implementing Panel Data Analysis

Cem Ersöz [1]* Hüseyin Önder Aldemir [2]

[1] Özyeğin University, Faculty of Aviation and Aeronautical Sciences, Department of Aviation Management, Istanbul, Türkiye
cem.ersoz@ozyegin.edu.tr
[2]Özyeğin University, Faculty of Aviation and Aeronautical Sciences, Department of Aviation Management, Istanbul, Türkiye
onder.aldemir@ozyegin.edu.tr
Orcid: 0000-0001-8010-5205[1] Orcid: 0000-0002-8083-0447[2]
*Correspondence: cem.ersoz@ozyegin.edu.tr

Abstract: This study investigates the impact on airline profitability of different aircraft types that airlines choose as a strategic decision for their flight operations. Datasets were gathered from the MIT Airline Data Project for ten airlines operating in the USA for the five-year period between 2015 and 2019. Three different panel data models- Pooled, Fixed Effects, and Random Effects- were employed to examine the effects of aircraft types (small narrow-body, large narrow-body, and wide-body) on profitability. The plm package of R language was used to create panel data models. In conclusion, the Fixed Effects Panel Data Model proved to be the most successful in explaining profit variation in all datasets. Variables determining airline profits change according to the airline specifications and are not time-dependent.

Keywords: Airline, profitability, aircraft type, strategic management, strategy, panel data model

1. Introduction

Due to the fragile nature of the airline industry, which can easily face financial crises, profitability stands out as the main factor for airlines to survive in their competitive environments. Political turmoil, outbreak of regional clashes, terrorist attacks, global pandemics, etc. create immediate economic fluctuations which firstly affect the operations of aviation industry [1]. Therefore, the airlines have had difficulty coming up with creative strategies to increase revenue. Furthermore, one of the key elements influencing profitability is the fierce and even unfair competition among airline businesses.

Statistics and firsthand knowledge from top airlines with varying business strategies in the aviation industry demonstrate that outsourcing business operations during a global pandemic has enabled carriers to better manage the negative effects of the external logistics environment, adapt quickly to changes in customer demand, and optimize costs based on workload [2]. All strategic decisions of airlines, from route planning to fleet formation, are aimed at increasing the profitability of the business and ensuring its survival. The profitability of airlines varies depending on many different factors. In the scope of this research, it is aimed to assess the impact of fleet structure, which is one of the most significant strategic decisions of the airlines, on profitability by applying different panel data models.

Forming the fleets of the airlines is a part of strategic management. One of the most important elements in the prosperity and profitability of an airline is effective fleet utilization [3]. The fleet structure of Full Service Carriers (FSCs) is different from that of Low Cost Carriers (LCCs) since LCCs operate short distance. According to MIT Airline Data Project, the aircraft in these fleets are three categories [3]:

- Wide-body (WB): Two-aisle configuration
- Large Narrow-body (LN): Typically, 151 seats or more in a two-class configuration (e.g. Boeing 737-800/900/Max 8/Max 9, Boeing 757, Airbus A321/A320 NEO/A321NEO)

Citation: C. Ersöz and H.Ö. Aldemir , "Assessing The Impact of Airlines' Strategic Decisions in Fleet Planning on Profitability by Implementing Panel Data Analysis", Journal of Statistics and Applied Sciences, no. 8, pp. 15-21, Dec. 2023, doi:10.52693/ jsas.1355520

• Small Narrow-body (SN): Typically, 150 seats or less in a two-class configuration (e.g. Boeing 737-700, Airbus A319)

The research question of this study: Are aircraft types and airline profitability significantly correlated statistically?

2. Literature review

Reference [4] searched how airlines raise their profitability towards liberalization. According to the findings of this research, profitable airlines have younger, more efficient fleets, high passenger load factors, and a small percentage of capacity-related costs. They also add freight to their passenger loads. Reference [5] indicated that labor productivity was the most important determinant of the profitability while on-time performance had no impact on profitability. According to their research, the average annual maintenance cost, labor productivity, gas price, and employee wage are all important indicators of profitability.

Reference [6] searched the influential factors on three largest Chinese airlines' (Air China, China Southern Airlines, and China Eastern Airlines) profitability between 2006 and 2019 by applying LASSO model. In the conclusion, the influential factors on airline profitability emerged as crude oil prices, exchange rate, volume of the passenger transportation while Chinese airlines' profitability did not rise in tandem with rises in GDP and per capita disposable income. Reference [7] examined the financial performances of sixteen airlines between 2004 and 2017 to determine the factors on profitability of Low Cost Carriers (LCCs) by employing panel data analysis. The results demonstrated that profitability is influenced by growth prospects, asset structure, and degree of debt. Reference [8] investigated the factors that contributed to Copa Airlines' long-term financial stability and profitability while other Latin American airlines experienced losses. Based on this research, five factors contributed to Copa Airlines' profitability: the airline's geographic location, which allows it to use narrow body aircraft throughout America; operations similar to low-cost carriers (LCCs) with a single aircraft; low market concentration of competitors on its routes; a cooperative and productive relationship with its hub airport; a dollarized domestic economy with strong GDP growth.

