Bol. Soc. Paran. Mat.
©SPM –ISSN-2175-1188 on line
SPM: www.spm.uem.br/bspm

(3s.) **v. 2024 (42)** : 1–8.
ISSN-0037-8712 in press
doi:10.5269/bspm.63521

Theta Models for Daily Pandemic Data

Maria Acim, Mehdi Zahid and Adil Ez-Zetouni

abstract: Forecasting techniques are critical for developing better strategies and making timely judgments in times of crisis. Recently, both epidemiologists and statisticians have become interested in anticipating the COVID-19 pandemic using time series forecasting methods. In this work, we will investigate the performance of Theta models for predicting short-term pandemic data. The major goal of this research is to determine which of the statistical Theta-methods is appropriate for predicting COVID-19 trends for a set of selected countries, namely: the United Kingdom, South Africa, Malaysia, Morocco, and Russia. The results suggest that the standard theta approach is more accurate for data from the United Kingdom, which has a lot of variability. For the other analyzed countries, the dynamic optimized theta model performs better in forecasting.

Key Words: Time series, forecasting techniques, Theta method, COVID-19, pandemic data.

Contents

1. Introduction

The current COVID-19 epidemic, which began in Wuhan, China in December 2019 and is caused by the novel SARS-CoV-2 virus, has posed a major threat to world public health. Several waves have arisen in different nations around the world since the beginning of virus transmission, despite preventative and immunization attempts.

Numerous statistical models have been developed to anticipate the pattern of epidemic evolution through time [22,23,7,8]. The authors in [9,10,11,12,14] realized several works that used COVID-19 case patterns to make decisions about pandemic crisis containment and management. Forecasting epidemic trends has served as an early warning system for public health officials, allowing them to respond by enacting relevant NPIs such as mask use, movement control orders, physical separation, and frequent hand washing to prevent disease outbreaks. Several recent studies have used econometric models to estimate the evolution of the COVID-19 pandemic (see for instance [12,13,16]). Obviously, time series approaches have played a crucial role in forecasting COVID-19 instances, as evidenced in ([13,16,15,17]). In this context, several researchers have recently employed the ARIMA time series method to forecast the COVID-19 pandemic evolution ([16,15,17]). During the third wave of COVID-19 in Malaysia, Tan et al. [17] devised a time series approach to forecast reported case trends using SARIMA Models. In many previous works, hybrid methodologies combining time series and artificial intelligence models have been employed to build systems for forecasting COVID-19 cases (see [18,19,20]).

In recent years, the Theta model has attracted the attention of many researchers and forecasting specialists due to its simplicity and accuracy ([2,3,4,5,6]). Forecasting using the Theta method has created

2010 *Mathematics Subject Classification:* 91B82, 91G70.
Submitted May, 11, 2022. Published August 19, 2022

a lot of interest in many fields, such as in economics and finance [26]. The method entails removing the evident trends in the observable phenomenon (growth, periodicity) and concentrating on what remains after they are removed. The Theta approach is applied to deseasonalize non-seasonal or deseasonalized time series using the multiplicative classical decomposition. It performs well for high-frequency data (daily and monthly data) [24].

The aim of this study is to analyze a family of existing Theta methods in the literature using various error measurement metrics for five separate sets of pandemic data from the United Kingdom, South Africa, Malaysia, Morocco, and Russia. Then, using data from 5 March, 2020 to 24 December, 2021, we try to forecast daily confirmed COVID-19 cases for the studied countries. The remainder of the paper is laid out as follows: The proposed Theta-models are detailed and formulated in Section 2. The data and their statistical features are described in Section 3. The results and their practical consequences are discussed in Section 4. Finally, in Section 5, we end the work with a conclusion.

