**A Modeling Study on The Estimation of COVID-19 Daily and Weekly Cases and Reproduction Number Using The Adaptive Kalman Filter: The example of Ziraat Bank, Turkey**

1Dr. İlker MET, Technology Management, KHO, Head of Human Resources, Ziraat Bank, Ankara, Turkey, imet@ziraatbank.com.tr

Dr. Levent ÖZBEK, Department of Statistics, Ankara University Faculty of Science, Ankara, Turkey, Levent.Ozbek@science.ankara.edu.tr

Himmet AKSOY, Business Administration, Selçuk University, Human Resources, Ziraat Bank, Ankara, Turkey, hiaksoy@ziraatbank.com.tr

Ayfer ERKOÇ, Statistics, Ankara University, Career Management Officer, Ziraat Bank, Ankara, Turkey, ayfcelik@ziraatbank.com.tr

**1Corresponding Author:** Dr. İlker Met,

 Ziraat Bank, Anafartalar Street, No:8, 06080, Ankara/ Turkey, +903125842000

# Abstract

**Objective:** Since the beginning of 2020, the world has been struggling with a viral epidemic (COVID-19), which poses a serious threat to the collective health of the human race. In order to overcome and combat this virus, coming up with effective policies faster than the reproduction rate of the disease is of the utmost importance.  Mathematical modeling of epidemics is critical for developing such policies, especially during these uncertain times. As with previous epidemics, we, like the scientists before us, are attempting to model disease case data to support policy makers in this period.

**Methods:** In this study, the reproduction number and model parameters were predicted using AR(1) (autoregressive time-series model of order 1) and the adaptive Kalman filter (AKF). The data sample used in the study consists of the weekly and daily number of cases amongst the Ziraat Bank personnel between March 11, 2020, and April 19, 2021. The reason we selected bank data as a sample was because they operate in all provinces of Turkey and the trend movement of their cases mirrored that of the country. This sample was modeled in the state space, and the AKF was used to estimate the number of cases per day.

**Results:** Using the AR(1) stochastic process for estimation is a common approach as it does not require any other modeling assumptions. Due to the simplicity of the AR models, it is a highly recommended approach for reproduction number estimation. It is quite simple to model the daily and weekly case number time series with the time-varying parameter AR(1) stochastic process and to estimate the time-varying parameter with online AKF. Overall, we found that the weekly case number prediction was more accurate than the daily case number(R2 = 0.97), especially in regions with a low number of cases.

**Conclusion**: We suggest that the simplest method for reproduction number estimation can be obtained by modeling the daily cases using an AR(1) model.

**Key Words:** COVID-19, Modeling, Reproduction number estimation, AR(1), Kalman filter

**Introduction**

The ability to make future predictions helps us to navigate uncertain times. Mathematical, deterministic, and statistical modeling methods are invaluable tools at the disposal of policymakers to do just that, which is especially salient during these trying times. Just as it was with previous viral outbreaks, scientists have conducted numerous studies modeling the rate of disease transmission and the dates when we will reach maximum capacity. These studies are not specific to the COVID-19 outbreak, and these epidemic models are scientifically important for policy development (1).

Trend and peak estimations are tough obstacles due to the changes and restrictions in the data disclosure policies of different countries. The deviation and margin of error are seen to be high in studies with peak estimation (2). Owing to these constraints, predictive models generally rely on the reproduction numbers to determine model trends. As other researchers have noted, the uncertainty of available official data, especially regarding the actual number of infected cases, may lead to uncertain results and false estimates (3).

A total of 244 oft-cited articles since January 2020 showed that 46% of studies used compartmental models, 32% used statistical models, and 1% used individual-based models. A majority of the studies were conducted in Asian (78.93%) and European (59.09%) countries. A majority of them used compartmental models (SIR and SEIR) (46.1%) and statistical models (growth models and time series) (31.8%). The remaining studies employed artificial intelligence (6.7%), the Bayesian approach (4.7%), network models (2.3%), and agent-based models (1.3%) (4).

The biggest contribution of modeling studies is their ability to show the number of reproductions and the progress of the epidemic trend. Highlighting the true impact of the epidemic at an early stage is an important step towards effective planning on restrictions, hospital SOPs, vaccination policy, etc. As the COVID-19 pandemic continues, mathematical epidemiologists have continued to share their models on how the disease has spread, the current state of play, and what still needs to be done. Since the beginning of COVID-19, nonparametric methods have been used in modeling studies more frequently, especially logarithmic and exponential models (5).