The relationship between service quality and profitability in airlines has been the subject of many academic studies and the effect of service quality on profitability has been analyzed. Reference [9] concluded that there is no meaningful correlation between customer rankings on SkyTrax (The World Airline and Airport Star Rating programme classifying airlines and airports by the quality of product and staff service standards) and operating profit margins for airlines. Thus, an airline that experiences high levels of customer satisfaction could also have low profit margins, and vice versa. This implies that airlines' short-term profit-oriented decision-making process may place a low value on service and customer satisfaction. Load factors and yields have even greater effects on airline profit margins, because they are mutually dependent. Reference [10] firstly studied the quality-profitability relationship in the US airline business by depending on Airline Quality Rating (AQR) Index. The study's findings demonstrate the AQR's considerable impact on US airlines' profitability. The profitability is also highly impacted by customer complaints, mishandled baggage, and on-time performance; on the other hand, the AQR component's denied boarding has a negligible impact on profitability. Then reference [10] showed and validated the positive and significant impact of service quality on the return on investment (ROI) of US airline companies, while quality was found to have a non-significant effect on airline passenger revenues by utilizing all four quality related indexes (American Customer Satisfaction Index-ACSI; Airline Quality Rating-AQR; JD Power Airline Satisfaction Index; Net Promoter Score-NPS) applied in the US airline industry.

Reference [12] searched the impact of global alliances on the profitability increase of founding member airlines by employing a difference-in-difference analysis; however, they couldn't find any proof that the establishment of international alliances increased the profitability of the founding member airlines or gave them a competitive advantage over non-founding members. Reference [13] conducted an empirical investigation of the combined benefits of code-sharing agreements and global alliances on airline profitability using a sample of 81 airlines between 2007 and 2012. The findings indicated that an airline's profit margin increases from code-sharing when a larger percentage of its code-sharing partners are members of the same global alliance. However, there was no discernible correlation between profit margin and the percentage of comprehensive code-sharing partnerships to total partnerships.

Reference [14] examined the airline profitability change by using panel dataset consisted of 53 airlines in the 1983-2010 period. They demonstrated that technical development has been a consistent primary driver of productivity change since 1990s. Additionally, over the past ten years, changes in input prices have mostly determined changes in profitability and have followed a similar pattern to changes in output prices. The fact that the increase in output prices is less than the increase in input prices when productivity growth is present suggests that some productivity benefits are passed on from airlines to customers.

Reference [15] investigated the operational performance and profitability of nine U.S. airlines between 2015 and 2019 by applying a two-stage network data envelopment analysis (DEA) model and a truncated regression. The results of this study demonstrated that airline companies may evaluate their resource allocation strategies regarding revenue structures, cost management, and the availability of various funding choices in order to improve efficiency at the profitability stage. Airlines using the low-cost business model outperformed their full-service counterparts in terms of efficiency. Although the size of an airline has an advantageous impact on operating efficiency, having more full-time employee equivalents has a negative impact on efficiency results, highlighting the significance of improving labor efficiency among carriers.

Reference [16] examined the impact of outsourced maintenance on eight U.S. passenger airlines' profitability by utilizing the datasets from Air Carrier Financial Reports between 1995-2019. employing four panel data methods: POLS, an individual fixed effects model, a two-way fixed effects model, and a time fixed effects model. They concluded that there was no meaningful correlation between airline profitability and outsourced maintenance.

3. Methodology

Data for this research is in a panel data format as shown in Table 1. It was gathered from MIT Airline Data Project website in xlsx format. The dataset consists of a balanced panel, which means it contains data for 10 airlines (cross-sectional units) over 5 years (time periods). In total, there are 50 observations.

Table 1. Data Sample

Airline	Year	Operating Income (loss) (millions USD)	SN Aircraft in Fleet	LN Aircraft in Fleet	WB Aircraft in Fleet
Alaska Airlines	2015	1,298	41	88	0
Alaska Airlines	2016	1,349	33	115	0
Alaska Airlines	2017	1,260	22	132	0
…	…	…	…	…	…

Income and fleet data were gathered for the operational airlines in the United States namely Alaska Airlines, Allegiant Air, American Airlines, Delta Airlines, Frontier Airlines, Hawaiian Airlines, Jetblue Airways, Southwest Airlines, Spirit Airlines and United Airlines between the 2015-2019 period.