2. Methods

In this section, we review some of the literature's preliminary definitions of Theta models. Then we present some error metrics in order to select the appropriate model for the used pandemic data. For time series forecasting the Theta model was presented for the first time by Assimakopoulos and Nikolopoulos [1] . The Theta method is based on the concept of changing the local slopes of the time series using Simple Exponential Smoothing approach [1]. Many authors' investigations have been developed in order to optimize the Theta model. Basically, the original time series is recreated from the individual lines by searching for the weights of Theta lines. The authors in [25] presented a new approach for choosing the optimal value for the θ-parameter when a single 'θ-line' is used, rather than a combination of two 'θ-lines' as in [1]. Then Fiorucci et al. [24] provided a generalization of the Theta model, namely the Dynamic Optimised Theta Model. In the following we will present the mathematical and methodological framework of the different main Theta models.

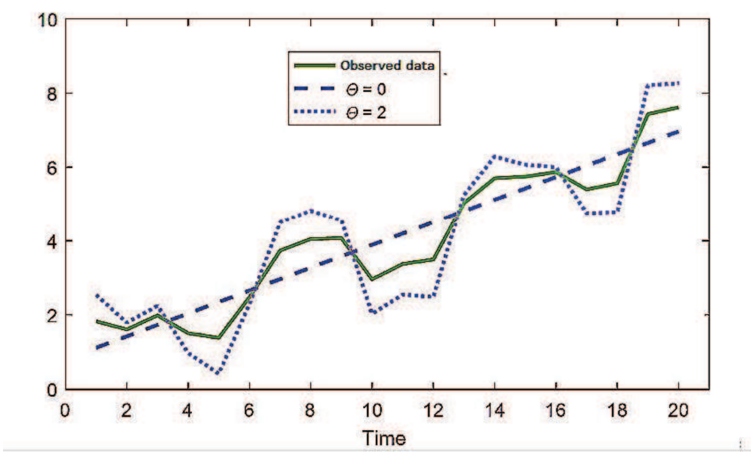

Figure 1: Observed data and θ-lines for $\theta = 0$ and $\theta = 2$ (see [2])

2.1. Standard Theta Model [1]

Assimakopoulos and Nikolopoulo [1] proposed the Theta line as the solution of the equation

$$D^2\zeta_t(\theta) = \theta D^2 Y_t,\ t = 1, \dots T, \tag{2.1}$$

where $Y_1, \dots, Y_T$ represent the original time series data and $DX_t = (X_t - X_{t-1})$. The initial values ζ_1 and ζ_2 are obtained by minimizing $\sum_{i=1}^{T}[Y_t - \zeta_t(\theta)]^2$. However, the analytical solution of (2.1) is given by

$$\zeta_t(\theta) = \theta Y_t + (1-\theta)(\mathcal{A}_T + \mathcal{B}_T t), t = 1, \dots T, \tag{2.2}$$

where $\mathcal{A}_T$ and $\mathcal{B}_T$ are the minimum square coefficients of a simple linear regression over $Y_1, \ldots, Y_T$ against $1, \ldots, T$ which are only dependent on the original data and given as follow

$$\mathcal{A}_T = \frac{1}{T}\sum_{i=1}^{T} Y_t - \frac{T+1}{2}\mathcal{B}_T; \tag{2.3}$$

$$\mathcal{B}_T = \frac{6}{T^2-1}(\frac{2}{T}\sum_{t=1}^{T} tY_t - \frac{T+1}{T}\sum_{t=1}^{T} Y_t). \tag{2.4}$$

Theta lines can be understood as functions of the linear regression model directly applied to the data from this perspective. Indeed, the Theta method's projections for h steps ahead are an ad hoc combination (50 percent - 50 percent) of the linear extrapolations of $\zeta(0)$ and $\zeta(2)$.

- When $\theta < 1$ is applied to the second differences of the data, the decomposition process is defined by a theta coefficient, which reduces the second differences and improves the approximation of series behavior.
- If $\theta = 0$, the deconstructed line is turned into a constant straight line (Fig. 1).
- If $\theta > 1$ then the short term movements of the analyzed series show more local curvatures (Fig. 1).