Along with the growth models, nonlinear methods were also used for parameter estimation. The spread of the epidemic in China was estimated using a logistic growth model, while non-linear least squares (NLS) were used to estimate model parameters. In this study, we found that the growth rate of Covid-19 is significantly different between China, South Korea, and Iran (6).

In Nigeria, estimations were made using logistic and exponential models, while model parameters were estimated using ordinary least squares (7).

The logistic growth model, generalized logistic growth model, generalized Richards model, and generalized growth model to the reported number of infected cases for the whole of China, 29 provinces in China, and 33 countries and regions that have been or are undergoing major outbreaks are used to estimate peak times (8).

Nine non-linear models (Brody, Bertalanffy, Logistic, Generalized Logistic, Richard, Negative Exponential, Stevens, Tanaka, Gompertz) for the US, Brazil, Germany, India, Russia, Italy, Spain, France, the United Kingdom, and Turkey were studied to model the estimation of reproduction number and daily number of cases by estimating parameters using a Kalman filter (9).

A support vector machine (SVM) with fuzzy granulation was used to predict the growth range of confirmed new cases, new deaths, and new cured cases in China. The experimental results showed that the Elman neural network and SVM can predict the development trend of cumulative confirmed cases, deaths, and cured cases, whereas LSTM (long short-term memory) is more suitable for predicting cumulative confirmed cases (10).

Using mathematical and statistical methods, we attempted to estimate the reproduction number and the time range between waves in South Korea. The findings of the model study support the effectiveness of control measures against COVID-19 in Korea (11). The purpose of managing and controlling COVID-19 in Iran mortality trends was modeled using regression, spatial modeling, risk mapping, and change detection using the random forest machine learning technique (12). The [maximum likelihood](https://www.sciencedirect.com/topics/medicine-and-dentistry/maximum-likelihood-method) (ML) value of reproductive number (R0) can be estimated by the Poisson distribution determined by daily infectiousness (13).

It is seen in the literature that the types of kalman filter are also used in the prediction of the transmission of epidemic diseases. An adaptive unscented Kalman filter (AUKF) -based optimal controller has been designed to control unknown tuberculosis dynamics in individuals treated with active tuberculosis, at home or in hospital. In this way, even in the presence of a small group of infectious people, the long-term persistence of the disease is thought to be prevented (14).Our study propose a method different from the models of epidemic diseases so far.

**Material and Methods:**

This study consists of four steps:

1. Data collection
2. Trend Analyses with Moving Average Graphs Modeling
3. Parameter Estimation with time series (AR) models

Parameter estimation with adaptive Kalman filters (AKF).**The Flowchart of Methodologies:**



Figure 1. The flowchart of methodologies modeling the spread of [coronavirus](https://www.sciencedirect.com/topics/immunology-and-microbiology/coronavirinae) in Ziraat Bank

**1.Data Collection Phase:** In this study, we used the data of Ziraat Bank as a sample because of their wide presence throughout Turkey. Another reason for using bank data is the policy change affecting data disclosure within the country. The data was collected through an application that was developed by the Bank's technology team, starting from the day the first case was observed on March 11, 2020 to April 19, 2021. Data collected in the timespan of these 325 days was analyzed. (timespan is >365 and 325 days modelled , because of no cases days are excluded.) The total number of cases used in study represents 21% of bank employees.

**2. Trend Analyses:**

|  |  |
| --- | --- |
|  |  |

Figure 2. COVID-19 cases trend graphs between 11 March 2020 – 19 April 2020 in Ziraat Bank Case

As shown in Figure 2, we can see the 3rd peak occurring in the moving average (MA) charts. In this study, MA was used as a checkpoint. Daily cases are more smoothing than weekly cases.

**3.Parameter Estimation with Time Series AR1 Models**

In this study, time series data (number of daily cases) was modeled. It is assumed that the number of daily cases $i\_{t}$ is in the form of AR(1) is given by Eq. 1.

 $i\_{t}=θi\_{t-1}+v\_{t}$ (1)

where θ is a constant and $v\_{t}$ is $v\_{t} $~ N(0,$σ\_{1}^{2}$) . The random variables $v\_{1}$,$ v\_{2 },…, v\_{n}$ are assumed to be uncorrelated. It is also assumed that the θ parameter in Eq. (1) is time-varying and is a stochastic process in the form of a random walk process. In this case, the θ random walk process can be expressed as in Eq. (2)

$θ\_{t}=θ\_{t-1}+w\_{t}$ (2)

$w\_{t}$ is normally distributed with N(0,$σ\_{1}^{2}$), and the random variables $w$,$ w\_{2 },…, w\_{n}$ are assumed to be uncorrelated. Considering Eq.(1) and (2), the following state-space model can be written:

$θ\_{t}=θ\_{t-1}+w\_{t}$ (2)

$i\_{t}=θi\_{t-1}+v\_{t}$ (3)

 (9).

