Time Indexed Covariate Analysis of the Administered Vaccine Doses on Covid-19 Spread

Mohsen Maleki, Mohammad Reza Mahmoudi, Hamid Bidram and Amir Mosavi

April 18, 2022

Time indexed covariate analysis of the administered vaccine doses on Covid-19 spread

Mohsen Maleki [a], Mohammad Reza Mahmoudi [b, *] Hamid Bidram [a], Amir Mosavi [c, d, *]

[a] *Department of Statistics, Faculty of Mathematics and Statistics, University of Isfahan, Isfahan, 81746-73441, Iran*

[b] *Department of Statistics, Faculty of Science, Fasa University, Fasa, Fars, Iran*

[c] *Faculty of Civil Engineering, Technische Universität Dresden, 01069 Dresden, Germany;*

[d] *John von Neumann Faculty of Informatics, Obuda University, 1034 Budapest, Hungary*

Abstract

In this study, we focus on the spread of COVID–19 on besides of vaccination of this disease. Total infected cases of COVID–19 have considered and model it by a light/heavy tailed auto–regressive model with innovations belong on the flexible class of the two-piece scale mixtures of normal (*TP–SMN*) family. Also considering the covariate variables which are indexed by time are considered in the model to more accuracy of modeling. An EM type algorithm has considered for finding the maximum likelihood estimations of the model parameters. Modelling and prediction of infected numbers of COVID–19 in the U.S. has considered and vaccinated numbers of COVID–19 is considered as auxiliary (covariate) in the model.

Keywords: Auto–regressive model with covariate; COVID–19 vaccinated cases; ECME–algorithm; Infected numbers of COVID–19; ML estimates; Two–piece scale mixtures of normal family; Vaccinated numbers of COVID–19.

1. Introduction

The rapid spread of the infectious COVID–19 disease has affected whole of the world, and preparation and confronting for the prevalence of COVID–19 diseases requires careful plans and operating. Some mathematically modeling of the spreading of COVID–19, have considered by several mathematician authors, which some of the recent researches are given as follows. The confirmed and recovered cases and also death rate of COVID–19 have

modeled and forecasted via using asymmetry time series by Maleki et al. [1,2]. The cumulative confirmed, recovered and deaths cases of COVID–19 cases for top–16 countries, have forecasted via using statistical machine learning models has done by ArunKumar et al. [3]. ArunKumar et al. [4] have used gated recurrent units and long short-term memory cells and forecast the COVID–19 using the deep layer recurrent neural networks. Lau et al. [5] considered SARIMA model on the infectious diseases. Kalantari [6] forecast the COVID–19 using optimal singular spectrum analysis. Al-qaness et al. [7] forecast the COVID–19 in the Russia and Brazil, using efficient artificial intelligence forecasting models. Amruta et al. [8] have a study on the SARS–CoV–2 mediated neuro-inflammation and the impact of COVID–19 in neurological disorders. Luo et al. [9] considered a novel based on traditional Richards model and its application in COVID–19. The interested reader can refer to [10-15] to see some recent papers published on the COVID–19.

Time series are useful statistical model to forecast and control the datasets which are gathered by time. We can also use some flexible family of distributions with light/heavy tails and symmetry/asymmetry behaviors in the structure of the time series models which can be more expedient for datasets with outliers or/and skewed behavior, (see [16-24]). In this paper we consider a wide class of distributions in the structure of Auto–Regressive model with covariate, which is known as two-piece scale mixtures of normal (*TP–SMN*) family of distributions. The useful *TP–SMN* distributions were introduced by the authors of [25,26] and have used in the several statistical models to show the flexibility and performances of the family (see [27-30]).

Using auxiliary (covariate) on the statistical models we can improve the obtained results, especially in time series models. One of the main application of the time series modelling related to the COVID–19 is to single out the role plaid by the variables of COVID–19 indexed by time, for example by modelling and forecasting of the number of daily confirmed cases. Obviously by increasing the daily number of vaccination it will effect on the daily number of confirmed cases, thus expecting an extensively decay on the confirmed cases in the near future. We will model the infected numbers of COVID–19 by a robust time series models with vaccinated numbers of COVID–19 as a covariate. We have shown in the application section of the work that by using this covariate improves the modelling and predictions.

The paper is set as follows. Some required aspects of the *TP–SMN* distributions are provided in Section 2. The *TP–SMN* Auto–Regressive (AR) model with time indexed covariate is given in Section 3. In Section 4, the proposed model is fitted on the infected numbers of COVID–19 in the U.S. with vaccinated numbers of COVID–19 covariate. Finally in in Section 5, some conclusions of the work are provided.

2. The TP–SMN distributions

Herein, we briefly review main features of the *TP–SMN* distributions.

Let random variable X has renown *SMN* distributions given in [31] denoted by $X \sim SMN\,(\mu, \sigma, \boldsymbol{v})$. Then X possesses the probability density function (PDF) given by,

$$f_{SMN}(x|\mu, \sigma, \boldsymbol{v}) = \int_0^\infty \phi(x|\mu, u^{-1}\sigma^2)\, dH(u|\boldsymbol{v}), \quad x \in \mathbb{R}, \tag{1}$$

in which $\phi(\cdot\,|\mu, \sigma^2)$ is the density of $N(\mu, \sigma^2)$ distribution, $H(\cdot\,|\boldsymbol{v})$ indicates the cumulative distribution function (CDF) related to the scale mixing variable u for each *TP–SMN* members which indexed by the parameters $\boldsymbol{v}$.

The applicable *TP–SMN* family is defined via the *SMN* family, which involves asymmetric light–tailed two–piece–normal (*TP–N*), and asymmetric heavy–tailed two–piece–t (*TP–T*), two–piece–slash (*TP–SL*) and two–piece–contaminated–normal (*TP–CN*) distributions and also covers their symmetrically forms. A random variable Y distributed with *TP–SMN* distributions denoted by $Y \sim TP\text{–}SMN\,(\mu, \sigma, \boldsymbol{v}, \gamma)$ with the PDF

$$g(y|\mu, \sigma, \gamma, \boldsymbol{v}) = \begin{cases} 2(1-\gamma)\, f_{SMN}(y|\mu, \sigma(1-\gamma), \boldsymbol{v}), & y \le \mu, \\ 2\gamma\, f_{SMN}(y|\mu, \sigma\gamma, \boldsymbol{v}), & y > \mu, \end{cases} \tag{2}$$

where $0 < \gamma < 1$ is the skewness parameter, $f_{SMN}(\cdot\,|\mu, \sigma, \boldsymbol{v})$ is the *SMN* PDF given in (1).

Mean and variance of $Y \sim TP\text{–}SMN\,(\mu, \sigma, \boldsymbol{v}, \gamma)$ are $E(Y) = \mu - b\sigma(1 - 2\gamma)$ and $\mathrm{Var}(Y) = \sigma^2[c_2 k_2 - b^2 c_1^2]$, respectively, where $b = \sqrt{2/\pi}\, k_1$, $c_r = \gamma^{r+1} + (-1)^r(1-\gamma)^{r+1}$ and k_r is given in Table 1. It is noteworthy that for $\gamma = 0.5$, the well-known *SMN* family (which is symmetric) has extracted from the *TP–SMN* family.

Table 1. Some main TP−SMN distributions.

