Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A close look at 2019 novel coronavirus (COVID 19) infections in Turkey using time series analysis & efficiency analysis

Harun Knaci a, Mehmet Güray Ünsal b, Reşat Kasap c

a Erciyes University, Faculty of Economics and Administrative Sciences Department of Statistics, Kayseri, Turkey
b Usak University, Faculty of Art and Science, Department of Statistics, Usak, Turkey
c Gazi University, Faculty of Science, Department of Statistics, Ankara, Turkey

A R T I C L E   I N F O

Article history:
Received 16 May 2020
Revised 17 November 2020
Accepted 10 December 2020
Available online 23 December 2020

Keywords:
COVID 19
Pandemic
Time series analysis
Efficiency analysis

A B S T R A C T

2019 novel coronavirus (COVID 19) infections detected as the first official records of the disease in Wuhan, China, affected almost all countries worldwide, including Turkey. Due to the number of infected cases, Turkey is one of the most affected countries in the world. Thus, an examination of the pandemic data of Turkey is a critical issue to understand the shape of the spread of the virus and its effects. In this study, we have a close look at the data of Turkey in terms of the variables commonly used during the pandemic to set an example for possible future pandemics. Both time series modeling and popular efficiency measurement methods are used to evaluate the data and enrich the results. It is believed that the results and discussions are useful and can contribute to the language of numbers for pandemic researchers working on the elimination of possible future pandemics.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

At the end of 2019, 2019 novel coronavirus (COVID 19) infections, which is thought to have emerged in the Wuhan region of China and spread to the whole world in a short time, became one of the crucial secretions threatening the world health. Due to its infectious feature, it spreads very quickly in Asia, Europe, America and Africa continents.

COVID 19 is a contagious disease and the virus is called as SARS-CoV-2. The symptoms of COVID 19 is similar to influenza. However, old and/or people with certain illness are generally worse affected. Because of having no cure for this virus, the number of COVID 19 cases have been increasing and this disease becomes one of the world’s major health problems. Prevention strategies, such as keeping social distance, frequent hand washing, and wearing a medical masks, can help protect people from the 2019 novel coronavirus infections [1].

COVID 19 affects the health system of many developed and developing world countries negatively. Turkey, the United States of America, Spain, Italy, Russia, England, France, Iran Germany and Brazil are some of the most negatively affected countries in the world. Although Turkey started a long process of resistance at the beginning of the epidemic, a long pandemic process started by detecting the first case on 10th March. Because of being a geographical bridge between Asia and Europe, having many international airports and socioeconomic structures based on international relations, the spread of the pandemic in Turkey is quite fast. Turkey is the 9th most-affected country from the virus according to the data collected on the 11th May. Thus, the examination of the pandemic data of Turkey is a very important issue to understand the shape of the spread of the virus and its effects. In this study, we have a close look at the data of Turkey in terms of variables commonly used during the pandemic to set an example for possible future pandemics. Furthermore, the predictions are obtained for the coming fourteen days period by using exponential smoothing and Box-Jenkins models in time series analysis. In terms of some variables, the efficiency measurement issue is considered by using Stochastic Frontier Analysis (SFA), Data Envelopment Analysis (DEA) and Cross efficiency evaluation (CE) methods. In this way, the efficiency values of each day in the pandemic are tried to be calculated. Efficiency results are obtained by using the real data and the forecast-

Abbreviations: AICA, Akaike Information Criterion; ARIMA, Autoregressive Integrated Moving Average; CE, Cross Efficiency Evaluation; DEA, Data Envelopment Analysis; DNOC, Daily Number of Cases; DNOD, Daily Number of Deaths; DNOI, Daily Number of Recovered Cases; DNOT, Daily Number of Tests; DROC, Daily Rate of Cases; MAE, Mean Absolute Error; SC, Speed of Cases; SD, Speed of Deaths; SFA, Stochastic Frontier Analysis; RC, Rate of Cases; RD, Rate of Deaths; TNOC, Total Number of Cases; TNOD, Total Number of Deaths; TNOI, Total Number of The Cases In Intensive Care Unit; TNOI, Total Number of Intubated Cases; TNOR, Total Number of Recovered Cases; TNORC, Total Number of Remaining Cases.