The state variable is unobservable, the time-varying $θ\_{t}$ parameter, and can be estimated using the AKF.

Model codes are written in Matlab 2013a program.





Figure 3. Daily estimated cases and estimated-reproduction number.





Figure 4. Weekly cases and estimated-reproduction number.

Figure 3 shows the estimation of the number of cases per day and the estimation of the reproduction number with the Kalman filter. In Figure 4, the estimation of the weekly case numbers and the estimation of the reproduction number with the kalman filter are shown. The graphs are created using the Matlab. Because of daily cases variability, weekly cases are averaged and smoothed. Calculated MSE, MAPE and R2 shown in Table1. İf there is no variation in the daily cases, calculated values can be expected ​​close to the weekly cases calculated values. As can be seen from Table.1 calculated values different between daily and weekly cases. Estimations made by averaging weekly cases are suitable for using R(t).

**Table 1.** Calculated MSE, MAPE,R2

|  |  |  |  |
| --- | --- | --- | --- |
| Time | MSE | MAPE |  R2 |
| Daily | 84 | 154.19 | 0.69 |
| Weekly | 6 | 62.23 | 0.97 |

MSE: Mean squared error, MAPE: Mean absolute percentage error.

It is quite simple to model the daily and weekly case number time series with the time-varying parameter AR(1) stochastic process and estimate the time-varying parameter with online AKF..It can be seen weekly case number prediction was more precise than the daily case number (R2 = 0.97).

**4.Estimating the Reproduction Number with Adaptive Kalman Filtering**

The instantaneous reproduction number, Rt, at time t can be estimated using Eq 4. (15)

$R\_{t}=\frac{E(i\_{t})}{\sum\_{s=1}^{t}i\_{t-s}-w\_{s}}$ (4)

In equation (4), $i\_{t}$ stands for the number of new infections generated at time step t. $w\_{s}$ is the probability distribution of the infectivity profile, which is dependent on time elapsed since the infection of the case. In practice, w is approximated by the distribution of serial intervals. Let us express the value of $R\_{t}$ calculated using the AR(1) model with $R\_{t}^{AR}$. İf s=1 and $w\_{1}$=1 are given in (4). then Eq 4. can be written as

$R\_{t}^{AR}=\frac{E(i\_{t})}{i\_{t-1}}$ = $\frac{î\_{t}}{i\_{t-1}} $= $\frac{θ\_{t}i\_{t-1}}{i\_{t-1}}$ = $θ\_{t}$, t=1,2,3,…,n (5)

The Kalman filter is a popular estimation method used to solve the state estimation problem in dynamic systems. As long as the system characteristics are known correctly, Kalman filter works with the best prediction performance. However, in cases where the system characteristics are partially known or uncertain, it is inevitable that there will be serious losses in the prediction performance of the filter. In order to overcome the performance loss problem in the Kalman filter, the adaptive kalman filter method has been adapted. İn adaptation forgetting factor proposed by Özbek and Aliev is used (15).

The estimated $R\_{t}$ value using the AR(1) model is equal to the estimate of the time-varying parameter of the AR(1) model. The value of $R\_{t}^{AR}$calculated using Equation 5. are shown in Figure 3 and Figure 4. (16).

**Conclusion and Results**

Theta in the AR(1) model called it the 'multiplication' factor. In time series models, this is called a parameter. In classical time series methods, this parameter is assumed to be constant and estimated accordingly. Here we assumed the parameter as time varying and random walk stochastic process. When this assumption is made, classical time series methods are not used. We took this assumption and AR(1) model together and turned it into a state-space model and estimated the parameter based on time using AKF. This method does not require all the data, only the arrival of the last observation is sufficient to obtain an estimate of the parameter and is an on-line estimation method. This is one of the advantages of the method. When the classical AR(1) model is used, the stationarity condition must be met in order to make the predictions. In the time-varying parameter assumption we use, this condition does not need to be met. This is the second advantage of the method we use. Since the Adaptive Kalman Filter is a self-adaptive estimation method and the variance of the noise processes in the model is not known exactly, it makes better predictions than the normal Kalman filter. This is stated in the references.In addition to generally known methods, it is quite simple to model the daily case number time series with the time-varying parameter AR(1) stochastic process and estimate the time-varying parameter with online AKF. It does not require any assumptions and it produces good results in cumulative weekly data when the model results are examined. The model is so simple that it can be reproduced and applied without detailed knowledge of epidemiology or programming languages. We must stress again that the purpose of this study is merely to show that it is possible to model epidemic trends using simple methods.