Using natural logarithm of income (log(income)) as a dependent variable and small narrow-body (SN_Aircraft_in_Fleet), large narrow-body (LN_Aircraft_in_Fleet) and wide-body (WB_Aircraft_in_Fleet) aircraft numbers as independent variables, several panel data models were created using R's plm package.

4. Analysis

We would like to understand the effects of fleet size on the incomes of airlines. In particular, we would like to know how incomes change over time and across US airlines and how fleet size relates to this change.

4.1. Pooled Model

We start with a general pooled regression model where the coefficients of the regression equation are assumed to apply for all airlines for all years. Regression coefficients are displayed in Table 2. The intercept (5.79) is the estimated log income when all other variables are zero. In this context, it doesn't have a practical interpretation but serves as a reference point. A one-unit increase in the number of small narrow-body aircraft in the fleet is associated with an approximately 0.36% increase in income. Airlines with more small narrow-body aircraft tend to have higher incomes. A one-unit increase in the number of large narrow-body aircraft in the fleet is associated with approximately a 0.24% increase in income. This suggests that having larger narrow-body aircraft is positively related to airline income. Airlines with a greater number of widebody aircraft in their fleet tend to have higher incomes. A one-unit increase in widebody aircraft is associated with approximately a 0.53% increase in log income. All the coefficients are statistically significant with p-values less than 0.01, indicating a strong relationship between the number of aircraft in each category and airline income. The R-squared value of 0.852 suggests that the model explains approximately 85.2% of the variation in log-transformed income, indicating a good fit for the data. This means that the model captures a substantial portion of the variation in income explained by the number and type of aircraft in the fleet. The F-statistic is highly significant (p-value < 0.001), indicating that at least one of the independent variables is relevant in explaining log income. The model, as a whole, is statistically significant.

Table 2. Pooled Model Coefficients

	Estimate	Std. Error	t value	Pr(|>t|)
(Intercept)	5.792677743	0.111252711	52.06774474	1.53075E-42
SN_Aircraft_in_Fleet	0.003648559	0.000449679	8.113700519	2.01162E-10
LN_Aircraft_in_Fleet	0.002363979	0.000777106	3.042027123	0.00387397
WB_Aircraft_in_Fleet	0.005268682	0.001952085	2.699002076	0.009695056

4.2. Fixed Effects Models

In the fixed effects within model, individual-specific (airline-specific) effects are taken into account by calculating the differences within each airline over time. Regression results are displayed in Table 3. Examining the coefficients, the relationships between the number and type of aircraft in an airline's fleet and its log-transformed income are less pronounced compared to the pooled model. The number of small narrow-body aircraft in an airline's fleet is associated with a marginal increase of about 0.16% in log income, but this effect is not statistically significant. Similarly, the number of widebody aircraft in the fleet has a minimal impact on log income, with a coefficient that is not statistically significant. The number of large narrow-body aircraft, while showing a negative relationship with log income, is also statistically insignificant. The model, as indicated by the low R-squared value of 0.12174 and an F-statistic that is not statistically significant, suggests that the fixed effects within this context might not adequately capture the dynamics of the data, as the adjusted R-squared even becomes negative, raising concerns about its appropriateness.

Table 3. Fixed Effects Within Model Coefficients

	Estimate	Std. Error	t value	Pr(|>t|)
SN_Aircraft_in_Fleet	0.001596986	0.003001682	0.532030413	0.597885333
LN_Aircraft_in_Fleet	-0.00201158	0.001783633	-1.127799243	0.266665197
WB_Aircraft_in_Fleet	0.00514338	0.014442256	0.356134108	0.723762868

On the other hand, in the fixed effects between model, individual-specific (airline-specific) effects are captured by including a fixed effect for each airline. Regression results are displayed in Table 4. The model considers the differences between airlines but does not capture time-specific effects. Examining the coefficients, we find that the number of small narrow-body aircraft in an airline's fleet has a statistically significant positive effect on log income, with an estimated coefficient of approximately 0.35%. This suggests that for each additional small narrow-body aircraft in the fleet, the log-transformed income tends to increase. However, the number of large narrow-body aircraft and widebody aircraft in the fleet do not show statistically significant effects on log income. The model, as indicated by the R-squared value of 0.91095, explains a substantial portion of the variation in log income. The adjusted R-squared is also relatively high at 0.86642, indicating that the model provides a good fit to the data. The F-statistic is statistically significant, suggesting that the overall model is a good fit for the data. In this context, the Between Model appears to capture a more relevant and robust relationship between aircraft fleet composition and log-transformed income for the dataset, especially concerning the number of small narrow-body aircraft.