Distributions	$f_{SMN}(\cdot),\ y \in \mathbb{R}$	k_r
TP–N	$\phi(y; \mu, \sigma^2)$	---
TP–T	$\dfrac{\Gamma\left(\frac{v+1}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi v}\sigma}\left(1+\dfrac{d}{v}\right)^{-\frac{v+1}{2}};\ \boldsymbol{v} = v$	$\left(\dfrac{v}{2}\right)^{r/2}\dfrac{\Gamma\left(\frac{v-r}{2}\right)}{\Gamma\left(\frac{v}{2}\right)}, v > r$
TP–SL	$v\int_0^1 u^{v-1}\phi(y; \mu, u^{-1}\sigma^2)\,\mathrm{d}u;\ \boldsymbol{v} = v$	$\dfrac{2v}{2v-r}, v > \dfrac{r}{2}$
TP–CN	$v\phi(y; \mu, \tau^{-1}\sigma^2) + (1-v)\phi(y; \mu, \sigma^2);\ \boldsymbol{v} = (v, \tau)^{\mathsf{T}}$	$\dfrac{v}{\tau^{r/2}} + 1 - v$

A key feature of using the well-known algorithm to estimate the statistical models using the *TP–SMN* distributions is that we can utilize the *TP–SMN* distributions in the form mixture distribution with two–components and distinct supports in (3). So let $Y \sim TP\text{--}SMN\,(\mu, \sigma, \boldsymbol{v}, \gamma)$, then the PDF given in (2) is represented as:

$$g(y|\mu, \sigma_1, \sigma_2, \boldsymbol{v}) = 2\pi f_{SMN}(y|\mu, \sigma_1, \boldsymbol{v})I_{(-\infty,\mu]}(y) + 2(1-\pi)f_{SMN}(y|\mu, \sigma_2, \boldsymbol{v})I_{(\mu,+\infty)}(y), \quad (3)$$

where $\pi = \dfrac{\sigma_1}{\sigma_1+\sigma_2}$, $\sigma_1 = \sigma(1-\gamma)$ and $\sigma_2 = \sigma\gamma$, and the parameters σ and γ (respectively scale and slant) in (2) are meliorated as $\sigma = \sigma_1 + \sigma_2$ and $\gamma = \sigma_2/(\sigma_1 + \sigma_2)$.

By employing the latent variables $Z_j; j = 1,2$, in terms of the mixture components (3), the proposed *TP–SMN* variable Y can be represented in the stochastic form

$$\begin{cases} Y|Z_1 = 1 \sim SMN(\mu, \sigma_1, \boldsymbol{v})I_T(y_i), \\ Y|Z_2 = 1 \sim SMN(\mu, \sigma_2, \boldsymbol{v})I_{T^c}(y_i), \end{cases} \quad (4)$$

where $T = (-\infty, \mu)$ and $SMN(\cdot)I_T(\cdot)$ denotes the *SMN*–distributions which are truncated on the interval T, and the random vector $\boldsymbol{Z} = (Z_1, Z_2)^{\mathsf{T}}$ with the probability mass function (PMF) given by:

$$P(\boldsymbol{Z} = \boldsymbol{z}) = \pi^{z_1}(1-\pi)^{z_2}; \quad z_1, z_2 = 0,1, \quad (5)$$

and is denoted by $\mathbf{Z} \sim M(1, \pi, 1 - \pi)$. Note that each component-label variable $Z_i, i = 1,2$ has also a Bernoulli distribution and $Z_1 + Z_2 = 1$.

3. TP-SMN Auto-Regressive with covariates model

The Auto-Regressive model with covariates and order p based on the *TP-SMN* innovations, indicated using *TP-SMN-ARX(p)* model via intercept is considered and expressed by

$$Y_t = \mu + \sum_{j=1}^{p} \alpha_j Y_{t-j} + \sum_{k=1}^{m} \beta_k X_{kt} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots, \tag{6}$$

where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_p)^{\top}$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)^{\top}$ are respectively the Auto-Regressive and covariates coefficients vectors, and $\{\varepsilon_t\}$ is a i.i.d. noises distributed sequence as

$$\varepsilon_t \overset{iid}{\sim} TP - SMN\ (0, \sigma, \boldsymbol{v}, \gamma), \quad t = 0, \pm 1, \pm 2, \dots \ . \tag{7}$$

Considering the *ARX* model (6) with parameters $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu, \sigma_1, \sigma_2, \boldsymbol{v})^{\top}$, to have a more appropriate form, the model can be represented by $Y_t = \mu + \boldsymbol{\alpha}^{\top} \boldsymbol{Y}_{t-1} + \boldsymbol{\beta}^{\top} \boldsymbol{X}_t + \varepsilon_t$, for which $\boldsymbol{Y}_{t-1} = (Y_{t-1}, \dots, Y_{t-p})^{\top}$, $\boldsymbol{X}_t = (X_{1t}, \dots, X_{mt})^{\top}$ and $\boldsymbol{Y} = (Y_{-p+1}, \dots, Y_0, Y_1, \dots, Y_n, \boldsymbol{X}_1^{\top}, \dots, \boldsymbol{X}_n^{\top})^{\top}$ as the sample vector.

Considering the non-stochastic part $Y_{-p+1}, \dots, Y_0$ of the sample $\boldsymbol{Y}$, the (conditional) likelihood function related to the *TP-SMN-ARX(p)* model is given by

$$\ell(\boldsymbol{\Theta}|Y, X) = \sum_{t=1}^{n} \log g_{Y_t|Y_{t-1} = y_{t-1}, X_t = x_t}(y_t | \boldsymbol{\alpha}^{\top} \boldsymbol{y}_{t-1} + \boldsymbol{\beta}^{\top} \boldsymbol{x}_t + \mu, \sigma_1, \sigma_2, \boldsymbol{v}). \tag{8}$$

Because of complexities of the *SMN*-densities (1), finding the *ML* estimates related to the *TP-SMN-ARX(p)* model parameters via (8) are difficult process. But considering an appropriate hierarchically representation of the *TP-SMN* distributions on besides of the established *ARX* model, obtaining the *ML* estimates of parameters can be desired by an *EM*-type scheme at the follows.

Consider the complete data $\boldsymbol{D} = (\boldsymbol{Y}^{\top}, \boldsymbol{X}^{\top}, \boldsymbol{U}^{\top})^{\top}$, with observed part of the data $(\boldsymbol{Y}^{\top}, \boldsymbol{X}^{\top})^{\top}$ and missing (latent) part of the data $\boldsymbol{U} = (U_1, \dots, U_n)^{\top}$ and $\boldsymbol{Z}_t = (Z_{t1}, Z_{t2})^{\top}$; $t = 1,2, \dots, n$. Then, regarding to the mixture representation (4), the PMF in (5) and the well-

known stochastic representation of *SMN* distributions in Andrews and Mallows [31], the *TP–SMN–ARX* model using (6) can be represented in the hierarchically manner

$$Y_t|Y_{t-1} = y_{t-1}, X_t = x_t, U_t = u_t, Z_{ti} = 1 \overset{ind.}{\sim} N(\alpha^\top y_{t-1} + \beta^\top x_t + \mu, u_t^{-1}\sigma_i^2)I_{T_t}(y_t)^{2-i}I_{T_t^c}(y_t)^{i-1},$$

$$U_t|Z_{ti} = 1 \overset{ind.}{\sim} H(u_t|v),$$

$$Z_t \overset{iid.}{\sim} M(1, \pi, 1 - \pi), \tag{9}$$

where $i = 1,2$, $T_t = (-\infty, \alpha^\top y_{t-1} + \beta^\top x_t + \mu)$ for $t = 1,2,\ldots,n$ and $N(\cdot)I_T(\cdot)$ is the normal distribution truncated on the domain T.

