

SKEWED AUTO-REGRESSIVE PROCESS WITH EXOGENOUS INPUT VARIABLES: AN APPLICATION IN THE ADMINISTERED VACCINE DOSES ON COVID-19 SPREAD

MOHSEN MALEKI,* MOHAMMAD REZA MAHMOUDI,^{†,**,‡‡}

HAMID BIDRAM* and AMIR MOSAVI^{‡,§,¶,||,††,‡‡}

**Department of Statistics*

Faculty of Mathematics and Statistics

University of Isfahan, Isfahan 81746-73441, Iran

†Department of Statistics, Faculty of Science

Fasa University, Fasa, Fars, Iran

‡Faculty of Civil Engineering

Technische Universität Dresden

01069 Dresden, Germany

§John von Neumann Faculty of Informatics

Obuda University, 1034 Budapest, Hungary

¶Institute of Information Society, University of Public Service

1083 Budapest, Hungary

||Institute of Information Engineering

Automation and Mathematics

Slovak University of Technology in Bratislava

Bratislava, Slovakia

***mahmoudi.m.r@fasau.ac.ir*

††amir.mosavi@mailbox.tu-dresden.de

^{‡‡}Corresponding authors.

This is an Open Access article in the “Special Issue Section on Fractal AI-Based Analyses and Applications to Complex Systems: Part III”, edited by Yeliz Karaca (University of Massachusetts Medical School, USA), Dumitru Baleanu (Cankaya University, Turkey), Majaz Moonis (University of Massachusetts Medical School, USA), Yu-Dong Zhang (University of Leicester, UK) & Osvaldo Gervasi (Perugia University, Italy) published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Received October 22, 2021
 Accepted December 10, 2021
 Published May 25, 2022

Abstract

This study focuses on the prevalence of COVID-19 disease along with vaccination in the United States. We have considered the daily total infected cases of COVID-19 with total vaccinated cases as exogenous input and modeled them using light/heavy tailed auto-regressive with exogenous input model based on the innovations that belong to the flexible class of the two-piece scale mixtures of normal (TP-SMN) family. We have shown that the prediction of COVID-19 spread is affected by the rate of vaccine injection. In fact, the presence of exogenous input variables in time series models not only increases the accuracy of modeling, but also causes better and closer approximations in some issues including predictions. An Expectation-Maximization (EM) type algorithm has been considered for finding the maximum likelihood (ML) estimations of the model parameters, and modeling as well as predicting the infected numbers of COVID-19 in the presence of the vaccinated cases in the US.

Keywords: Auto-Regressive with Exogenous Inputs; COVID-19; Coronavirus; COVID-19 Vaccine; Time Series; Two-Piece Scale Mixtures.

1. INTRODUCTION

The rapid spread of the infectious COVID-19 disease has affected the whole world. Meanwhile, preparing and confronting the prevalence of COVID-19 diseases require careful plans and operations. Some mathematical modeling of the spread of COVID-19 has been considered by several mathematics authors and some of the recent researches are given as follows. The confirmed and recovered cases along with the death rate of COVID-19 have been modeled and forecast via using asymmetry time series by Maleki *et al.*^{1,2} The cumulative confirmed, recovered and death cases of COVID-19 cases for the top 16 countries have been forecast via using statistical machine learning models that have been designed 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 examined Seasonal Autoregressive Integrated Moving Average (SARIMA) model on infectious diseases. Kalantari⁶ forecast the COVID-19 using optimal singular spectrum analysis. Al-qaness *et al.*⁷ forecast the COVID-19 in Russia and Brazil using efficient artificial intelligence forecasting models. Amruta *et al.*⁸ conducted a study on the SARS-CoV-2 mediated neuroinflammation and the impact

of COVID-19 in neurological disorders. Luo *et al.*⁹ examined a novel based on Richard's traditional model and probed its application in COVID-19. Interested readers can refer to some recent papers published on the COVID-19.¹⁰⁻¹³

Time series are useful statistical models to forecast and control the datasets which are gathered by time. We can also use some flexible families 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.^{1,2,14} In this paper, we consider a wide class of distributions on the structure of Auto-Regressive model with exogenous input variables which is known as two-piece scale mixtures of normal (TP-SMN) family of distributions.^{1,2} The useful TP-SMN distributions have been used in several statistical models to show the flexibility and performance of the family.¹⁵⁻¹⁸