Using AR(1), the stochastic process for estimation is a common approach as it does not require any other modeling assumptions. Due to the simplicity of AR models, we highly recommend this approach for reproduction number estimation. As seen in our study, it is simple enough to model the daily and weekly case number time series with the time-varying parameter AR(1) stochastic process and estimate the time-varying parameter with online AKF. We also noted that the weekly case number prediction was more precise than the daily case number prediction (R2 = 0.97).

**Appendix :State-Space Model and Adaptive Kalman Filter (AKF)**

The optimum linear filtering and estimations methods introduced by Kalman (1960) have been considered one of the greatest achievements in estimation theory.

Discrete-time linear state-space models and Kalman filtering (KF) have been employed since the 1960s, mostly in the control and signal processing areas. The KF has been extensively employed in many areas of estimation the extensions and applications of discrete-time linear state-space models can be found in almost all disciplines [17-25].

Let us consider a general discrete-time stochastic system represented by the state and measurement models given by

 (A1)

 (A2)

where  is an *n*×1 system vector,  is an *m*×1 observation vector,  is an *n×n* system matrix, is an *m×n* matrix,  an *n*×1 vector of zero mean white noise sequence and  is an *m*×1 measurement error vector assumed to be a zero mean white sequence uncorrelated with the  sequence. The covariance matrices  and are defined by , . The filtering problem is the problem of determining the best estimate of its  condition, given its observations  [17-25]. When  observations are given, the prediction of state  with



and the covariance matrix of the error with



when  observations are given, the prediction of state  with

and the covariance matrix of the error are shown with

.

Let the initial state be assumed to have a normal distribution in the form of . The optimum update equations for KF are,

 (A3)

 (A4)

 (A5)

 (A6)

 (A7)

In the above equations,  is the a priori estimation and  is the a posteriori estimation of . Also,  and  are the covariance of a priori and a posteriori estimations respectively [17],[18]. In some cases, divergence problems may occur in the Kalman Filter due to the incorrect installation of the model. In order to eliminate divergence in the Kalman filter, adaptive methods are used [26], [27], [28]. One of these is the use of the forgetting factor. A forgetting factor is proposed by Ozbek and Aliev [26].

 (A8)