An *ECME* algorithm generated by Liu and Rubin [32] and the hierarchical form of the *TP–SMN–ARX* model in (9) is utilized in to extract the *ML* estimates of the parameters.

Regarding the hierarchical form (9), conclude the conditional completed log–likelihood function attributed to the *TP–SMN–ARX(p)* model given by

$$\ell(\Theta) = -n\log(\sigma_1 + \sigma_2) - \frac{1}{2}\sum_{t=1}^n \sum_{i=1}^2 W_{ti}\left(\frac{Y_t - \alpha^\top Y_{t-1} - \beta^\top X_t - \mu}{\sigma_i}\right)^2 + \sum_{t=1}^n \sum_{i=1}^2 Z_{ti}\log h(U_t|v),$$

where $\Theta = (\alpha^\top, \beta^\top, \mu, \sigma_1, \sigma_2, v)^\top$ and $W_{ti} = U_t Z_{ti}$ for $t = 1,2,\ldots,n$ and $i = 1,2$. Defining the quantities $\hat{z}_{ti} = E[Z_{ti}|\hat{\Theta}, Y, X]$ and $\hat{w}_{ti} = E[U_t Z_{ti}|\hat{\Theta}, Y, X]$, and according to the properties of the conditional expectation, $\hat{w}_{ti} = \hat{\kappa}_{ti}\hat{z}_{ti}$ is obtained where $\hat{z}_{t1} = I_{(-\infty, \hat{\alpha}^\top y_{t-1} + \hat{\beta}^\top x_t + \hat{\mu}]}(y_t)$ for $\hat{z}_{t2} = 1 - \hat{z}_{t1}$ and $\hat{\kappa}_{ti} = E[U_t|\hat{\Theta}, y_t, x_t, Z_{ti} = 1]$ for the *TP-SMN-ARX* members, are considered by

- *TP–N–ARX* model: $\hat{\kappa}_{ti} = 1,$
- *TP–T–ARX* model: $\hat{\kappa}_{ti} = \frac{\hat{v}+1}{\hat{v}+d_{ti}},$
- *TP–SL–ARX* model: $\hat{\kappa}_{ti} = \frac{2\hat{v}+1}{d_{ti}}\frac{P_1(\hat{v}+3/2, d_{ti}/2)}{P_1(\hat{v}+1/2, d_{ti}/2)},$
- *TP–CN–ARX* model: $\hat{\kappa}_{ti} = \frac{\hat{\tau}^2\hat{v}e^{-\hat{\tau}d_{ti}/2} + (1-\hat{v})e^{-d_{ti}/2}}{\hat{\tau}\hat{v}e^{-\hat{\tau}d_{ti}/2} + (1-\hat{v})e^{-d_{ti}/2}},$

where $d_{ti} = \left(\frac{y_t - \hat{\alpha}^\top y_{t-1} - \hat{\beta}^\top x_t - \hat{\mu}}{\hat{\sigma}_i}\right)^2$ and $P_x(a, b)$ denotes the evaluated cumulative distribution function (CDF) of the Gamma(a, b) distribution at point x.

Determination of Q–function as of $Q(\Theta|\widehat{\Theta}^{(k)}) = E_\theta[\ell(\Theta)|\widehat{\Theta}^{(k)}, Y, X]$ is the E–Step of the *ECME*–algorithm on the $(k+1)$th algorithm's iteration as

$$Q(\Theta|\widehat{\Theta}^{(k)}) = -n\log(\sigma_1 + \sigma_2) - \frac{1}{2}\sum_{t=1}^{n}\sum_{i=1}^{2}\widehat{w}_{ti}^{(k)}\left(\frac{Y_t - \alpha^\top Y_{t-1} - \beta^\top X_t - \mu}{\sigma_i}\right)^2$$
$$+ \sum_{i=1}^{n}\sum_{j=1}^{2}E[Z_{ti}\log h(U_t|\nu)|\widehat{\Theta}^{(k)}, Y, X],$$

where $\widehat{w}_{ti}^{(k)} = E[U_t Z_{ti}|\widehat{\Theta}^{(k)}, Y, X] = \widehat{\kappa}_{ti}^{(k)}\widehat{z}_{ti}^{(k)}$ and hereafter consider $\widehat{\omega}_t^{(k)} = \sum_{i=1}^{2}\widehat{w}_{ti}^{(k)}/\sigma_i^{2(k)}$.

Conditional maximizations of the ECME–algorithm on the $(k+1)$th algorithm's iteration as follows:

CM–step of the algorithm for coefficients:

$$\widehat{\alpha}^{(k+1)} = \left(\sum_{t=1}^{n}\widehat{\omega}_t^{(k)}Y_{t-1}Y_{t-1}^\top\right)^{-1}\sum_{t=1}^{n}\widehat{\omega}_t^{(k)}(Y_t - \widehat{\beta}^{\top(k)}X_t - \widehat{\mu}^{(k)})Y_{t-1},$$

$$\widehat{\beta}^{(k+1)} = \left(\sum_{t=1}^{n}\widehat{\omega}_t^{(k)}X_t X_t^\top\right)^{-1}\sum_{t=1}^{n}\widehat{\omega}_t^{(k)}(Y_t - \widehat{\alpha}^{\top(k+1)}Y_{t-1} - \widehat{\mu}^{(k)})X_t,$$

CM–step of the algorithm for intercept (location parameter):

$$\widehat{\mu}^{(k+1)} = \frac{\sum_{t=1}^{n}\widehat{\omega}_t^{(k)}(Y_t - \widehat{\alpha}^{\top(k+1)}Y_{t-1} - \widehat{\beta}^{\top(k+1)}X_t)}{\sum_{t=1}^{n}\widehat{\omega}_t^{(k)}}.$$

CM–step of the algorithm for scale parameters:

Solving the stressed cubic equations $\sigma_i^3 + p\sigma_i + q = 0$, conclude the update $\widehat{\sigma}_i^{(k+1)}$ for $i = 1,2$, where $p = -\frac{1}{n}\sum_{t=1}^{n}\widehat{w}_{ti}^{(k)}(X_t - \widehat{\alpha}^{\top(k+1)}Y_{t-1} - \widehat{\beta}^{\top(k+1)}X_t - \widehat{\mu}^{(k+1)})^2$, for which $q = p\sigma_2 I_{(i=1)} + p\sigma_1 I_{(i=2)}$. It is noteworthy that $p < 0$ implies $q < 0$, therefor the proposed cubic equation possesses a positive unique just root.

CML–step of the algorithm for scale mixing parameters:

$$\boldsymbol{v}^{(k+1)} = \text{argmax}_{\boldsymbol{v}}\ell\left(\widehat{\boldsymbol{\alpha}}^{\mathsf{T}(k+1)}, \widehat{\boldsymbol{\beta}}^{\mathsf{T}(k+1)}, \widehat{\mu}^{(k+1)}, \widehat{\sigma}_1^{(k+1)}, \widehat{\sigma}_2^{(k+1)}, \boldsymbol{v}\middle|\boldsymbol{Y}, \boldsymbol{X}\right),$$

where the log-likelihood function $\ell(\cdot|\boldsymbol{Y}, \boldsymbol{X})$ is in (8).

A sufficient convergence rule for the algorithm iterates is satisfied when $\left|\ell\left(\widehat{\boldsymbol{\Theta}}^{(k+1)}\middle|\boldsymbol{Y}, \boldsymbol{X}\right)/\ell\left(\widehat{\boldsymbol{\Theta}}^{(k)}\middle|\boldsymbol{Y}, \boldsymbol{X}\right) - 1\right| \leq \varepsilon$, for the known values of the tolerance ε. In the work we employed $\varepsilon = 10^{-2}$, while choosing the tolerance values may vary in other statistical models.