We will refer to the above setup as the standard Theta method. The steps for building the theta method are as follows:

1. **Deseasonalisation:** Firstly, the time series data is tested for statistically significant seasonal behaviour. A time series is seasonal if

$$|\rho_m| > q_{1-\frac{a}{2}}\sqrt{\frac{1+2\sum_{i=1}^{m-1}\rho_i^2}{T}} \tag{2.5}$$

   where ρ_k denotes the lag k autocorrelation function, m is the number of the periods within a seasonal cycle (for example, 12 for monthly data), T is the sample size, q is the quantile function of the standard normal distribution, and $(1-a)\%$ is the confidence level. Assimakopoulos and Nikolopoulo [1] opted for a 90% confidence level. If the time series is identified as seasonal, then it is deseasonalised via the classical decomposition method, assuming the seasonal component to have a multiplicative relationship.

2. **Decomposition**: The second step consits for the decomposition of the seasonally adjusted time series into two Theta lines, the linear regression line $\zeta(0)$ and the theta line $\zeta(2)$.
3. **Extrapolation:** $\zeta(2)$ is extrapolated using simple exponential smoothing (SES), while $\zeta(0)$ is extrapolated as a normal linear regression line.
4. **Combination:** the final forecast is a combination of the forecasts of the two θ lines using equal weights.
5. **Reseasonalisation:** In the presence of seasonality in first step, then the final forecasts are multiplied by the respective seasonal indices.

2.2. Optimized Theta Model [21]

The optimized form of Theta Method can be described by the following equation:

$$Y_t = w\zeta_t(\theta_1) + (1-w)\zeta_t(\theta_2),\ t = 1, \ldots, T \tag{2.6}$$

where $w \in [0,1]$ is the weight parameter as a function of the parameters θ_1 and θ_2 if $\theta_1 < 1$ and $\theta_2 > 1$ the weight parameter is given by

$$w := w(\theta_1, \theta_2) = \frac{\theta_2 - 1}{\theta_2 - \theta_1}, \tag{2.7}$$

these equations allow to construct a generalisation of the Theta method. Since we are interesting in modeling and predicting of daily pandemic data behavior, we fix $\theta_1 = 0$ and focus on the optimisation of the short-term component, $\theta_2 = \theta$ with $\theta > 1$, therefore the theta decomposition becomes:

$$Y_t = \left(1 - \frac{1}{\theta}\right)(\mathcal{A}_T + \mathcal{B}_T t) + \frac{1}{\theta}\zeta_t(\theta), \quad t = 1, \ldots, T. \tag{2.8}$$

The h-step-ahead forecasts calculated at origin T are given by

$$\widehat{Y}_{T+h|T} = \left(1 - \frac{1}{\theta}\right)[\mathcal{A}_T + \mathcal{B}_T(T+h)] + \frac{1}{\theta}\widetilde{\zeta}_{T+h|T}(\theta), \tag{2.9}$$

where $\widetilde{\zeta}_{T+h|T}$ is the extrapolation of $\zeta_t(\theta)$ by an SES method. In the case of $\theta = 2$, the Eq. (2.9) will correspond to the fourth step of the Standard Theta algorithm.

2.3. Dynamic standard and Dynamic optimised Theta model [21]

So far, we have considered that $\mathcal{A}_T$ and $\mathcal{B}_T$ are constant coefficients for all t. If we consider that these coefficients as dynamic functions; i.e., from the state t to the state $t+1$ we will only consider the prior information $\{Y_s, s \leq t\}$ when computing $\mathcal{A}_t$ and $\mathcal{B}_t$. Hence, we replace $\mathcal{A}_T$ and $\mathcal{B}_T$ in Eq. (2.9) with $\mathcal{A}_t$ and $\mathcal{B}_t$. Further, for $\theta = 2$, we can obtain a stochastic version of Standard Theta model, namely the dynamic standard Theta model.