Maleki *et al.*¹ considered the robust and asymmetry autoregressive models based on the TP-SMN distributions to model the index of the COVID-19 spread and then, Maleki *et al.*² extended this model to the autoregressive moving average (ARMA) model and improved the results with more accurate predictions. Recently, by increasing the number of vaccinations, the spread of COVID-19 indexes has

been affected and decreased. So, it seems that the effect of vaccinations must be considered in time series modeling to have more accurate inferences and predictions. Additionally, we have extended the time series model in Maleki *et al.*¹ and added the exogenous input terms on the structure of their robust and asymmetry autoregressive models to consider the effect of vaccinations on the spread of COVID-19 indexes.

This paper is organized as follows. Some required aspects of the TP-SMN distributions are provided in Sec. 2. The TP-SMN Auto-Regressive (AR) model with time indexed covariate is given in Sec. 3. In Sec. 4, the proposed model is fitted on the infected numbers of COVID-19 in the US. along with vaccinated numbers of COVID-19 covariate. Finally, in Sec. 5, some conclusions of the work are provided.

2. THE TP-SMN DISTRIBUTIONS

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

Let the random variable X have renown SMN distributions and denoted by $X \sim \text{SMN}(\mu, \sigma, \nu)$.¹⁹ Then, X possesses the probability density function (pdf) given by

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

in which $\phi(\cdot | \mu, \sigma^2)$ is the density of $N(\mu, \sigma^2)$ distribution, $H(\cdot | \nu)$ indicates the cumulative distribution function (cdf) related to the scale mixing variable u for each TP-SMN member which is indexed by the parameters ν .

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. In addition, it covers their

symmetrical forms of the maintained distributions. A random variable Y distributed with the TP-SMN distributions denoted by $Y \sim \text{TP-SMN}(\mu, \sigma, \nu, \gamma)$ and with the pdf is given by

$$g(y | \mu, \sigma, \gamma, \nu) = \begin{cases} 2(1 - \gamma)f_{\text{SMN}}(y | \mu, \sigma(1 - \gamma), \nu), & y \leq \mu, \\ 2\gamma f_{\text{SMN}}(y | \mu, \sigma\gamma, \nu), & y > \mu, \end{cases} \quad (2)$$

where $0 < \gamma < 1$ is the skewness parameter, $f_{\text{SMN}}(\cdot | \mu, \sigma, \nu)$ is the SMN pdf given in (1).

Mean and variance of $Y \sim \text{TP-SMN}(\mu, \sigma, \nu, \gamma)$ are $E(Y) = \mu - b\sigma(1 - 2\gamma)$ and $\text{Var}(Y) = \sigma^2[c_2k_2 - b^2c_1^2]$, respectively, where $b = \sqrt{2/\pi}k_1$, $c_r = \gamma^{r+1} + (-1)^r(1 - \gamma)^{r+1}$ and k_r are given in Table 1. It is noteworthy that for $\gamma = 0.5$, the well-known SMN family (which is symmetric) has been extracted from the TP-SMN family.

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 of mixture distribution with two components and distinct supports in (3). So, let $Y \sim \text{TP-SMN}(\mu, \sigma, \nu, \gamma)$, then the pdf given in (2) is represented as

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

where $\pi = \frac{\sigma_1}{\sigma_1 + \sigma_2}\sigma_1 = \sigma(1 - \gamma)$ 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

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

Table 1 Some Main TP-SMN Distributions.

Distributions	$f_{\text{SMN}}(\cdot), y \in \mathbb{R}$	$k_r = E[U^{-r/2}]$
TP-N	$\phi(y; \mu, \sigma^2)$	—
TP-T	$\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\sqrt{\pi\nu\sigma}}(1 + \frac{d}{\nu})^{-\frac{\nu+1}{2}}; \nu = \nu$	$(\frac{\nu}{2})^{r/2} \frac{\Gamma(\frac{\nu-r}{2})}{\Gamma(\frac{\nu}{2})}, \nu > r$
TP-SL	$\nu \int_0^1 u^{\nu-1} \phi(y; \mu, u^{-1}\sigma^2) du; \nu = \nu$	$\frac{2\nu}{2\nu-r}, \nu > \frac{r}{2}$
TP-CN	$\nu\phi(y; \mu, \tau^{-1}\sigma^2) + (1 - \nu)\phi(y; \mu, \sigma^2); \nu = (\nu, \tau)^\top$	$\frac{\nu}{\tau^{r/2}} + 1 - \nu$

where $T = (-\infty, \mu)$ and $\text{SMN}(\cdot)I_T(\cdot)$ denotes the SMN-distributions which are truncated on the interval T , and the random vector $\mathbf{Z} = (Z_1, Z_2)^\top$ with the probability mass function (pmf) given by