* Corresponding author.

E-mail addresses: hknaci@erciyes.edu.tr (H. Knaci), mgunsal@gazi.edu.tr (M.G. Ünsal), rkasap@gazi.edu.tr (R. Kasap).

https://doi.org/10.1016/j.chaos.2020.110583
0960-0779/© 2020 Elsevier Ltd. All rights reserved.
The data set is collected from the website of the Republic of Turkey, Ministry of Health [2].

This study has two purposes. The first is to try to present a general view of COVID 19 in Turkey and understand the shape of the spread of the virus and its effects within the framework of the statistical data recorded officially during the pandemic. The second is to show how the selected data for time series modeling and efficiency analysis can be used to serve the purpose of Sections 3 and 4. By considering these two goals of the study, it is aimed to provide inferences that can make different interpretations by using different techniques. Furthermore, the aim is to be able to explain the pandemic process to researchers from different perspectives. The rest of the paper is organized as follows: Section 2 includes some comments about the data used in this study, descriptive statistics and graphs are also given in the section. Section 3 deals with the modeling processes in the concept of time series analysis. Section 4 touches on the efficiency analysis by using efficiency measurement methods. Finally, Section 5 concludes the study.

2. General overview of data

In this section, the data of 2019 novel coronavirus (COVID 19) infections are investigated to understand the shape of the spread of the virus and its effects in Turkey. The daily data is collected from the day of the first detected case (3rd March) to the 10th May. The data overview process tries to summarize and visualize the updated situation as much as possible by using the graphs and giving descriptive statistics about the variables used. The data used in this study can be found on the website of the Republic of Turkey, Ministry of Health [2].

The variables examined in this section are daily number of tests, the daily number of cases, the daily number of recovered cases, the daily number of deaths, the daily rate of cases, rate of deaths, speed of spread, speed of deaths, the total number of cases, the total number of deaths, the total number of the cases in the intensive care unit, the total number of intubated cases, total number of recovered cases and total number of remaining cases. In the next sections, some of the important variables used in the paper are also examined in details. Since lots of variables are discussed in the overview and analysis processes of the paper, abbreviations are used for them. The list of abbreviations can be found at the end of the article.

Firstly, the descriptive statistics are given for each variable to summarize the features and structures of the variables used in the paper. Some of the essential descriptive statistics are given in Table 1. Furthermore, Figs. 1-5 are given to visualize the data by using the graphs in terms of total and daily statistics, rates and speeds.

It is needed to give some details about the remarkable points of the values in Table 1 and Figs. 1-5. As seen in Table 1 and Fig. 1, the day when most cases appeared as 513 is the 32nd day (10th April) of the pandemic in Turkey. It can be defined as a peak point of the pandemic. This period is really the most violent process in Turkey during the pandemic. While the data is in this way according to the detection process of the disease, different implications can be done about the death statistics. Depending on the progress and aggravation of the disease in the body, the highest value of the number of deaths appeared as 127 in the 40th day (18th April) of the pandemic, which is eight days after 10th April. In Fig. 2, all the indicators related to the total data such as the total number of cases, the total number of recovered cases, the total number of deaths, etc. tend to increase, naturally.

As seen in Fig. 3, the most variation time of the daily rate of cases is in the first days of the pandemic, then it starts to lie between the interval of 0.7 and 1.3. The high values of the data of the rate of cases are realized within the days between 16th and 37th. As an interesting point, the rate of death reaches high val-

![Line Graphs for Daily Data](image.png)

Fig. 1. Line Graphs for Daily Data of the COVID 19 variables.
Fig. 2. Line Graphs for Total Data of the COVID 19 variables.