2.4. Error metrics

Error metrics are used to quantify a forecasting model's error. They can give forecasters a way to compare the effectiveness competing models. There are many error metrics used in the literature, in this paper we have employed some common error metrics defined in the following. Let Y_t stands for the actual value and $\tilde{Y}_t$ is the predicted value by the model. T is the sample size of observed time series data.

Mean Absolute Error (MAE):

MAE is the Mean of Absolute value of Errors defined as the sum of absolute errors divided by the sample size T:

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|Y_t - \tilde{Y}_t\right|. \tag{2.10}$$

Root Mean Square Errors (RMSE):

RMSE can be defined as the standard deviation of the residuals values (differences between predicted values and the observed values). It is calculated as follows:

$$RMSE = \sqrt{\sum_{i=1}^{T}\frac{\left(\tilde{Y}_t - Y_t\right)^2}{T}}. \tag{2.11}$$

Mean Absolute Percentage Error (MAPE):

The mean absolute percentage error is given by the following formula:

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{Y_t - \tilde{Y}_t}{Y_t}\right|. \tag{2.12}$$

symmetric Mean Absolute Percentage Error (sMAPE):

Having defined the MAPE, we also take a look at one of the suggested alternatives to it - the symmetric MAPE in order to to overcome the asymmetry in MAPE measure. Then we define sMAPE as follows:

$$\text{sMAPE} = \frac{1}{T}\sum_{t=1}^{T}\frac{\left|\tilde{Y}_t - Y_t\right|}{\left(Y_t + \tilde{Y}_t\right)/2}. \tag{2.13}$$