4. Application

This section investigates the infected numbers of COVID–19 since 19–Jan–2021 to 27–Mar–2021 related to the U.S. (see Fig 1) and vaccinated numbers of COVID–19 in there and in the corresponding time (see Fig 2), and fit the *TP–SMN–ARX* to the proposed datasets.

After some suitable transformations in Brockwell and Davis [33], the stationary data which are ready to modeling are obtained and plotted in Fig 3. Using some model selection criteria and considering the PACF (partial auto-correlation function) diagram, shown in Fig 4 conclude that the best TP–T–ARX(5) model was fitted to the stationary series of the infected numbers of COVID–19 in the form of

TP–SMN–ARX model:

$$Y_t = -9139.5940 - 0.6390\, Y_{t-1} - 0.4393\, Y_{t-2} - 0.4123\, Y_{t-3} - 0.4340\, Y_{t-4} - 0.3338\, Y_{t-5}$$
$$+ 0.0029\, X_t + \varepsilon_t,$$

where $\varepsilon_t \sim TP\text{–}T(0, \sigma = 23211.9000, \gamma = 0.5582, v = 3.0939)$.

Gaussian–ARX model:

$$Y_t = -1995.2399 - 0.6627\, Y_{t-1} - 0.3799\, Y_{t-2} - 0.3730\, Y_{t-3} - 0.3878\, Y_{t-4} - 0.2617\, Y_{t-5}$$
$$+ 0.0036\, X_t + \varepsilon_t,$$

where $\varepsilon_t \sim N(0, \sigma = 17888.5400)$.

According to log-likelihood and Akaike information criterion (AIC) criteria, which are recorded in Table 2, the TP–T–ARX model is more reasonable than ordinary Gaussian–ARX model. The ACF (auto–correlation function) diagram of estimated innovations (residuals) is shown in Fig. 5 (Left), and also the histogram of estimated innovations for which the

estimated heavy–tailed TP–T density is superimposed on it, is given in Fig. 5 (Right). Fig. 5 shows some reasonability of the proposed estimated model to the stationary series of infected numbers of COVID–19 related to the U.S.
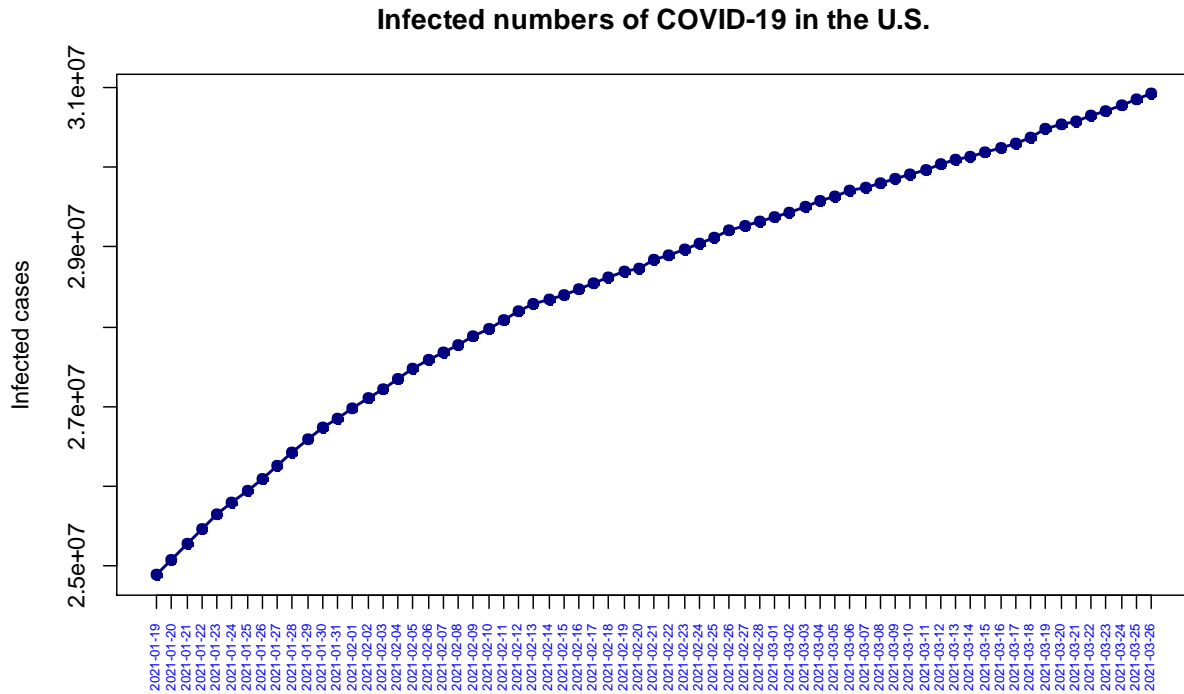
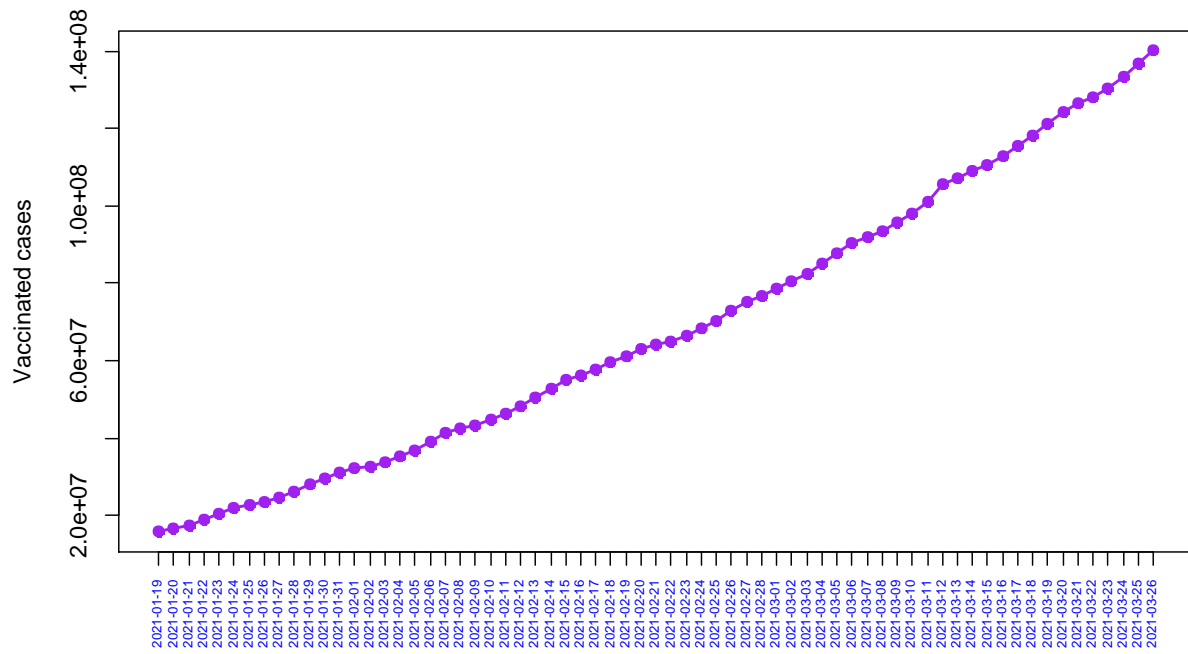**Fig 1.** Infected numbers of COVID-19 plot related to U.S. since 19-Jan to 27-Mar in 2021.