# References

1. Gog JR. How you can help with COVID-19 modelling. Nat Rev Phys (2020) 2(6):274–275 <https://doi.org/10.1038/s42254-020-0175-7>.
2. Tiwari A. Modelling and analysis of COVID-19 epidemic in India. *J Saf Sci Resil* (2020) **1**(2):135–140. ISSN 2666-4496. <https://doi.org/10.1016/j.jnlssr.2020.11.005>.
3. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PLOS ONE* (2020) **15**(3):e0230405. <https://doi.org/10.1371/journal.pone.0230405>.
4. Gnanvi JE, Salako KV, Kotanmi GB, Glèlè Kakaï RG. On the reliability of predictions on COVID-19 dynamics: A systematic and critical review of modelling techniques. *Infect Dis Modell* (2021) **6**:258–272. ISSN 2468-0427. <https://doi.org/10.1016/j.idm.2020.12.008>.
5. Sornette D, Mearns E, Schatz M. Interpreting, analysing and modelling COVID-19 mortality data. *Nonlinear Dyn* (2020) **101**:1751–1776. <https://doi.org/10.1007/s11071-020-05966>.
6. Shen CY. Logistic growth modelling of COVID-19 proliferation in China and its international implications. Int J Infect Dis (2020) 96:582–589. ISSN 1201-9712. <https://doi.org/10.1016/j.ijid.2020.04.085>.
7. Ayinde K, Lukman AF, Rauf RI, Alabi OO, Okon CE, Ayinde OE. Modeling Nigerian COVID-19 cases: A comparative analysis of models and estimators**.** *Chaos Solitons Fractals* (2020) **138**:article 2020. DOI: [10.1016/j.chaos.2020.109911](https://doi.org/10.1016/j.chaos.2020.109911).
8. Wu K, Darcet D, Wang Q, Sornette D. Generalized logistic growth modeling of the COVID-19 outbreak: Comparing the dynamics in the 29 provinces in China and in the rest of the world*. Nonlinear Dyn* (2020) **101**:1–21. <https://doi.org/10.1007/s11071-020-05862-6>.
9. Özbek L, Demirtaş H, Turkiye Klinikleri J Biostat. A Study on The Estimation of COVID-19 Daily Cases and Reproduction Number Using Adaptive Kalman Filter For USA, Brazil, Germany, India, Russia, Italy, Spain, United Kingdom, France, Turkey (2021) 13(1):3. doi: [10.5336/biostatic.2020-80186](https://doi.org/10.5336/biostatic.2020-80186).
10. Hao Y, Xu T, Hu H, Wang P, Bai Y. Prediction and analysis of Corona Virus Disease 2019. *PLOS ONE* (2020) **15**(10):e0239960. doi:[10.1371/journal.pone.0239960](https://doi.org/10.1371/journal.pone.0239960).
11. Shim E, Tariq A, Chowell G. Spatial variability in reproduction number and doubling time across two waves of the COVID-19 pandemic in South Korea, February to July 2020. *Int J Infect Dis* (2021) **102**:1–9, ISSN 1201-9712. <https://doi.org/10.1016/j.ijid.2020.10.007>.
12. Pourghasemi HR, Pouyan S, Heidari B, Farajzadeh Z, Fallah Shamsi SR, Babaei S, Khosravi R, Etemadi M, Ghanbarian G, Farhadi A et al. Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). *Int J Infect Dis* (2020) **98**:90–108, ISSN 1201-9712. <https://doi.org/10.1016/j.ijid.2020.06.058>.
13. Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. *Int J Infect Dis* (2020) **93**:201–204, ISSN 1201-9712. <https://doi.org/10.1016/j.ijid.2020.02.033>.
14. Cetin, M , Beyhan, S . (2020). Adaptive Kalman Filtering Based Optimal Control of Tuberculosis Dynamics with Exogenous Reinfections . Journal of Engineering Sciences and Design, 8 (4) , 1260-1268 . DOI: 10.21923/jesd.717130
15. Özbek L, Aliev FA. Comments on adaptive Fading Kalman Filter with an application. Available at:https://www.researchgate.net/publication/263595161\_Adaptive\_Fading\_Kalman\_Filter\_with\_an\_Application. Automatica (1998) 34(12):1663–1664. DOI: [10.1016/S0005-1098(98)80025-3](https://doi.org/10.1016/S0005-1098%2898%2980025-3).
16. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol (2013) 178(9):1505–1512. DOI: [10.1093/aje/kwt133](https://doi.org/10.1093/aje/kwt133), PMID: [24043437](http://www.ncbi.nlm.nih.gov/pubmed/24043437), PMCID: [PMC3816335](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3816335).
17. Jazwinski, AH. Stochastic Processes and Filtering Theory. Academic Press, 1970
18. Anderson, BDO, Moore JB. Optimal Filtering. Prentice Hall, 1979.
19. Chui, C.K, Chen G. Kalman Filtering with Real-time Applications. Springer Verlag, 1991.
20. Ljung, L, Söderström T. Theory and Practice of Recursive Identification. The MIT Press,1993.
21. Chen, G. Approximate Kalman Filtering. World Scientific, 1993.
22. Grewal S, Andrews AP. Kalman Filtering Theory and Practice. Prentice Hall, 1993.
23. Öztürk F, Özbek L. Mathematical Modelling and Simulation, Pigeon Yay, 2016 (in Turkish).
24. Özbek L. Kalman Filtresi, Akademisyen Yay. 2017 (in Turkish).
25. Kalman RE. A new Approach to linear Filtering and Prediction Problems”. Journal of Basic Engineering. (1960);82:35-45.
26. Özbek L, Aliev FA. Comments on Adaptive Fading Kalman Filter with an Application. Automatica (1998); 34(12): 1663-1664.
27. Efe M, Özbek L. Fading Kalman Filter for Manoeuvring Target Tracking. Journal of the Turkish Statistical Assocation (1999); 2(3):193-206.
28. Özbek L, Efe M. An Adaptive Extended Kalman Filter with Application to Compartment Models. Communications In Statistics-Simulation And Computation (2004); 33(1): 145-158.