Fig. 3. Line Graphs for Rates of the COVID 19 variables.
uses in two different periods during the pandemic. Both between the 11th and 15th days and after the 25th day, the rate of death exhibits high values. In Fig. 4, speed values of the number of cases and deaths are given, respectively. The speed value is the ratio of the value of relevant day to the value of the previous day. In other words, it is a kind of rate showing the speed or frequency of an event that occurs per unit of time, population, or other standards of comparison. Depending on the scarcity of the initial data in the first days, the highest speed rates of detected cases and deaths are observed in the first days of the pandemic as 5th and 10th days, respectively. Actually, this is an indicator about the disease and this situation also shows that the pandemic tends to spread rapidly from the early days. In Fig. 5, the data about death statistics can be seen clearly. The total number of deaths appears to increase polynomially. As mentioned above, the highest value of the number of deaths is 127 and it appeared on the 40th day (18th April) of the pandemic. If we have a close look at daily number of deaths, it is likely to see that it increases until the 40th day of the pandemic and tends to decrease afterwards. At the beginning, the initial variance changes are valid. Then, the speed of deaths has values between 1 and 1.5, especially after the 11th day of the pandemic.

Depending on the process of the disease and the effect of the treatment process, the first recovered cases are detected on the 9th day of the pandemic. The maximum value of the daily number of recovered patients (43498 cases) occurs on the 50th day (28th April) of the pandemic. Especially between the 50th and 53rd days, the number of recovered cases is quite high. If we consider the structures of the data in terms of the values of descriptive statistics in Table 1, it is likely to say that daily rate of cases, speed of cases and speed of deaths show significant positive (right) skewed and pointed distribution structure in terms of statistics and standard error (std. error) values of skewness and kurtosis, respectively.

3. Time series modeling process

Time series analysis is a subject of Statistical Science, which includes techniques for modeling variable values depending on time and making predictions for the future. Unlike regression analysis, time series analysis may contain lagged values of variables in the model. Similar to regression models, linear and non-linear forms of time series models are available. In this section, basic time series models and forecasting methods are explained. Additionally, modeling and forecasting applications are discussed by using COVID-19 data of some variables mentioned in this study.

3.1. Time series analysis

Time series analysis is one of the important tools of applied statistics used for modeling the time series data and forecasting the future values. The modeling process includes expressing data graphically, using autocorrelation and partial autocorrelation functions (ACF and PACF) and modeling them is a task in itself. As a general definition of observed values of time series, \( Z_t, t = 1, 2, \ldots, n \) with \( n \) sample sizes and, \( \varepsilon_t \) observed data over time is expressed by \( Z_t \) [3].

The Box-Jenkins modeling ARIMA(\( p, d, q \)) processes include model selection, parameter estimation and model diagnostic checking. The model is considered according to historical observation values (\( p \)) and past error values (\( q \)). The degree of the parameter \( p \) and \( q \) are the degrees of the autoregressive and the moving average parameters, respectively. The degree of difference is considered as \( d \). This model is shown as \( \text{ARIMA}(p, d, q) \) and form formulated as \( \phi_p(B)(1 - B)^d Z_t = \theta_q(B) + \varepsilon_t \). Here, \( \phi(B) \)
and $\theta(B)$ are polynomials of $p$ and $q$ degrees, respectively $[3,4]$. 

$$\phi(B) = 1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p$$

$$\theta(B) = 1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q$$

(1)

The best model is determined by considering the value of the AIC (Akaike Information Criterion), the AIC is given as below $[3]$:

$$AIC(M) = nln\hat{\sigma}_M^2 + 2M$$

(2)

Here, $M$ is the number of estimated parameters. The Box – Jenkins approach for forecasting is Minimum mean square error forecasts for of $Z_{n+1}$ is given as the conditional expectation $[3,5]$. 

$$\hat{Z}_n(t) = E(Z_{n+1} | Z_n, Z_{n-1}, \ldots)$$

(3)

In this section, classical moving averages and exponential smoothing methods for forecasting is considered, as well. Exponential smoothing technique involves averaging (smoothing) the past values of a series in an exponentially manner $[5]$. 