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

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 EXOGENOUS INPUT MODEL

The auto-regressive with exogenous input variable model of order p for autoregressive and order m for exogenous input terms based on the TP-SMN innovations and indicated using TP-SMN-ARX(p, m) model is considered and expressed by

$$Y_t = \mu + \sum_{j=1}^p \alpha_j Y_{t-j} + \sum_{j=1}^m \beta_j X_{t-j} + \varepsilon_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (6)$$

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

$$\varepsilon_t \stackrel{\text{iid}}{\sim} \text{TP-SMN}(0, \sigma, \nu, \gamma), \quad t = 0, \pm 1, \pm 2, \dots \quad (7)$$

Considering the ARX model (6) with parameters $\boldsymbol{\Theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \mu, \sigma_1, \sigma_2, \nu)^\top$ to have a more appropriate form, the model can be represented by $Y_t = \boldsymbol{\alpha}^\top \mathbf{Y}_{t-1} + \boldsymbol{\beta}^\top \mathbf{X}_{t-1} + \varepsilon_t$, for which $\mathbf{Y}_{t-1} = (Y_{t-1}, \dots, Y_{t-p})^\top$ and $\mathbf{X}_{t-1} = (X_{t-1}, \dots, X_{t-m})^\top$. So, we consider the sample vector $(\mathbf{Y}^\top, \mathbf{X}^\top)^\top = (Y_{-p+1}, \dots, Y_0, X_{-m+1}, \dots, X_0, Y_1, \dots, Y_n, X_1, \dots, X_n)^\top$ with the non-stochastic parts Y_{-p+1}, \dots, Y_0 and X_{-m+1}, \dots, X_0 of the sample, and the (conditional) likelihood function related to the TP-SMN-ARX(p, m) model is given by

$$\begin{aligned} l(\boldsymbol{\Theta} | \mathbf{Y}, \mathbf{X}) &= \sum_{t=1}^n \log g_{Y_t} | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \mathbf{X}_{t-1} = \mathbf{x}_{t-1} \\ &\quad \times (y_t | \boldsymbol{\alpha}^\top \mathbf{y}_{t-1} + \boldsymbol{\beta}^\top \mathbf{x}_{t-1} + \mu, \sigma_1, \sigma_2, \nu) \end{aligned} \quad (8)$$

Because of complexities of the SMN-densities (1), finding the ML estimates related to the TP-SMN-ARX(p, m) model parameters via (8) is a difficult process. But considering an appropriate hierarchical representation of the TP-SMN distributions besides the established ARX model, obtaining the ML estimates of parameters can be derived by an EM-type scheme as follows:

Because of complexities of the SMN-densities (1), finding the ML estimates related to the TP-SMN-ARX(p, m) model parameters via (8) is considered a difficult process. However, considering an appropriate hierarchical representation of the TP-SMN distributions along with the established ARX model, obtaining the ML estimates of parameters can be derived by an EM-type scheme as follows:

Consider the complete data $\mathbf{D} = (\mathbf{Y}^\top, \mathbf{X}^\top, \mathbf{U}^\top)^\top$ with observed part of the data $(\mathbf{Y}^\top, \mathbf{X}^\top)^\top$ and the missing (latent) part of the data $\mathbf{U} = (U_1, \dots, U_n)^\top$ and $\mathbf{Z}_t = (Z_{t1}, Z_{t2})^\top; t = 1, 2, \dots, n$. Then, regarding the mixture representation (4), the pmf in (5) and the well-known stochastic representation of SMN distributions in Andrews and Mallows,¹⁹ the TP-SMN-ARX model using (6) can be represented in a hierarchical manner