Fig 2. Vaccinated numbers of COVID-19 plot related to U.S. since 19-Jan to 27-Mar in 2021.
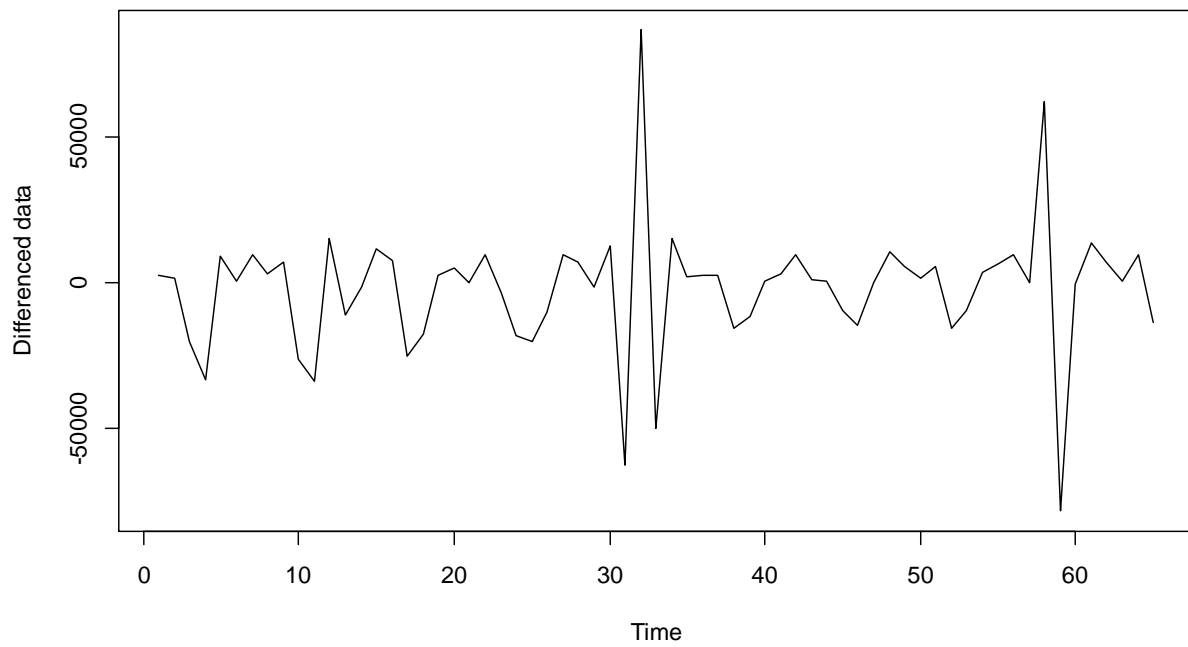
Fig 3. Differenced transform the infected numbers of COVID-19 plot.

PACF of differenced data
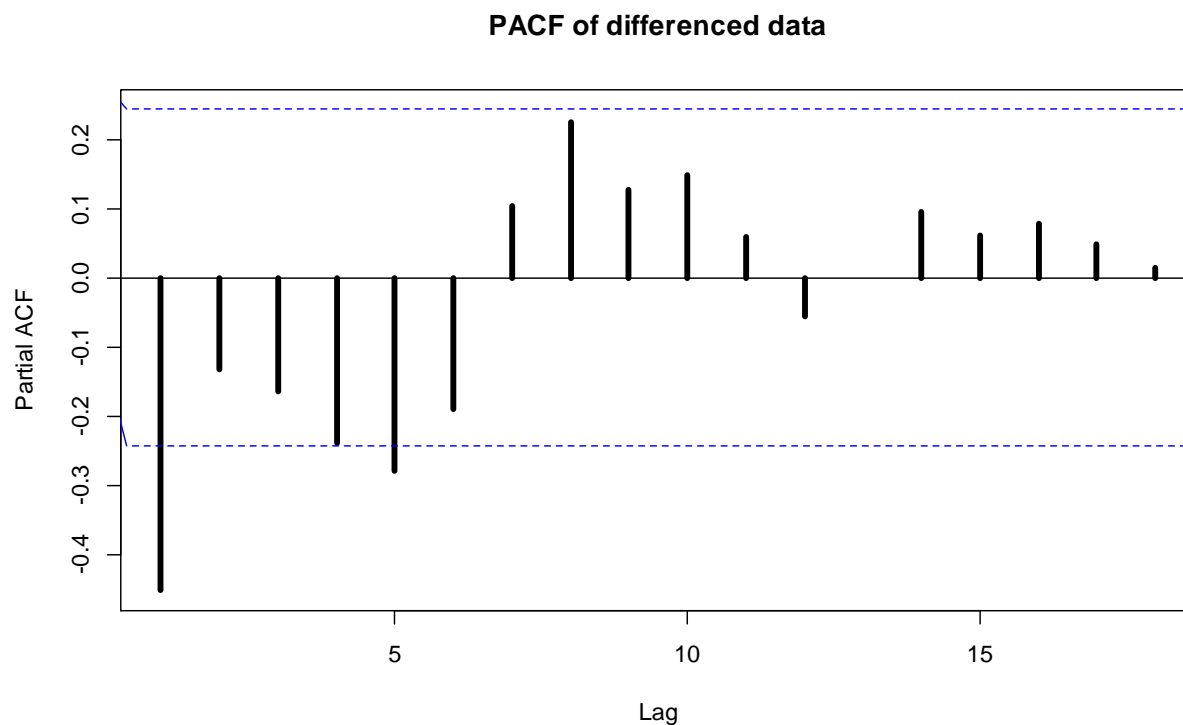
Fig 4. PACF of the differenced data (stationary).

To investigate the suitability and goodness of fit tests for the proposed fitted TP–SMN–ARX model, P–values of the Box–Pierce and Ljung–Box tests on the estimated innovations are recorded in Table2. Note that the large P–values (near to one) of the proposed tests on the estimated innovations of TP–T–ARX model, demonstrate the reasonability proposed fitted TP–SMN–ARX model. Also comparing the corresponding P–values of the proposed tests on the ordinary Gaussian–ARX model, show the better performances of the TP–SMN–ARX model (see Table 2).

Table 2. P-values of the Box–Pierce and Ljung–Box tests on the residuals of the TP–SMN–ARX and ordinary Gaussian–ARX models.