Table 4. Fixed Effects Between Model Coefficients

	Estimate	**Std. Error**	**t value**	**Pr(|>t|)**
(Intercept)	5.753380169	0.236072669	24.3712252	3.13747E-07
SN_Aircraft_In_Fleet	0.003473405	0.000972053	3.573266357	0.011737893
LN_Aircraft_In_Fleet	0.00303367	0.00180633	1.679466208	0.14406443
WB_Aircraft_In_Fleet	0.003849919	0.004417283	0.87155827	0.416955641

4.3. Random Effects Model

In this model, individual-specific (airline-specific) effects are captured as random effects. The model estimates two types of effects: idiosyncratic and individual. Idiosyncratic Effects: These represent the unexplained variation within individuals (airlines) over time. Idiosyncratic effects explain 35.5% of the total variation. These idiosyncratic effects capture unobservable airline-specific factors that influence income and are not accounted for by the variables in the model. Individual Effects: These common effects, which can be thought of as shared industry-wide factors, explain about 64.5% of the total variation in log income. The estimated theta parameter of 0.6849 indicates the proportion of the total variation that can be attributed to these common effects. A higher θ suggests that individual-specific differences play a more significant role in explaining the variation in the dependent variable, while a lower θ indicates that random fluctuations within individuals have a greater impact. Our result indicates that approximately 68.49% of the total variance in the operating income is attributed to systematic differences between individual airlines, while the remaining 31.51% is due to random fluctuations or idiosyncratic effects within each airline.

Coefficients of the random effects model are displayed in Table 5. We find that the number of small narrow-body aircraft in an airline's fleet has a statistically significant positive effect on log income. Each additional small narrow-body aircraft in the fleet is associated with an average increase of around 0.35% in income. In other words, expanding the fleet with small narrow-body aircraft tends to lead to higher log income. However, the number of large narrow-body aircraft and widebody aircraft in the fleet does not appear to have a statistically significant impact on log income, similar to the findings in the fixed effects between Model. The model overall explains a moderate portion of the variation in log income, as indicated by the R-squared value of 0.56739. This value suggests that a substantial part of the variation is still unexplained, and other factors beyond the variables included in the model are influencing income. The adjusted R-squared value of 0.53918, which is slightly lower, accounts for the number of model parameters and penalties for model complexity, providing a more conservative estimate of model fit. The chi-squared statistic is statistically significant, supporting the overall goodness of fit for the model. This Random Effect Model with Swamy-Arora's transformation allows us to consider both individual-specific and shared industry-wide effects when analyzing the relationship between log income and the composition of airline fleets. It highlights the importance of small narrow-body aircraft in influencing airline income, while also acknowledging the existence of individual-specific and unobservable factors that impact income variations among airlines.

Table 5. Random Effects Model Coefficients

	Estimate	Std. Error	z value	Pr(|>z|)
(Intercept)	5.926263658	0.223039203	26.5705023	1.4887E-155
SN_Aircraft_In_Fleet	0.003943312	0.000842003	4.683252815	2.82358E-06
LN_Aircraft_In_Fleet	0.000581265	0.001022091	0.568701937	0.569558433
WB_Aircraft_In_Fleet	0.00894311	0.003057269	2.925195458	0.003442399

4.4. Testing for Heteroscedasticiy and Serial Correlation

In the context of panel data analysis, the presence of heteroscedasticity and serial correlation can significantly affect the validity of our regression models. These issues were meticulously examined in the pooled, fixed effects, and random effects models to ensure the reliability of our findings.

In the pooled model, a Breusch-Pagan test was employed to test for heteroscedasticity. The results revealed no significant evidence of heteroscedasticity, as indicated by a p-value exceeding the conventional significance level of 0.05. This finding suggests that the assumption of constant error variance across observations is likely met. Consequently, the coefficients of the pooled model are robust and remain interpretable, enhancing the credibility of our analysis. Turning our attention to the fixed effects model, the Breusch-Pagan test revealed no statistically significant heteroscedasticity (p-value > 0.05). This implies that the variances of the idiosyncratic errors do not systematically vary with the predictor variables, and the assumption of constant variance holds for this model. Similarly, in the random effects model, the Breusch-Pagan test found no compelling evidence of heteroscedasticity (p-value > 0.05). These results indicate that the variances of the idiosyncratic errors across different groups (individuals) and time periods remain approximately constant. This reassures us that the assumptions underlying the random effects model are satisfied.