$$\begin{aligned} Y_t | \mathbf{Y}_{t-1} = \mathbf{y}_{t-1}, \quad \mathbf{X}_{t-1} = \mathbf{x}_{t-1}, \quad U_t = u_t, \\ Z_{ti} = 1 \stackrel{\text{ind.}}{\sim} N(\boldsymbol{\alpha}^\top \mathbf{y}_{t-1} + \boldsymbol{\beta}^\top \mathbf{x}_{t-1} + \mu, u_t^{-1} \sigma_i^2) \\ \quad \times I_{T_t}(y_t)^{2-i} I_{T_t^c}(y_t)^{i-1} \\ U_t | Z_{ti} = 1 \stackrel{\text{ind.}}{\sim} H(u_t | \nu), \\ \mathbf{Z}_t \stackrel{\text{iid.}}{\sim} M(1, \pi, 1 - \pi), \end{aligned} \quad (9)$$

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

A special EM-type algorithm, called Expectation-Conditional Maximization Either (ECME) algorithm, was generated by Liu and Rubin,²⁰ and the hierarchical form of the TP-SMN-ARX model in (9) has been utilized 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, m) model given by

$$l(\boldsymbol{\Theta}) = -n \log(\sigma_1 + \sigma_2) - \frac{1}{2} \sum_{t=1}^n \sum_{i=1}^2 W_{ti}$$

$$\times \left(\frac{Y_t - \alpha^\top \mathbf{Y}_{t-1} - \beta^\top \mathbf{X}_{t-1} - \mu}{\sigma_i} \right)^2 + \sum_{t=1}^n \sum_{i=1}^2 Z_{ti} \log h(U_t | \nu),$$

where $\Theta = (\alpha^\top, \beta^\top, \mu, \sigma_1, \sigma_2, \nu)^\top$ and $W_{ti} = U_t Z_{ti}$ for $t = 1, 2, \dots, n$ and $i = 1, 2$. Defining the quantities $\hat{z}_{ti} = E[Z_{ti} | \hat{\Theta}, \mathbf{Y}, \mathbf{X}]$ and $\hat{w}_{ti} = E[U_t Z_{ti} | \hat{\Theta}, \mathbf{Y}, \mathbf{X}]$ and according to the properties of the conditional expectation, $\hat{w}_{ti} = \hat{\kappa}_{ti} \hat{z}_{ti}$ is obtained, where $\hat{z}_{t1} = I_{(-\infty, \hat{\alpha}^\top \mathbf{y}_{t-1} + \hat{\beta}^\top \mathbf{x}_{t-1} + \hat{\mu}]}(y_t)$ 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: $\hat{\kappa}_{ti} = \frac{\hat{\nu} + 1}{\hat{\nu} + d_{ti}}$,
- TP-SL-ARX model: $\hat{\kappa}_{ti} = \frac{2\hat{\nu} + 1}{d_{ti}} \frac{P_1(\hat{\nu} + 3/2, d_{ti}/2)}{P_1(\hat{\nu} + 1/2, d_{ti}/2)}$,
- TP-CN-ARX model: $\hat{\kappa}_{ti} = \frac{\hat{\tau}^2 \hat{\nu} e^{-\hat{\tau} d_{ti}/2} + (1 - \hat{\nu}) e^{-d_{ti}/2}}{\hat{\tau} \hat{\nu} e^{-\hat{\tau} d_{ti}/2} + (1 - \hat{\nu}) e^{-d_{ti}/2}}$,

where $d_{ti} = \left(\frac{y_t - \hat{\alpha}^\top \mathbf{y}_{t-1} - \hat{\beta}^\top \mathbf{x}_{t-1} - \hat{\mu}}{\hat{\sigma}_i} \right)^2$ and $P_x(a, b)$ denote the evaluated cdf of the Gamma (a, b) distribution at point x .