Simple exponential smoothing has a single level parameter, it can be given as below:

$$L(t) = \alpha Y(t) + (1 - \alpha)L(t)$$

$$\hat{Y} (k) = L(t)$$

(4)

Brown's exponential smoothing has level and trend parameters, it can be defined in the equations below:

$$L(t) = \alpha Y(t) + (1 - \alpha)L(t - 1)$$

$$T(t) = \alpha (L(t) - L(t - 1)) + (1 - \alpha) T(t - 1)$$

$$\hat{Y} (k) = L(t) + (k - 1) + \alpha T(t)$$

(5)

Lastly, Holt’s exponential smoothing has level and trend parameters, it can be considered by the following equations:

$$L(t) = \alpha Y(t) + (1 - \alpha)(L(t - 1) + T(t - 1))$$

$$T(t) = \gamma (L(t) - L(t - 1)) + (1 - \gamma) T(t - 1)$$

$$\hat{Y} (k) = L(t) + kT(t)$$

(6)

Here, the best forecasting model is determined by using the Mean Absolute Error (MAE) value. In other words, if the relevant model has the smallest MAE, it can be considered as the best one $[5]$. 

$$MAE = \frac{1}{F} \sum_{l=1}^{F} |e_l|$$

(7)

3.2. Results of time series analysis

In this Section, DNOC, TNOC, DNOD, TNOD, DNOR, TNOR and TNORC as COVID 19 data for Turkey in terms of time series modeling and forecasting are obtained for the next 14 days (until 24th May) of the pandemic. Table 2 presents the models obtained for each series.

The model statistics for the models obtained for each series are given in Table 3. Accordingly, it can be said that all models are statistically sufficient and the best model at 0.01 significance level. Table 4 gives the parameter estimates of the models obtained from the series as exponential smoothing models. Accordingly, all parameter estimates except Gamma (trend) parameter of Model 3 are statistically significant at the level of 0.01 significance. Table 5 gives the parameter estimates of the models obtained from the series as exponential smoothing models.
the series as ARIMA models. Accordingly, all parameter estimates are statistically significant at 0.05 significance level. In Table 6, 14-day prediction values obtained by using forecasting techniques for each COVID 19 data of Turkey are given based on the models given in Table 4 and 5. The column of the day no is the number of days since the virus was first detected in Turkey. According to the forecasting results, the daily number of deaths and cases are expected to decrease in the short term. Although there is a decrease in daily data, normally, total data are expected to increase for number of deaths and cases. So, this case is an indicator which shows that the intensity of the pandemic decreases in the short term. The total number of recovered cases and intubated cases tend to decrease in the short term, which is a pleasant situation for the pandemic.

Using time series analysis to forecast the pandemic values of Turkey is not common subject in the literature. Furthermore, these forecasting values are used (obtained from both linear and non-linear forms) in the efficiency measurement process in Section 4, which is a novel application. By using these forecasting values and calculating the efficiency scores (with a non-linear model, SFA in Section 4.1) the days are considered as decision-making units in the study, and this approach has never been used in the literature for Turkish data.

4. Efficiency evaluation process

The two most popular methods used in the solution of benchmarking and ranking problems are Data Envelopment Analysis
(DEA) and Stochastic Frontier Analysis (SFA). Both methods draw a production frontier (efficiency frontier) for the production possibility set. While DEA draws the frontier with mathematical modeling sets, SFA draws the limit with mathematical functions. DEA is a nonparametric method and SFA is a parametric method, so the analysis performed is expected to be more appropriate since it contains a random error term in the function that SFA uses. Decision-making units (DMUs) on the efficiency frontier achieved with DEA are called efficient. In SFA, there are no DMUs on the efficiency frontier, its efficiency is ranked according to its proximity to this frontier.

4.1. Stochastic frontier analysis

SFA models contain an error term. There are some assumptions for this distribution of error term in SFA. The most important of them is that we know the character of this error term before starting the analysis. In other words, this error term has a distribution and it is assumed that the distribution is known in advance. In addition, SFA needs some assumptions about the structure of the data generated process and production possibility set before the implementation. That is, it assumes that we know the stochastic relationship between the inputs and outputs produced. Since this will cause a cost in the application, it can be said that this situation constitutes a disadvantage according to DEA.