3. Data description

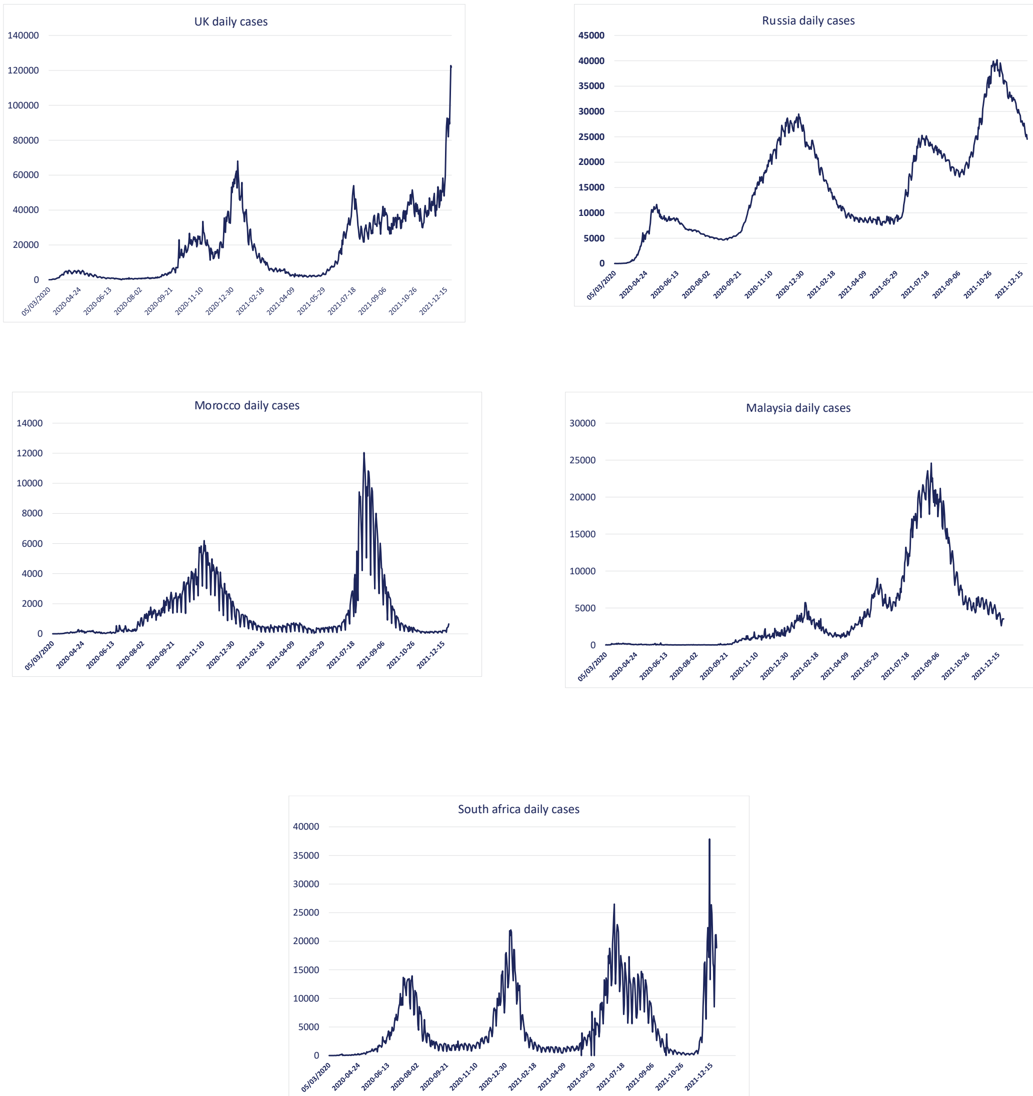

Figure 2: Daily Pandemic Confirmed cases of COVID-19 for United Kingdom (UK), South Africa (SA), Morocco, Malaysia, and Russia countries from 5 March, 2020 to 24 December, 2021.

	UK	**South Africa**	**Malaysia**	**Morocco**	**Russia**
Mean	18071	5114	4144	1447	15401
Median	10174	2316	1750	505	12599
Max.	122760	37875	24599	12039	40210
S.D.	19218	5650	5853	2082	9950

Table 1: Statistic features of studied data.

We observe that the countries investigated have varying levels of data variability. With a standard deviation (SD) of about 19218, the UK has the most variability in daily cases, while Morocco has the least fluctuation with a standard deviation of nearly 2082 (Table ??). Furthermore, we can remark that each country is characterized by its own disease confirmed daily cases curve according to the pandemic evolution in the period of observation. (See Fig. ??). In certain cases (the United Kingdom, South Africa, and Morocco), pandemic data fluctuates on a regular basis, whereas in others, a consistent pattern emerges (Malaysia, Russia).

Data	Theta Model	MAE	sMAPE	MAPE	RMSE
UK data	Standard Theta Method	**1916.127**	**0.1421**	0.1905	**3209.989**
	Dynamic Standard Theta Model	1916.764	0.1428	**0.1708**	3211.401
	Optimized Theta Model	1918.874	0.1477	0.2473	3211.009
	Dynamic Optimized Theta Model	1921.219	0.1443	0.1875	3211.484
Russia data	Standard Theta Method	492.3969	0.0643	0.1883	**712.8101**
	Dynamic Standard Theta Model	494.8383	0.0586	0.0740	714.5737
	Optimized Theta Model	491.9469	**0.0580**	0.5850	713.5040
	Dynamic Optimized Theta Model	**491.3338**	0.0582	**0.0670**	713.6362
South Africa data	Standard Theta Method	1134.446	0.288	16.420	**2178.052**
	Dynamic Standard Theta Model	1134.339	**0.281**	16.368	2178.851
	Optimized Theta Model	1134.624	0.290	16.737	2178.170
	Dynamic Optimized Theta Model	**1134.198**	0.282	**16.338**	2178.181
Moroccan data	Standard Theta Method	356.7994	0.3401	0.4082	**725.1188**
	Dynamic Standard Theta Model	356.9537	**0.3350**	0.3975	725.2397
	Optimized Theta Model	356.7405	0.3379	0.3985	725.1196
	Dynamic Optimized Theta Model	**356.7401**	0.3385	**0.3964**	725.1196
Malaysia data	Standard Theta Method	395.8821	0.2860	0.5354	**675.5662**
	Dynamic Standard Theta Model	395.3078	0.2631	**0.3366**	675.8996
	Optimized Theta Model	395.1413	0.2633	0.3497	675.5707
	Dynamic Optimized Theta Model	**395.0984**	**0.2609**	0.3381	675.5717