To scrutinize serial correlation, we utilized the Breusch-Godfrey/Wooldridge test for all three models. Remarkably, none of the models displayed statistically significant serial correlation (p-value > 0.05). This outcome supports the assumption that there is no autocorrelation in the idiosyncratic errors over time. The absence of serial correlation implies that the observations at different time periods are independent, reinforcing the reliability of our models.

Overall, the results of these tests offer strong reassurance regarding the integrity of the pooled, fixed effects, and random effects models. The absence of both heteroscedasticity and serial correlation underscores the suitability of our models for the panel data at hand. Consequently, the results and coefficients derived from these models are more likely to accurately reflect the underlying economic relationships, contributing to the robustness and trustworthiness of our panel data analysis.

5. Conclusion

In this study, we investigated the impact of fleet size on airline incomes, employing various panel data models to capture diverse effects across US airlines over the period of 2015-2019. Our analysis began with a pooled model, revealing significant positive relationships between income and the number of small narrow-body, large narrow-body, and wide-body aircraft in the fleet. The high R-squared value (0.852) indicated a strong model fit. Moving to Fixed Effects Models, the within model demonstrated limited significance in the relationships, questioning its appropriateness. Conversely, the between model highlighted a strong positive relationship for small narrow-body aircraft, suggesting it as a more robust representation of the dataset dynamics.

The Random Effects Model incorporated both idiosyncratic and individual effects, revealing that approximately 68.49% of the total variance in operating income is attributed to systematic differences between individual airlines. This model emphasized the significance of small narrow-body aircraft in influencing airline income. Additionally, we conducted thorough tests for heteroscedasticity and serial correlation across all models. The absence of significant findings in these tests enhances the credibility of our results, affirming the reliability of our models in reflecting the true economic relationships.

Our findings suggest that the fixed effects between model and the random effects model are more suitable for understanding the relationship between fleet composition and airline income. We recommend further exploration into airline-specific characteristics and strategies that contribute to the observed variations. Additionally, future research could delve into the potential impact of external factors, such as economic conditions or industry regulations, on airline profitability. The results also underscore the importance of considering both individual-specific and shared industry-wide effects in analyzing airline income dynamics.

Author Contributions: All authors contributed equally to this study.

Funding: This research did not receive external funding.

Conflicts of interest: The authors declare no conflict of interest.

References

[1] F. Yiğit and İ. Canöz, “Avrupa’daki̇ Hava yolu pay Senetleri̇ni̇n covid-19 pandemi̇si̇ne tepki̇si̇: Olay çalişmasi metodolojisi,” İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi, vol. 19, no. 39, pp. 1309–1326, 2020. doi:10.46928/iticusbe.831057

[2] Semeriahina, M., Hryhorak, M., & Bugayko, D. (2021). The Analysis of the Impact of Business Process Outsourcing on the Profitability of Commercial Airline in the Conditions of a Global Pandemic. Logistics and Transport, 49(1), 31-45.

[3] Massachusetts Institute of Technology (MIT) (2020). Airline Data Project- Aircraft and Related Retrieved from https://web.mit.edu/airlinedata/www/Aircraft&Related.html

[4] Antoniou, A. (1992). The Factors Determining the Profitability of International Airlines: Some Econometric Results. Managerial and Decision Economics, 13, 503-514.

[5] Parast, M. M., & Fini E. H. (2010). The effect of productivity and quality on profitability in US airline industry-An empirical investigation. Managing Service Quality, 20(5), 458-474.

[6] Xu X., McGrory, C.A., Wang, Y.G., & Wu, J. (2021). Influential factors on Chinese airlines’ profitability and forecasting methods. Journal of Air Transport Management, 91, 101969.

[7] Kiracı, K. (2020). The Factors Determining the Profitability of Low Cost Airlines. Romanian Statistical Review, 1, 41-53.

[8] O'Connell, J.F., Avellana, R.M., Warnock-Smith, D., & Efthymiou, M. (2020). Evaluating drivers of profitability for airlines in Latin America: A case study of Copa Airlines. Journal of Air Transport Management, 84, 101727.

[9] Merkert, R., & Pearson, J. (2015). A Non-parametric Efficiency Measure Incorporating Perceived Airline Service Levels and Profitability. Journal of Transport Economics and Policy, 49(2), 261–275.

[10] Kalemba, N., & Campa-Planas, F. (2017). The Quality – Profitability Link in the US Airline Business: A Study Based on the Airline Quality Rating Index. Procedia Engineering, 187, 308-316.

[11] Kalemba, N., & Campa-Planas, F. (2018). The quality effect on the profitability of US airline companies. Tourism Economics, 24(3), 251-269.