The model proposed by Aigner et al. [6], Meesun and Van den Broeck [7] almost at the same time is as follows.

\[ \ln(y_i) = x_i \beta - u_i + v_i, \quad i = 1, 2, ..., N \quad (8) \]

where \( v \) is the error term corresponding to statistical noise and contains the margin of error caused by all kinds of uncontrollable situations. \( v \) and \( u \) in the Eq. (8) are independent variables. \( u \) is a non-negative random variable, that is, it is the factor that makes the situation stochastic and is called inefficiency effect. In other words, it is a variable that reflects the inefficient of the unit of interest. If \( u = 0 \), the firm is called efficient, and if \( u > 0 \) it is concluded that firm is to a certain extent inefficient. Meesun and Van den Broeck [7] considered \( u \) as exponential distribution, while Battese and Corra [8] considered half-normal distribution. In addition to Eq. (8), the model with constants is given in Eq. (9).

\[ \ln(y_i) = \beta_0 + x_i \beta - u_i + v_i, \quad i = 1, 2, ..., N \quad (9) \]

Eq. (9) is called the stochastic frontier production function. With this function, the output values are enveloped from the top by the stochastic variable \( e^{\hat{\beta}^T x_i + v_i} \). By using the above definitions, the technical efficient score of decision-making units is obtained

| Variable (Model no) | Model Type | Estimate | SE | t | Sig. |
|---------------------|------------|----------|----|----|------|
| DNOE (Model 2)      | Alpha (Level and Trend) | 0.942    | 0.064 | 14.625 | 0.000 |
| DNOE (Model 6)      | Alpha (Level)      | 1.000    | 0.140 | 7.162 | 0.000 |
| DNOE (Model 7)      | Gamma (Trend)     | 0.199    | 0.106 | 1.881 | 0.065 |
| TNOR (Model 5)      | Alpha (Level)     | 1.000    | 0.149 | 6.701 | 0.000 |
| TNOR (Model 6)      | Alpha (Level and Trend) | 1.000    | 0.073 | 13.404 | 0.000 |
| TNORC (Model 7)     | Alpha (Level and Trend) | 1.000    | 0.065 | 15.497 | 0.000 |

| Variable (Model no) | Model Type | Estimate | SE | t | Sig. |
|---------------------|------------|----------|----|----|------|
| DNOE (Model 1) Natural Logarithm | 1 | 25.897 | 5.300 | 4.886 | 0.000 |
| DNOE (Model 4) Natural Logarithm | 1 | 18.264 | 4.696 | 3.889 | 0.000 |
| TNOD (Model 4) Natural Logarithm | 2 | −0.086 | 0.042 | −2.057 | 0.044 |
| TNOD (Model 4) Natural Logarithm | 2 | −0.378 | 0.067 | −5.620 | 0.000 |

| Date      | Day no | DNOE | DNOE | DNOE | TNOD | TNOD | TNOR | TNOR |
|-----------|--------|------|------|------|------|------|------|------|
| 5/11/2020 | Day 63 | 1066 | 54   | 3089 | 140,935.07 | 3895 | 98,869 | 42,000.04 |
| 5/12/2020 | Day 64 | 1009 | 53   | 3089 | 142,097.69 | 3948 | 101,958.01 | 40,025.08 |
| 5/13/2020 | Day 65 | 935  | 52   | 3089 | 143,260.3 | 4000 | 105,047.01 | 38,050.12 |
| 5/14/2020 | Day 66 | 765  | 51   | 3089 | 144,422.92 | 4051 | 108,136.01 | 36,075.15 |
| 5/15/2020 | Day 67 | 760  | 50   | 3089 | 145,585.53 | 4010 | 111,225.01 | 34,100.19 |
| 5/16/2020 | Day 68 | 723  | 49   | 3089 | 146,748.15 | 4150 | 114,314.02 | 32,125.23 |
| 5/17/2020 | Day 69 | 645  | 48   | 3089 | 147,910.76 | 4198 | 117,403.02 | 30,150.27 |
| 5/18/2020 | Day 70 | 608  | 47   | 3089 | 149,073.38 | 4245 | 120,492.02 | 28,175.31 |
| 5/19/2020 | Day 71 | 623  | 46   | 3089 | 150,235.99 | 4291 | 123,581.03 | 26,200.35 |
| 5/20/2020 | Day 72 | 598  | 45   | 3089 | 151,398.6 | 4336 | 126,670.03 | 24,225.39 |
| 5/21/2020 | Day 73 | 584  | 44   | 3089 | 152,561.22 | 4380 | 129,759.03 | 22,250.43 |
| 5/22/2020 | Day 74 | 617  | 43   | 3089 | 153,723.83 | 4423 | 132,848.03 | 20,275.46 |
| 5/23/2020 | Day 75 | 655  | 42   | 3089 | 154,886.45 | 4465 | 135,937.04 | 18,300.5 |
| 5/24/2020 | Day 76 | 684  | 41   | 3089 | 156,049.06 | 4506 | 139,026.04 | 16,325.54 |
by Eq. (10).