Model	TP–T–ARX		Gaussian–ARX	
Log-likelihood	-664.5078		-729.1800	
AIC	1350.0790		1474.3600	
Test	Box–Pierce	Ljung–Box	Box–Pierce	Ljung–Box
P–values	0.9882	0.9625	0.9613	0.8821
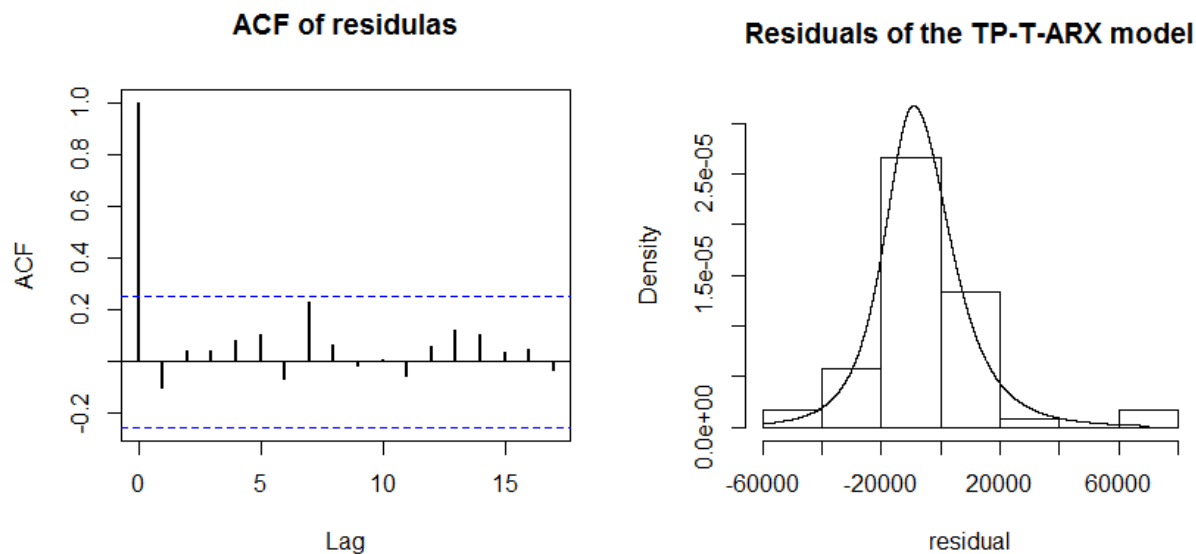
Fig 5. Sample ACF diagram of the estimated innovations of the fitted model (Left) Histogram of the estimated innovations of the fitted model via superimposed estimated TP-T density (Right).

To shoe further reasonability of the fitted model, we predicted the 10 data of the infected cases (2021-Mar-15 up to 2021-Mar-24), then fitted the TP–SMN–ARX models on them and forecast these data. Table 3 contains the predictions with and without covariate and 95% confidence intervals for them. Also Fig. 6, show the forecasted values based on the modelling with and without covariate which superimposed the diagram of the real values of infected numbers of COVID-19 related to U.S. Percent Relative Prediction Error (PRPE) of the predicted values via the model with covariate is 0.0762 % while is 0.1469 % for the model without covariate. So the results and Fig. 6 shows the importance of the covariate (considering the vaccination) on the modelling and predictions of the infected cases.

Table 3. Real values of the infected numbers of COVID-19 related to the U.S. since 2021-Mar-15 to 2021-Mar-24 with and without vaccinated numbers covariate and predictions with 95% confidence interval.

COVID-19 in the U.S.	Date	Real value	Prediction without covariate	Prediction with covariate	Lower C.I.	Upper C.I.
infected cases	2021-03-15	30181273	30154131	30175035	30148628	30214710
	2021-03-16	30234911	30219435	30222759	30184850	30263368
	2021-03-17	30298419	30256412	30285499	30248798	30324452
	2021-03-18	30361971	30338946	30355790	30318956	30394125

12

	2021-03-19	30487505	30415297	30414364	30341297	30492385
	2021-03-20	30534590	30648341	30593349	30537136	30672910
	2021-03-21	30581123	30541449	30596553	30547796	30659999
	2021-03-22	30641164	30581496	30621390	30582911	30663694
	2021-03-23	30708302	30672908	30686643	30649897	30727696
	2021-03-24	30776056	30756192	30769650	30703639	30810684

5. Conclusion

Coronaviruses are a wide category that affects gastrointestinal, hepatic, neurological and respiratory systems. Extended vaccination might reduce the prevalence of COVID–19 diseases. In our methodology, by considering a robust class of time series models with covariate, to model the infected numbers of COVID–19 in the U.S. with the vaccinated COVID–19 cases as the covariate. Performances of the proposed fitted model show that it would be better considering the vaccinated numbers of COVID–19 as the covariate, to predict infected numbers of COVID–19. Also there exist other desirable results that can be obtained by the model described in this paper, for example mobility data (indexed by time, for example in the daily form) can be a covariate to better modelling and prediction of the death rate of COVID–19. The free source *R* software [34] version 4.0.5 with a core i7 760 processor 2.8 GHz is used to implements the algorithms.

Predicted and real values
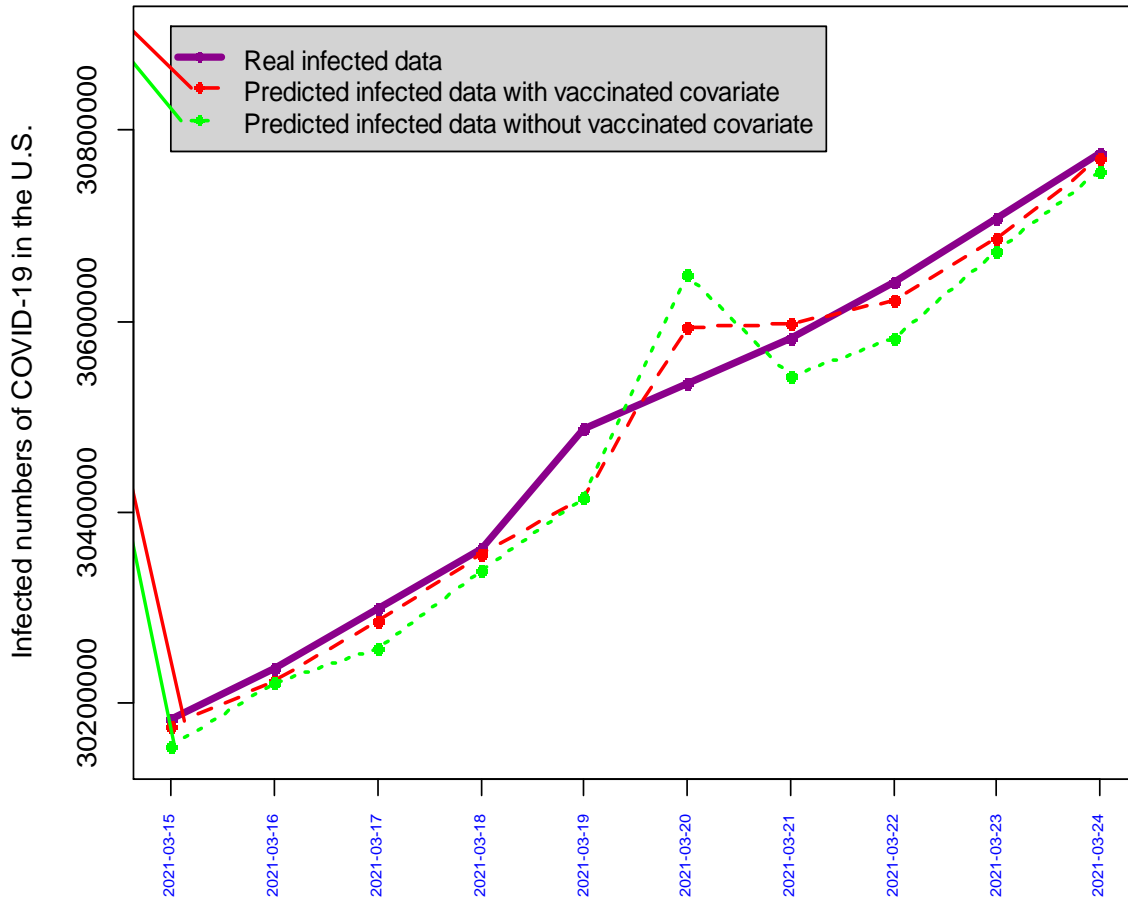
Fig 6. The experimental values and predicted infected numbers of COVID-19 related to the U.S. time series plots since 2021-Mar-15 up to 2021-Mar-24 with and without vaccinated numbers covariate.