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

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

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

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

CM-step of the algorithm for coefficients

$$\hat{\alpha}^{(k+1)} = \left(\sum_{t=1}^n \hat{w}_t^{(k)} \mathbf{Y}_{t-1} \mathbf{Y}_{t-1}^\top \right)^{-1} \sum_{t=1}^n \hat{w}_t^{(k)} \times (Y_t - \hat{\beta}^{\top(k)} \mathbf{X}_{t-1} - \hat{\mu}^{(k)}) \mathbf{Y}_{t-1},$$

$$\hat{\beta}^{(k+1)} = \left(\sum_{t=1}^n \hat{w}_t^{(k)} \mathbf{X}_{t-1} \mathbf{X}_{t-1}^\top \right)^{-1} \sum_{t=1}^n \hat{w}_t^{(k)} \times (Y_t - \hat{\alpha}^{\top(k+1)} \mathbf{Y}_{t-1} - \hat{\mu}^{(k)}) \mathbf{X}_{t-1}.$$

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

$$\hat{\mu}^{(k+1)} = \frac{\sum_{t=1}^n \hat{w}_t^{(k)} (Y_t - \hat{\alpha}^{\top(k+1)} \times \mathbf{Y}_{t-1} - \hat{\beta}^{\top(k+1)} \mathbf{X}_{t-1})}{\sum_{t=1}^n \hat{w}_t^{(k)}}.$$

CM-step of the algorithm for scale parameters:

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

CML-step of the algorithm for scale mixing parameters:

$$\nu^{(k+1)} = \operatorname{argmax}_{\nu} l(\hat{\alpha}^{\top(k+1)}, \hat{\beta}^{\top(k+1)}, \hat{\mu}^{(k+1)}, \hat{\sigma}_1^{(k+1)}, \hat{\sigma}_2^{(k+1)}, \nu | \mathbf{Y}, \mathbf{X}),$$

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

A sufficient convergence rule for the algorithm iterates is satisfied when $|l(\hat{\Theta}^{(k+1)} | \mathbf{Y}, \mathbf{X}) / l(\hat{\Theta}^{(k)} | \mathbf{Y}, \mathbf{X}) - 1| \leq \varepsilon$, for the known values of the tolerance ε . In the present study, 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 from 19 January 2021 to 27 March 2021 related to the US (see Fig. 1) and vaccinated numbers of COVID-19 there and in the corresponding time (see Fig. 2) and fit the TP-SMN-ARX to the proposed datasets. The free source R software²¹ version 4.0.5 with a core i7 760 processor 2.8 GHz is used to implement the algorithms

After some suitable transformations given in Brockwell and Davis,²² the stationary data which are ready for 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, it was concluded

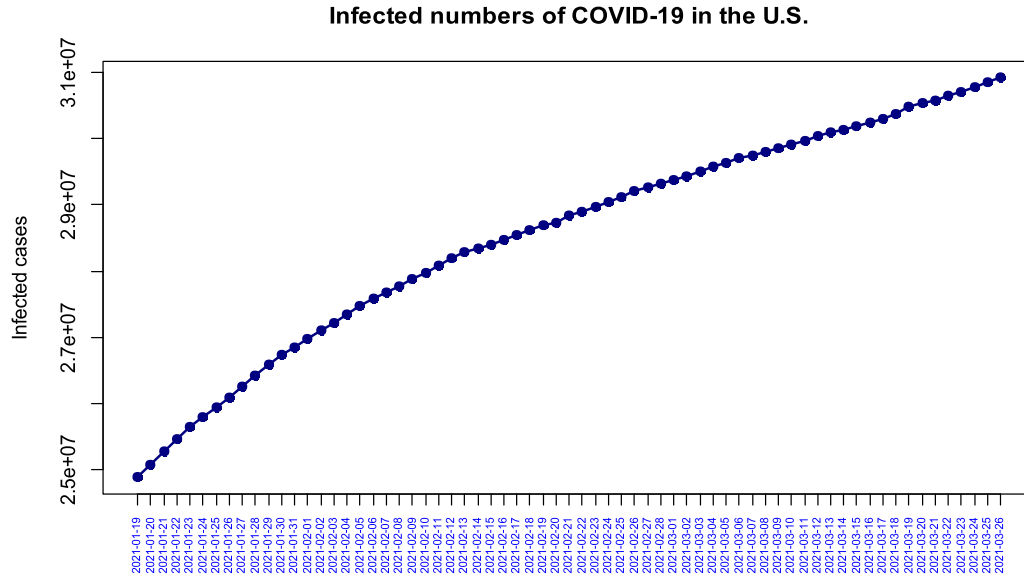


Fig. 1 Infected numbers of COVID-19 plot related to US since 19 January to 27 March 2021.