\[
TE_i = \frac{y_i}{e x_i^\beta} = e^{x_i^\beta - u_i} = e^{-u_i}, \quad 0 < TE_i < 1
\]  

(10)

Another name for this efficiency score is the Farrell measure. This is an output-oriented measure. After the parameter estimate is made, the technical efficiency is estimated by Eq. (11).

\[
\hat{TE}_i = \exp\left[-E(u_i|e_i)\right]
\]  

(11)

Where \(e\) is defined as \(v - u\).

4.2. Data envelopment analysis and cross efficiency evaluation

Data Envelopment Analysis (DEA) is a nonparametric performance measurement technique commonly used in various branches of science to measure the efficiencies of DMUs and classify them as efficient or inefficient ones. In the efficiency measurement process, some variables are considered as inputs and some of the variables are considered as outputs. One of the classical DEA models called CCR can be given by the following mathematical programming model:

\[
\begin{align*}
\text{max} \theta_p &= \sum_{r=1}^{s} u_r y_{rp} \\
\text{s. t.} & \sum_{r=1}^{m} v_{r} x_{rp} = 1 \\
& \sum_{r=1}^{m} u_r y_{rp} - \sum_{r=1}^{m} v_{r} x_{rj} \leq 0, \quad j = 1, 2, ..., n \\
& u_r \geq 0, \quad r = 1, ..., s \\
& v_{r} \geq 0, \quad r = 1, ..., m
\end{align*}
\]

(12)

\(n\): the number of DMUs

\(m\): the number of inputs

\(s\): the number of outputs

\(x_{ij}(i = 1, ..., m)\): input values of DMU \(j\) \((j = 1, ..., n)\)

\(y_{ij}(r = 1, ..., s)\): output values of DMU \(j\) \((j = 1, ..., n)\)

\(u_r\): weights assigned to \(r\)th input

\(v_{r}\): weights assigned to \(r\)th output

\(\theta_p\): efficiency score of the decision making unit under evaluation (DMU\(_p\)) [9].

In CCR model, if the optimum value of \(\theta_p\) equals to 1 (\(\theta_p^\ast = 1\), DMU\(_p\) is defined as efficient unit, otherwise, it is referred to as an inefficient unit. To solve some of the drawbacks such as discrimination power problem and assignment of zero weights of classical DEA models, cross efficiency evaluation method (CE) is developed by Sexton et al. [10]. CE provides an alternative and effective efficiency measurement and ranking. Its idea can be defined by considering two stages. Firstly, the optimal weight values are calculated for each DMU by using above-mentioned CCR model. Secondly, the cross-efficiency score of DMU\(_p\) (\(\theta_{p,j}\)) is calculated as below [10]:

\[
\theta_{p,j} = \frac{\sum_{r=1}^{s} u_r y_{rp}}{\sum_{i=1}^{m} v_{i} x_{ij}}
\]

(13)

4.3. Results of efficiency analysis

The first COVID 19 cases have been detected in Turkey on 10th March, 2020. The first death from COVID 19 occurred on 16th March, 2020, six days after the first case was reported. The first to recover (42 cases) happened on 26th March, 2020, fourteen days after the first case appeared. Since then, the number of new cases, the number of deaths, and the number of recovered cases has continued to increase. In this study, efficiency analysis is performed for two different situations (by excluding the forecasting values and by including the forecasting values) with three different methods (SFA, DEA, and CE). Firstly, real data between 26th March and 10th May are used. Secondly, data from 26th March to 24th May are used. However, the data between 11th May and 24th May is the forecasting data. Thus, taking the values obtained by times series modeling into account, COVID 19 performance measurement is considered for the next days.