Table 2: Theta models Benchmark Results.

4. Results

To forecast the spread of the COVID-19 pandemic in the United Kingdom (UK), South Africa (SA), Morocco, Malaysia, and Russia, we have performed pandemic prediction using four statistical Theta models. Among these are the Standard Theta, Dynamic Standard Theta, Optimized Theta, and Dynamic Optimized Theta models. A set of performance metrics were used to evaluate the models' performance. From the results presented in Table ??, we conclude that in the cases of Russia, South Africa, and Moroccan pandemic data, the Dynamic Optimized Theta model has the lowest MAE and MAPE values, as well as the lowest MAE and sMAPE values. On the other hand, the Standard Theta Method is more accurate for UK pandemic data and has the lowest RMSE, MAE, and sMAE values.

5. Conclusion

In this study, we used Theta algorithms to forecast COVID-19 data cases in both standard and dynamic optimized variations. Pandemic data from five nations is used as study cases from March 5, 2020, to December 24, 2021. Different metric errors were used to assess each method's performance. The Standard Theta Method is proven to be more accurate for highly variable data, such as the UK pandemic data. On the other hand, Dynamic Optimized Theta models are more accurate and have less fluctuation when anticipating pandemic data. Combining theta-models time series approaches with artificial intelligence methods to explore the dynamical behavior of pandemic disorders will be interesting for future research investigations.