[12] Douglas, I., & Tan, D. (2017). Global airline alliances and profitability: A difference-in-difference analysis. Transportation Research Part A, 103, 432–443.

[13] Zou, L., & Chen, X. (2017). The effect of code-sharing alliances on airline profitability. Journal of Air Transport Management, 58, 50-57.

[14] Scotti, D., & Volta, N. (2017). Profitability change in the global airline industry. Transportation Research Part E, 102, 1-12.

[15] Huang, C.C., Hsu, C.C., & Collar, E. (2021). An Evaluation of the Operational Performance and Profitability of the U.S. Airlines. International Journal of Global Business and Competitiveness, 16, 73–85.

[16] Jin, L., Mott, J., Yang, C., & Lu, C. (2022). A Panel Study of Outsourced Maintenance Impact on Major U.S. Passenger Airlines' Profitability (1995-2019). Journal of Aviation/Aerospace Education & Research, 31(2), 1-32.

JJournal of Statistics & Applied Science, Issue - 8
ISSN 2718-0999

Received: date: 25.07.2023
Accepted: date: 28.10.2023
Published: date: 31.12.2023

Book Review

Araştırma Sanatı
Wayne C. Booth, Gregory G. Colomb ve Joseph M. Williams

(Çev. Ayşenur Aslan Fidan ve Deniz Aydın)

A**nkara: Nobel Yayınları, 2018, 336 sayfa, ISBN: 978-605-320-858-7**

İbrahim Hakan Göver [1]

[1]Abdullah Gül University, Department of Political Science and International Relations, Kayseri, Turkey; hakan.gover@agu.edu.tr
Orcid: 0000-0001-8877-3267[1]

Değerlendiren: **İbrahim Hakan Göver**

Bu çalışmada Wayne C. Booth, Gregory G. Colomb ve Joseph M. Williams'ın "Araştırma Sanatı" adlı eseri incelenmiştir [1]. Şikago Üniversitesi Yayınları tarafından "The Craft of Research" adıyla yayımlanan bu kitap Haziran 2018 tarihinde Ayşenur Aslan Fidan ve Deniz Aydın'ın çevirisi ve Doç. Dr. Necati Kayaalp'in editörlüğü altında Nobel Akademik Yayıncılık tarafından Türkçe'ye kazandırılmıştır. Toplam 336 sayfalık bu kitabın amacının sadece akademik çalışmalar için değil, aynı zamanda siyasi ve ticari amaçlı rapor ve proje yazımı için de yol gösterici bir rehber olduğu yazarlar tarafından belirtilmektedir.

Kitap, beş bölümden oluşmaktadır. "Araştırma, Araştırmacılar ve Okurlar" başlıklı ilk bölümde araştırma ve yazmanın önemine vurgu yapılıyor. Araştırmanın öneminden dolayı akademisyenlerin kendilerini araştırmaya adadıkları, hükümetler ve şirketlerin ise araştırmalar için milyonlarca dolar harcadıkları belirtiliyor. Yazarlar; araştırmayı, en geniş anlamıyla, bir sorunu çözmeye yönelik bilgi toplama faaliyeti olarak tanımlayıp araştırmanın her yerde örneğin laboratuvarda, evde, ofiste, ormanda, mağarada, hatta uzayda bile yapılan bir faaliyet olduğunu belirtiyorlar. Ancak, yazarlara göre bir araştırmayı değerli kılan onun yazıya dökülmesidir. Çünkü bilgi yazıldığı takdirde başkalarıyla paylaşır ve diğerlerinin buna yeni bilgiler eklemesi sağlanır. Bununla birlikte, araştırma yaparken sadece yazmak yeterli değildir, yazdıklarımızın belli bir düzen içerisinde de olması gerekir. Bunun nedeni ise kendimiz için değil başkaları için düşünmemiz ve yazmamızdır. Başkaları için düşünmek ve yazmak ise diğer her türlü düşünme ve yazma faaliyetinden daha fazla dikkat ve çaba gerektirir.