Consent to Participate and Consent to Publish: Not applicable.

References

[1] M. Maleki, M.R. Mahmoudi, M.H. Heydari, and K.H. Pho. Modeling and forecasting the spread and death rate of coronavirus (COVID-19) in the world using time series models. *Chaos, Solitons & Fractals*, 110151. https://doi.org/10.1016/j.chaos.2020.110151, 2020.

[2] M. Maleki, M.R. Mahmoudi, D. Wraith, and K.H. Pho. Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel Medicine and Infectious Disease*, 101742. https://doi.org/10.1016/j.tmaid.2020.101742, 2020.

[3] K.E. ArunKumar, Dinesh V. Kalaga, Ch. Mohan Sai Kumar, Govinda Chilkoor, Masahiro Kawaji, Timothy M. Brenza. Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Applied Soft Computing*,107161, https://doi.org/10.1016/j.asoc.2021.107161, 2021.

[4] K.E. ArunKumar, Dinesh V. Kalaga, Ch. Mohan Sai Kumar, Masahiro Kawaji, Timothy M Brenza. Forecasting of COVID-19 using deep layer Recurrent Neural Networks (RNNs) with Gated Recurrent Units (GRUs) and Long Short-Term Memory (LSTM) cells. *Chaos, Solitons & Fractals*, 110861, https://doi.org/10.1016/j.chaos.2021.110861, 2021.

[5] K. Lau, I. Dorigatti, M. Miraldo, and K. Hauck. SARIMA-modelled greater severity and mortality during the 2010/11 post-pandemic influenza season compared to the 2009 H1N1 pandemic in English hospitals. *International Journal of Infectious Diseases*, https://doi.org/10.1016/j.ijid.2021.01.070, 2021.

[6] M. Kalantari. Forecasting COVID-19 pandemic using optimal singular spectrum analysis. *Chaos, Solitons & Fractals*, 110547, https://doi.org/10.1016/j.chaos.2020.110547, 2021.

[7] M.A.A. Al-qaness, A.I. Saba, A.H. Elsheikh, M.A. Elaziz, R.A. Ibrahim, S. Lu, A.A. Hemedan, S. Shanmugan,and A.A. Ewees. Efficient artificial intelligence forecasting models for COVID-19 outbreak in Russia and Brazil. *Process Safety and Environmental Protection*, https://doi.org/10.1016/j.psep.2020.11.007, 2021.

[8] N. Amruta, W.H. Chastain, M. Paz, R.J. Solch, I.C. Murray-Brown, J.B. Befeler, T.E. Gressett, M.T. Longo, E.B. Engler-Chiurazzi, and G. Bix. SARS-CoV-2 mediated neuroinflammation and the impact of COVID-19 in neurological disorders. *Cytokine & Growth Factor Reviews*, 58,1-15, 2021.

[9] X. Luo, H. Duan, and K. Xu. A novel grey model based on traditional Richards model and its application in COVID-19, Chaos, *Solitons & Fractals*,110480, https://doi.org/10.1016/j.chaos.2020.110480,

[10] T. Omay, and D. Baleanu. Fractional unit-root tests allowing for a fractional frequency flexible Fourier form trend: predictability of Covid-19. *Advances in Difference Equations*, 167, https://doi.org/10.1186/s13662-021-03317-9, 2021.

[11] A.S. Alshomrani, M.Z. Ullah, and D. Baleanu. Caputo SIR model for COVID-19 under optimized fractional order. *Advances in Difference Equations*, 185 (2021). https://doi.org/10.1186/s13662-021-03345-5, 2021.

[12] N.H. Sweilam, S.M. AL-Mekhlafi, A. Almutairi, and D. Baleanu. A hybrid fractional COVID-19 model with general population mask use: Numerical treatments. *Alexandria Engineering Journal*, 60: 3219-3232, 2021.

[13] M.R. Mahmoudi, D. Baleanu, S.S. Band, and A. Mosavi. Factor analysis approach to classify COVID-19 datasets in several regions. *Results in Physics*, 104071, https://doi.org/10.1016/j.rinp.2021.104071, 2021.

[14] S. Yadav, D. Kumar, J. Singh, and D. Baleanu, Analysis and dynamics of fractional order Covid-19 model with memory effect. Results in Physics, 104017, https://doi.org/10.1016/j.rinp.2021.104017, 2021.

[15] A. Shadloo-Jahromi, O. Bavi, M.H. Heydari, M. Kharati-Koopaee, and Z. Avazzadeh. Dynamics of respiratory droplets carrying SARS-CoV-2 virus in closed atmosphere. *Results in Physics*, 103482, https://doi.org/10.1016/j.rinp.2020.103482, 2021.

[16] M. Maleki, and A.R. Nematollahi. Auto–Regressive models with mixture of scale mixtures of Gaussian innovations. Iranian Journal of Science and Technology, Transactions A: Science, 41(4):1099–1107, 2017.

[17] P. Zarrin, M. Maleki, Z. Khodadadi, and R.B. Arellano-Valle. Time series process based on the unrestricted skew normal process. *Journal of Statistical Computation and Simulation*, 89 (1): 38-51, 2018.

[18] A. Hajrajabi, and M. Maleki. Nonlinear semiparametric Auto–Regressive model with finite mixtures of scale mixtures of skew normal innovations. *Journal of Applied Statistics*, 46 (11): 2010-2029, 2019.

[19] S. Ghasami, Z. Khodadadi, and M. Maleki. Auto–Regressive processes with generalized hyperbolic innovations. *Communications in Statistics - Simulation and Computation*. https://doi.org/10.1080/03610918.2018.1535066, 2019.

[20] S. Ghasami, Z. Khodadadi, and M. Maleki. Leptokurtic and platykurtic class of robust symmetrical and asymmetrical time series models. *Journal of Computational and Applied Mathematics*, 112806. https://doi.org/10.1016/j.cam.2020.112806, 2020.

[21] M. Maleki, D. Wraith, M.R. Mahmoudi, and J.E. Contreras-Reyes. Asymmetric heavy-tailed vector auto-regressive processes with application to financial data. *Journal of Statistical Computation and Simulation*, 90(2): 324-340, 2020.

[22] M. Maleki, A. Hajrajabi, and R.B. Arellano-Valle. Symmetrical and asymmetrical mixture Auto–Regressive processes. *Brazilian Journal of Probability and Statistics*, 34(2); 273-290, 2020.

[23] M.R. Mahmoudi, M. Maleki, D. Baleanu, V.T. Nguyen, and K.H. Pho. A Bayesian Approach to Heavy-Tailed Finite Mixture Auto–Regressive Models. *Symmetry,* 12(6): 929, 2020.

[24] M.R. Mahmoudi, M. Maleki, K. Borodin, K.H. Pho, and D. Baleanu. On comparing and clustering the spectral densities of several almost cyclostationary processes. *Alexandria Engineering Journal*, 59: 2555-2565, 2020.

[25] M. Maleki, and M.R. Mahmoudi. Two-Piece Location-Scale Distributions based on Scale Mixtures of Normal family. *Communications in Statistics_Theory and Methods,* 46(24): 12356-12369, 2017.

[26] B. Moravveji, Z. Khodadadi, and M. Maleki. A Bayesian Analysis of Two-Piece Distributions Based on the Scale Mixtures of Normal Family. *Iranian Journal of Science and Technology, Transactions A: Science*, 43(3): 991–1001, 2019.

[27] M. Maleki, and A.R. Nematollahi. Bayesian approach to epsilon-skew-normal family. *Communications in Statistics-Theory and Methods,* 46 (15): 7546-7561, 2017.