The variables used in the efficiency analysis are daily data for number of the recovered cases (output), number of deaths (input 1) and number of cases (input 2). The efficiency scores obtained for the first situation are given in Table 7 by three different methods. The graph of these scores is shown in Fig. 6. In the results of SFA, it is likely to see that the highest performance occurred between 31st

| Day | SFA | DEA | CE |
|-----|-----|-----|----|
| Day 17 | 0.5203 (44) | 0.0225 (39) | 0.0225 (39) |
| Day 18 | 0.7782 (40) | 0.0162 (42) | 0.0162 (42) |
| Day 19 | 0.8048 (39) | 0.0147 (44) | 0.0147 (44) |
| Day 20 | 0.8866 (28) | 0.0165 (41) | 0.0165 (41) |
| Day 21 | 0.7202 (43) | 0.0178 (40) | 0.0178 (40) |
| Day 22 | 0.8996 (27) | 0.0160 (43) | 0.0160 (43) |
| Day 23 | 0.9066 (26) | 0.0191 (45) | 0.0191 (45) |
| Day 24 | 0.8644 (30) | 0.0109 (46) | 0.0109 (46) |
| Day 25 | 0.8641 (31) | 0.0432 (32) | 0.0432 (32) |
| Day 26 | 0.8338 (35) | 0.0404 (34) | 0.0404 (34) |
| Day 27 | 0.8418 (34) | 0.0436 (33) | 0.0436 (33) |
| Day 28 | 0.1306 (46) | 0.0383 (35) | 0.0383 (35) |
| Day 29 | 0.4675 (35) | 0.0327 (38) | 0.0327 (38) |
| Day 30 | 0.7659 (42) | 0.0332 (36) | 0.0332 (36) |
| Day 31 | 0.9999 (1) | 0.0301 (38) | 0.0301 (38) |
| Day 32 | 0.9999 (1) | 0.0658 (29) | 0.0658 (29) |
| Day 33 | 0.9999 (1) | 0.0572 (31) | 0.0572 (31) |
| Day 34 | 0.7747 (41) | 0.0601 (30) | 0.0601 (30) |
| Day 35 | 0.8498 (33) | 0.0907 (27) | 0.0907 (27) |
| Day 36 | 0.8519 (32) | 0.0877 (28) | 0.0877 (28) |
| Day 37 | 0.8131 (38) | 0.1305 (26) | 0.1305 (26) |
| Day 38 | 0.8861 (29) | 0.1411 (23) | 0.1411 (23) |
| Day 39 | 0.9141 (24) | 0.1718 (20) | 0.1718 (20) |

| Mean | 0.8625 | 0.3375 | 0.3180 |
| Std. Deviation | 0.1505 | 0.3533 | 0.3340 |
and 33rd days. According to the results of DEA, the best performing days are 55th and 56th days, and CE consider this interval between 54th and 58th days. Briefly, it can be seen that the general situation is troubled in the first days of the epidemic, but then improved in the following days.

The values given in parentheses in Table 7 are the rank of efficiency scores. Rank (1) expresses the day of the best performance. According to the results, the efficiency scores are not good in the first days, but then the scores gradually increase. This is more evident in Fig. 6. As the days progressed, an increase in efficiency scores is detected. This is an indicator of the measures taken by the scientific committee of Ministry of Health and the increasing awareness of people.

The efficiency scores obtained for the second situation are given in Table 8 and the graph of these scores is shown in Fig. 7, where the days corresponding to the forecasting days are marked with ∗.

Similar results are obtained in Fig. 7, as well. Based on the three methods, it is likely to see that the efficiency scores gradually increase. The results of SFA show that the efficiency scores increase step by step and that the recovery process continues in the coming short term. Especially, after 3rd day, SFA scores tend to be high. Both DEA and CE show similar results. It is a noticeable point that 56th, 73rd, 74th and 76th days have the best performances according to DEA results. Briefly, it can be seen that the general situation is troubled in the first days of the epidemic, but then it improves in the following days. The reason for that can be considered
as the practices and the strict measures taken so far have worked. Turkey gradually increases its performance in the next days if it continues to follow this policy. Spearman’s rank correlations are also obtained to evaluate the consistency of the models used and are given in Table 9.