Declaration of Competing Interest

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

References

1. Assimakopoulos, V., & Nikolopoulos, K., The theta model: a decomposition approach to forecasting. International journal of forecasting, 16(4), 521-530, 2000.
2. Dudek, G., Short-term load forecasting using Theta method. In E3S Web of Conferences, 84, 01004, 2019.
3. Spiliotis, E., Assimakopoulos, V., & Nikolopoulos, K., Forecasting with a hybrid method utilizing data smoothing, a variation of the Theta method and shrinkage of seasonal factors. International Journal of Production Economics, 209, 92-102, 2019.
4. Hossain, M. R., & Ismail, M. T., Empirical mode decomposition based on theta method for forecasting daily stock price. Journal of Information and Communication Technology, 19(4), 533-558, 2020.
5. Petropoulos, F., & Spiliotis, E., The wisdom of the data: Getting the most out of univariate time series forecasting. Forecasting, 3(3), 478-497, 2021.
6. Spiliotis, E., Assimakopoulos, V., & Makridakis, S., Generalizing the theta method for automatic forecasting. European Journal of Operational Research, 284(2), 550-558, 2020.
7. Reich, Nicholas G., et al. "Accuracy of real-time multi-model ensemble forecasts for seasonal influenza in the US." PLoS computational biology 15.11 2019: e1007486.
8. Merkord, C. L., Liu, Y., Mihretie, A., Gebrehiwot, T., Awoke, W., Bayabil, E., ... & Wimberly, M. C., Integrating malaria surveillance with climate data for outbreak detection and forecasting: the EPIDEMIA system. Malaria journal, 16(1), 1-15, 2017.
9. Khanday, O. M., Dadvandipour, S., & Lone, M. A., Forecasting the spread of COVID19 in Hungary. medRxiv, 2020.
10. Shinde, G. R., Kalamkar, A. B., Mahalle, P. N., Dey, N., Chaki, J., & Hassanien, A. E., Forecasting models for coronavirus disease (COVID-19): a survey of the state-of-the-art. SN Computer Science, 1(4), 1-15, 2020.
11. Pal, R., Sarkar, S., & Mukhopadhyay, A., A Spreadsheet-Based Short Time Forecasting Method for the COVID-19 Pandemic. Transactions of the Indian National Academy of Engineering, 7(1), 185-196, 2022.
12. Yadav, S., Yadav, N., & Yadav, P. K. The Peak and Size of COVID-19 in India: SARIMA and Forecast. Demography India, 50(2), 75-86, 2022.
13. Thapa, P. "Time Series Analysis and Forecast of COVID-19 Pandemic." Decision Sciences for COVID-19. Springer, Cham, 2022. 97-106.
14. Paiva, H. M., Afonso, R. J. M., de Oliveira, I. L., & Garcia, G. F., A data-driven model to describe and forecast the dynamics of COVID-19 transmission. PloS one, 15(7), 2020, e0236386.
15. Sabry, I., Mourad, A. H. I., Idrisi, A. H., & ElWakil, M., Forecasting COVID-19 Cases in Egypt Using ARIMA-Based Time-Series Analysis. Eurasian Journal of Medicine and Oncology, 5(2), 123-131, 2021.
16. Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A., Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. Journal of infection and public health, 13(7), 914-919, 2020.
17. Tan, C. V., Singh, S., Lai, C. H., Zamri, A. S. S. M., Dass, S. C., Aris, T. B., ... & Gill, B. S., Forecasting COVID-19 Case Trends Using SARIMA Models during the Third Wave of COVID-19 in Malaysia. International journal of environmental research and public health, 19(3), 1504, 2022.
18. Devaraj, J., Elavarasan, R. M., Pugazhendhi, R., Shafiullah, G. M., Ganesan, S., Jeysree, A. K., ... & Hossain, E., Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant?. Results in Physics, 21, 103817, 2021.

19. Xu, Z., Su, C., Xiao, Y., & Wang, F., Artificial intelligence for COVID-19: Battling the pandemic with computational intelligence. Intelligent Medicine, 2022.

20. Alassafi, M. O., Jarrah, M., & Alotaibi, R., Time series predicting of COVID-19 based on deep learning. Neurocomputing, 468, 335-344, 2022.

21. Fiorucci J.A., Pellegrini T.R., Louzada F., Petropoulos F., The Optimised Theta Method, 2015, http://arxiv.org/abs/1503.03529.

22. Alzahrani, S. I., Aljamaan, I. A., & Al-Fakih, E. A., Forecasting the spread of the COVID-19 pandemic in Saudi Arabia using ARIMA prediction model under current public health interventions. Journal of infection and public health, 13(7), 914-919, 2020.

23. Khan, F., Saeed, A., & Ali, S., Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using Vector Autoregressive model in Pakistan. Chaos, Solitons & Fractals, 140, 110189, 2020.

24. Fiorucci J.A., Pellegrini T.R., Louzada F., Petropoulos F., Koehler, A. Models for optimising the theta method and their relationship to state space models, International Journal of Forecasting, 2016.

25. Thomakos, D., Nikolopoulos, K., Fathoming the theta method for a unit root process, IMA Journal of Management Mathematics, 25(1), 105–124, 2014, https://doi.org/10.1093/imaman/dps030

26. Nikolopoulos, K. I., & Thomakos, D. D., Forecasting with the theta method: theory and applications. John Wiley & Sons, 2019.

M. Acim, M. Zahid and A. Ez-Zetouni,
Department of Mathematics, Faculty of Sciences and Technology,
Cadi Ayyad University, B.P. 549, Marrakesh C.P. 40.000, Morocco.
E-mail address: maria.acim987@gmail.com, adilezzetouni@gmail.com & zahidmehdi10@gmail.com