Kitabın "Sorular Sorma, Cevaplar Bulma" başlıklı ikinci bölümünde yazarlar araştırmaya hazırlık aşamasıyla ilgili konular üzerinde duruyorlar. Bu kapsamda "konulardan sorulara"; "sorulardan probleme"; "problemlerden kaynaklara" ve "kaynaklarla etkileşim" alt başlıkları altında bir araştırma konusu nasıl seçilir, nasıl daha spesifik hale getirilir, bir problemle nasıl ilişkilendirilir, kaynaklara nasıl ulaşılır ve kaynaklar nasıl daha etkin kullanılır gibi temel sorulara aşamalı bir biçimde yanıt aramaya çalışıyorlar. Yazarlar problemi günlük hayatta kaçınılması gereken bir olgu; akademik hayatta ise tam tersine üstüne gidilmesi, araştırılması gereken bir olgu olarak görüyorlar. Ancak araştırmanın kurgusu iyi yapılmaz ve hangi soruya hangi cevabın verileceği önceden iyi planlanmazsa yapılan araştırma okur tarafından sadece "boş bir bilgi yığını" olarak görülecektir. Bu noktada araştırmada doğru soruları sormak ve bu sorulara doğru çözümler bulmak gerekir. Bazen tek başına doğru bir soruyu sormak bile o soruyu cevaplamak kadar önemlidir. Yazarlar, bu bölümde konu seçimi ve araştırma sorusu sormanın yanı sıra kaynakları bulma ve bunları etkin kullanmanın önemine de vurgu yapıyor ve bu konuda okuyuculara faydalı olabilecek bazı ipuçları veriyorlar.

"Bir İddiada Bulunma ve Onu Destekleme" başlıklı kısım, kitabın üçüncü bölümünü oluşturuyor. Yazarlar, bu bölümde bir araştırmanın en can alıcı noktası olan tartışma bölümü ile ilgili kritik bilgiler veriyorlar. Araştırmanın tartışma bölümü kaynaklardaki verilerin rasyonel bir biçimde bir araya getirildiği, okura yeteri kadar delil sunularak onların ikna edilmeye veya sunulan deliller üzerinden bir soruna çözüm üretilmeye çalışılan bölümüdür. Yazarlar bu bölümde iddialar nasıl gerekçelendirilir, iyi bir argüman nasıl oluşturulur ve delillerle nasıl desteklenir sorusunu yanıtlıyorlar. Bu noktada gerekçe ve delil/kanıt arasındaki ayrıma işaret ediyorlar. Gerekçelerin zihinsel eylemin; delillerin ise dış dünyanın bir ürünü olduğunu belirtiyorlar. Bilimsel bir araştırmada iddia ve gerekçe arasındaki mantıksal bağın açık, sunulan delillerin ise inandırıcı olması gerektiğini, aksi takdirde çalışmaya eleştirel gözle bakan okurların gerekçeleriyle birlikte iddiayı da baştan ret edeceklerini vurguluyorlar. Çalışmanın argümanı ne derece başarılı kurgulanırsa, araştırma da okurlardan o oranda kabul alacaktır. Yine yazarlar bu bölümde, bilim dünyasındaki yenilik ve gelişmeler nedeniyle 30-40 yıl önceki gerekçelerin bugün artık kabul göremeyebileceğini ve okurun ön yargılı olduğu, kabul etmek istemediği konularda (insanların deneyimlerine, gözlemlerine, bilgilerine, kültürel ve dini değerlerine karşı, onlarla çelişen) gerekçeler göstermenin zorluğundan bahsediyorlar ve bu konuda çok dikkatli olunması gerektiğini belirtiyorlar.

Kitapta yer alan dördüncü bölüm "Planlama, Taslak Oluşturma ve Revize Etme" başlığını taşıyor. Bu bölümde yazarlar; çalışmanın planlanması, taslağının oluşturulması, görsellerin sunulması, giriş ve sonuç bölümlerinin yazılması ile çalışmanın revize edilmesi konuları üzerinde duruyorlar. Yazarlara göre akademik bir çalışma şekillendirilmeden önce giriş, metot-gereçler, bulgular, tartışma ve sonuç bölümlerini içeren bir çalışma taslağı ya da planı hazırlanmalıdır. Bu kapsamda özellikle araştırmanın giriş ve sonuç bölümlerini hazırlamanın önemine vurgu yapılıyor ve bu konuda bazı ipuçları veriliyor. Örneğin giriş kısmının okurların çalışmaya olan ilgisini artıran ve çalışmanın ilerleyen bölümlerini daha iyi anlamalarını sağlayan kritik bir bölüm olduğu, bu bölümün çalışmaya başlanırken önce taslak halinde hazırlanması, çalışma bitirildikten sonra ise yeniden gözden geçirilip tekrar yazılması gerektiği gibi. Yine yazarlar bu bölümde araştırma sırasında zaman zaman geriye dönülüp çalışmayı bir bütün halinde değerlendirmek ve çalışmanın ana temadan ayrılıp ayrılmadığını kontrol etmemiz gerektiğini belirtiyorlar. Çünkü yazarların nazarında araştırma, her aşamasında üzerinde yeniden durup düşünmeyi ve yeniden değerlendirmeyi gerektiren bilimsel bir faaliyettir. Bu yönüyle araştırma, düz bir yolda ilerlemekten çok inişli-çıkışlı bir arazide zikzak çizerek gitmek gibidir. Bu nedenle üzerinde çalışılan araştırma, periyodik geri dönüşler ve eleştirel okumalar yapılmadan hiçbir zaman olgunlaşmış ve son halini almış bir araştırma olarak kabul edilmemelidir.