The correlation coefficients obtained are statistically significant ($\alpha = 0.05$). In the first situation, the correlation coefficients between SFA and DEA, and between SFA and CE are 0.6010 and 0.5910, respectively. Consequently, in both situations, SFA has higher correlations compared to DEA and CE. This section has a comparison between the efficiency measurement methods, and the scores and ranking results of each efficiency measurement method (for both linear approaches as DEA and CE and non-linear approach as SFA) and correlations between them are given in Tables 7-9.

5. Discussion and conclusion

Due to the number of infected cases, Turkey is one of the mostly influenced countries from COVID 19 in the world. Turkey is the 9th most-affected country from the virus according to data collected on the 11th May. Thus, the analysis of the data belonging to Turkey is important to understand the pandemic process. The results and discussions can be a beneficial indicator and a useful guide for the researchers who would like to work on the possibilities of future pandemics.

One of the purposes of this study is to try to present a general view of COVID 19 in Turkey and understand the shape of the spread of the virus and its effects within the framework of the statistical data recorded officially during the pandemic. The day when most cases appeared as 5138 is the 32nd day (10th April) of the pandemic in Turkey. It can be defined as a peak point of the pandemic. This period is really the most violent process in Turkey during the pandemic. The highest value of the number of deaths appeared as 127 in the 40th day (18th April) of the pandemic, which is eight days after 10th April. The maximum value of the daily number of recovered patients (43498 cases) occurred on the 50th day (28th April) of the pandemic. Especially between the 50th and 53rd days, the number of recovered cases are quite high. All these results indicate that the negative process is getting slowed down day by day, and after the day in the middle of the pandemic, the situation has started to improve. If we consider the structures of the data in terms of the values of descriptive statistics, it can be investigated that daily rate of cases, speed of cases and speed of deaths show significant positive (right) skewed and pointed distribution structure in terms of statistics and standard error (std. error) values of skewness and kurtosis, respectively.

The second purpose of the study is to show how the selected data for time series modeling and efficiency analysis can be used for serving the results for the policy makers. Forecasting results of time series analysis indicate that the daily number of deaths and cases are expected to decrease in the short term. Total data are expected to increase slowly for number of deaths and cases. So, this case is an indicator which shows that the intensity of the pandemic decreases in the short term. The total number of recovered cases and intubated cases tend to decrease in the short term, which is a pleasant situation for the pandemic. According to the results of efficiency analysis, the pandemic situation is not good in the first days of the pandemic, but then the scores gradually increased. As the days progressed, an increase in efficiency scores is detected. According to the results in the second scenario (by including the forecasting values) of the efficiency measures, it is expected to see this normalization and recovery process in coming short term, as well.
Declaration of Competing Interest

There is no conflict of interest and we confirm that the manuscript has been read and approved for submission. I hope you may find the manuscript suitable for publication and look forward to hearing from you in due course.

References

[1] website: https://www.medicalnewstoday.com/articles/COVID-19#incubation-period.
[2] website: https://COVID19.saglik.gov.tr/
[3] Box GEP. Time series analysis: forecasting and control. Englewood Cliffs: NJ Prentice Hall; 1994.

[4] Ünsal MG, Kasap R. Cases of residual types in diagnostic checking for ARMA model. Hacettepe J Math Stat 2014;43(3):1–10.
[5] Montgomery DC. Introduction to time series analysis and forecasting. New Jersey: Wiley; 2015.
[6] Aigner DJ, Lovell CAK, Schmidt P. Formulation and estimation of stochastic frontier production function models. J Econom 1977;6(1):21–37.
[7] Meeusen W, Van den Broeck J. Efficiency estimation from cobb-douglas production functions with composed error. Int Econ Rev 1977;18(2):435–44.
[8] Battese GE, Corra GS. Estimation of a production frontier model: with application to Pastoral Zone of Eastern Australia. Aust J Agric Econ 1977;21(3):169–79.
[9] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. Eur J Oper Res 1978;2:429–44.
[10] Sexton TR, Silkman RH, Hogan AJ. Data envelopment analysis: critique and extension. In: Silkman RH, editor. Measuring efficiency: an assessment of data envelopment analysis, 32. San Francisco: Jossey-Bass; 1986. p. 73–105.