[28] M. Maleki, Z. Barkhordar, Z. Khodadado, and D. Wraith. A robust class of homoscedastic nonlinear regression models. *Journal of Statistical Computation and Simulation,* 89(14): 2765-2781, 2019.

[29] A. Hoseinzadeh, M. Maleki, and Z. Khodadadi. Heteroscedastic nonlinear regression models using asymmetric and heavy tailed two-piece distributions. *AStA Advances in Statistical Analysis*, https://doi.org/10.1007/s10182-020-00384-3, 2020.

[30] Z. Barkhordar, M. Maleki, Z. Khodadadi, D. Wraith, and F. Negahdari. A Bayesian approach on the two-piece scale mixtures of normal homoscedastic nonlinear regression models. *Journal of Applied Statistics*, https://doi.org/10.1080/02664763.2020.1854203, 2020.

[31] D.R. Andrews, and C.L. Mallows. Scale mixture of normal distribution. *Journal of the Royal Statistical Society, Series B*, 36: 99–102, 1974.

[32] C. Liu, D.B. Rubin. The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. *Biometrika*, 81: 633-648, 1994.

[33] Brockwell, P.J., Davis, R.A. (1996). Time Series and Forecasting. Springer-Verlag New York, Inc.

[34] R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[35] Emadi, Mostafa, et al. "Predicting and mapping of soil organic carbon using machine learning algorithms in Northern Iran." Remote Sensing 12.14 (2020): 2234.

[36] Shamshirband, Shahaboddin, et al. "Developing an ANFIS-PSO model to predict mercury emissions in combustion flue gases." Mathematics 7.10 (2019): 965.

[37] Salcedo-Sanz, Sancho, et al. "Machine learning information fusion in Earth observation: A comprehensive review of methods, applications and data sources." Information Fusion 63 (2020): 256-272.

[38] Ehteram, Mohammad, et al. "Inclusive Multiple Model Using Hybrid Artificial Neural Networks for Predicting Evaporation.", Frontiers in Environmental Science 9 (2022)

[39] Samadianfard, Saeed, et al. "Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm." Energy Reports 6 (2020): 1147-1159.

[40] Taherei Ghazvinei, Pezhman, et al. "Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network." Engineering Applications of Computational Fluid Mechanics 12.1 (2018): 738-749.

[41] Mansoor, Khwaja, et al. "Securing IoT-based RFID systems: A robust authentication protocol using symmetric cryptography." Sensors 19.21 (2019): 4752.

[42] Kargar, Katayoun, et al. "Estimating longitudinal dispersion coefficient in natural streams using empirical models and machine learning algorithms." Engineering Applications of Computational Fluid Mechanics 14.1 (2020): 311-322.

[43] Qasem, Sultan Noman, et al. "Estimating daily dew point temperature using machine learning algorithms." Water 11.3 (2019): 582.

[44] Mosavi, Amirhosein, et al. "Ensemble boosting and bagging based machine learning models for groundwater potential prediction." Water Resources Management 35.1 (2021): 23-37.

[45] Seifi, Akram, et al. "Modeling and uncertainty analysis of groundwater level using six evolutionary optimization algorithms hybridized with ANFIS, SVM, and ANN." Sustainability 12.10 (2020): 4023.

[46] Mosavi, Amir, and Atieh Vaezipour. "Reactive search optimization; application to multiobjective optimization problems." Applied Mathematics 3.10A (2012): 1572-1582.

[47] Shabani, Sevda, et al. "Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis." Atmosphere 11.1 (2020): 66.

[48] Ghalandari, Mohammad, et al. "Aeromechanical optimization of first row compressor test stand blades using a hybrid machine learning model of genetic algorithm, artificial neural networks and design of experiments." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 892-904.

[49] Mosavi, Amir. "Multiple criteria decision-making preprocessing using data mining tools." arXiv preprint arXiv:1004.3258 (2010).

[50] Mahmoudi, Mohammad Reza, et al. "Principal component analysis to study the relations between the spread rates of COVID-19 in high risks countries." Alexandria Engineering Journal 60.1 (2021): 457-464.

[51] Karballaeezadeh, Nader, et al. "Prediction of remaining service life of pavement using an optimized support vector machine (case study of Semnan–Firuzkuh road)." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 188-198.

[52] Moeini, Iman, et al. "Modeling the time-dependent characteristics of perovskite solar cells." Solar Energy 170 (2018): 969-973.

[53] Shamshirband, Shahaboddin, et al. "Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines." Engineering Applications of Computational Fluid Mechanics 14.1 (2020): 805-817.

[54] Samadianfard, Saeed, et al. "Support vector regression integrated with fruit fly optimization algorithm for river flow forecasting in Lake Urmia Basin." Water 11.9 (2019): 1934.

[55] Lei, Xinxiang, et al. "GIS-based machine learning algorithms for gully erosion susceptibility mapping in a semi-arid region of Iran." Remote Sensing 12.15 (2020): 2478.

[56] Adejuwon, Adeyemi, and Amir Mosavi. "Domain driven data mining: application to business." International Journal of Computer Science Issues 7.4 (2010): 41-44.

[57] Riahi-Madvar, Hossien, et al. "Comparative analysis of soft computing techniques RBF, MLP, and ANFIS with MLR and MNLR for predicting grade-control scour hole geometry." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 529-550.

[58] Mosavi, Amir, and Mohammad Edalatifar. "A hybrid neuro-fuzzy algorithm for prediction of reference evapotranspiration." International conference on global research and education. Springer, Cham, 2018.

[59] Asadi, Esmaeil, et al. "Groundwater quality assessment for sustainable drinking and irrigation." Sustainability 12.1 (2019): 177.

[60] Mosavi, Amir, and Abdullah Bahmani. "Energy consumption prediction using machine learning; a review." (2019).

[61] Dineva, Adrienn, et al. "Review of soft computing models in design and control of rotating electrical machines." Energies 12.6 (2019): 1049.

[62] Mosavi, Amir, and Timon Rabczuk. "Learning and intelligent optimization for material design innovation." In International Conference on Learning and Intelligent Optimization, pp. 358-363. Springer, Cham, 2017.

[63] Torabi, Mehrnoosh, et al. "A hybrid machine learning approach for daily prediction of solar radiation." International Conference on Global Research and Education. Springer, Cham, 2018.

[64] Mosavi, Amirhosein, et al. "Comprehensive review of deep reinforcement learning methods and applications in economics." Mathematics 8.10 (2020): 1640.

[65] Ahmadi, Mohammad Hossein, et al. "Evaluation of electrical efficiency of photovoltaic thermal solar collector." Engineering Applications of Computational Fluid Mechanics 14.1 (2020): 545-565.

[66] Ghalandari, Mohammad, et al. "Flutter speed estimation using presented differential quadrature method formulation." Engineering Applications of Computational Fluid Mechanics 13.1 (2019): 804-810.

[67] Ijadi Maghsoodi, Abteen, et al. "Renewable energy technology selection problem using integrated h-swara-multimoora approach." Sustainability 10.12 (2018): 4481.

[68] Mohammadzadeh S, Danial, et al. "Prediction of compression index of fine-grained soils using a gene expression programming model." Infrastructures 4.2 (2019): 26.

[69] Sadeghzadeh, Milad, et al. "Prediction of thermo-physical properties of TiO2-Al2O3/water nanoparticles by using artificial neural network." Nanomaterials 10.4 (2020): 697.

[70] Choubin, Bahram, et al. "Earth fissure hazard prediction using machine learning models." Environmental research 179 (2019): 108770.