Kitabın beşinci ve son bölümü olan "Bazı Son Hususlar" başlıklı bölümde ise yazarlar araştırma yaparken bilinmesinde fayda olabilecek bazı spesifik konulara değiniyorlar. Araştırma etiği ve bibliyografik kaynaklar bu bölümde ele alınan konulardan bazıları olarak belirtilebilir. Yazarlara göre bilimsel bir çalışma yaparken başkalarının görüşlerinden yararlanarak kendi görüşlerimizi ortaya koymaktayız. Bu nedenle araştırma yaparken intihal yapmamak, sonuçları yanıltmamak, kaynakları ve verileri doğru göstermek, doğruluğuna güvenilmeyen verilere araştırmada yer vermemek, aksini kanıtlayamadığı itirazları gizlememek, karşı görüşleri küçümsememek ve çarpıtmamak, verilere zarar vermemek ve kaynakları gizlememek gibi temel etik konulara özellikle dikkat edilmesi gerektiği yazarlar tarafından belirtilmektedir. Bu etik konulara dikkat etmek, başkalarının ortak yararını kendi kişisel yararımızdan üstün tutmak anlamına gelmektedir. Ayrıca, kitabın bu bölümünde araştırma yapan ve araştırma konusunda kendi alanında spesifik bilgilere ihtiyaç duyan okurlara faydalı olabilecek nitelikte geniş ve zengin bir bibliyografik çalışmaya da yer veriliyor.

Üniversitelerin geçirdiği değişim ve 3. Nesil üniversiteler olarak günümüz üniversitelerinin artık eskisi gibi sadece eğitim vermek yerine araştırma yaparak bu araştırmaları toplumun faydasına sunmaları günümüz akademik camiasında araştırma yapmanın önemini daha da artırmıştır [2]. Ayrıca, çeşitlenen bilim dalları ve artan akademisyen sayıları nedeniyle günümüzde her zamankinden daha fazla bilimsel araştırma yapılmaktadır. Ancak, yapılan araştırmaların pek azının bilime açık bir katkısı olduğu ve usulüne uygun yapıldığı söylenebilir. İşte Booth, Colomb ve Williams'ın Araştırma Sanatı adlı bu kitabı, bu soruna parmak basmakta ve her kesimden araştırmacıya nasıl daha nitelikli bir çalışma yapabilecekleri ve bu konuda nelere dikkat etmeleri gerektiği konusunda temel ve önemli tavsiyelerde bulunmaktadır. Piyasada araştırma kitabı olarak Türk ve yabancı yazarlar tarafından yazılmış çok sayıda eser bulunmakla birlikte, bu çalışma çoğu kişinin araştırma yaparken göz ardı ettiği temel konulara yer vermesi, pratik ipuçları ve örnekler sunmasıyla benzerlerinden bir adım öne çıkmaktadır. Bu nedenle, bu kitabın özellikle yüksek lisans ve doktora tezi hazırlayacak olan lisansüstü öğrencilere faydalı olabileceğini belirtmek mümkündür. Ayrıca, kitabın akademik camiada görev yapan ve araştırma konusunda lisansüstü dersler veren öğretim elemanları için de bir başucu/referans kitap olarak katkı sağlayabileceği düşünülmektedir.

Diğer taraftan, kitapta bazı eksikliklerin olduğu da görülmektedir. Örneğin kitabın sonuna alanında öne çıkmış araştırmalardan birkaç örneğe yer verilmesi kitapta anlatılanları daha somut hale getirecektir. Son olarak, kitap araştırma konusunda temel bilgiler sunmanın ötesinde araştırma modelleri ve desenleri gibi daha ileri düzey konulara da yer vermiş olursa hem daha kapsamlı hale gelmiş hem de hitap ettiği okuyucu kitlesini genişletmiş olacaktır.

Kaynakça

[1] Booth, W. G., Colomb, G. G. ve Williams, J. (2018). Araştırma Sanatı, Çev: Ayşenur Aslan Fidan ve Deniz Aydın, Editör: Necati Kayaalp, Nobel Akademik Yayıncılık: Ankara.

[2] Wissema, J. G. (2009). Towards the Third Generation University: Managing the University in Transition, Edward Elgar Publishing

31.12.2023

Journal of Statistics & Applied Sciences, Issue – 8
ISSN 2718-0999