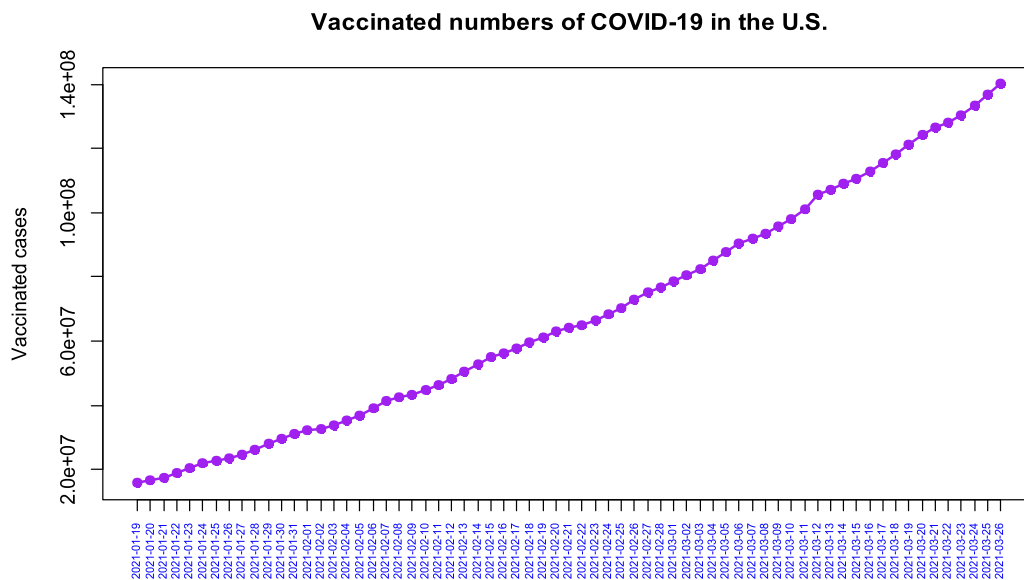


Fig. 2 Vaccinated numbers of COVID-19 plot related to the US since 19 January to 27 March in 2021.

that the best TP-T-ARX (5,1) model was fitted to the stationary series of the infected numbers of COVID-19 in the form of

TP-SMN-ARX model:

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

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

Gaussian-ARX model:

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

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

According to log-likelihood and Akaike information criterion (AIC) which are recorded in Table 2, the TP-T-ARX model is more reasonable than the ordinary Gaussian-ARX model. The auto-correlation function (ACF) diagram of estimated

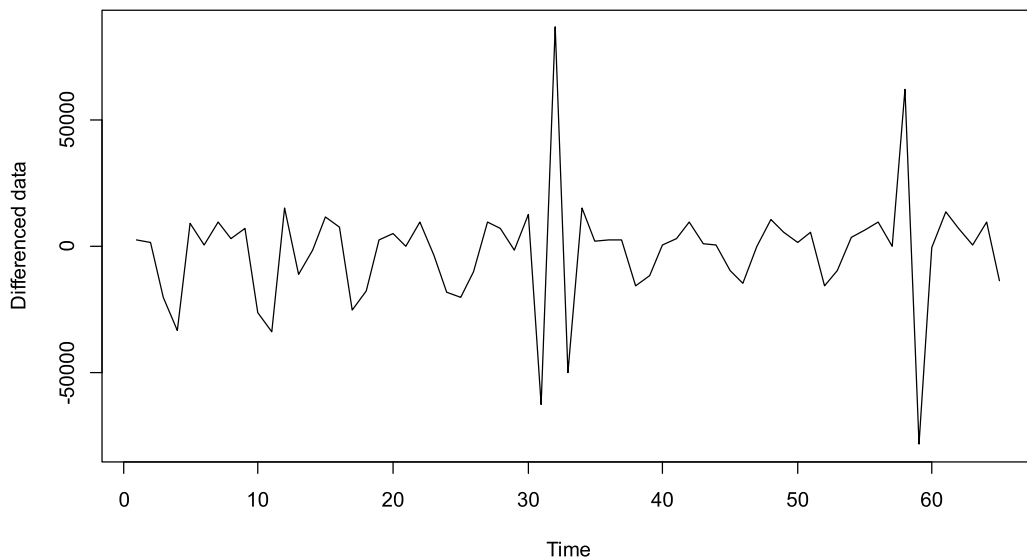


Fig. 3 Differenced transformation of the infected numbers of COVID-19 plot.

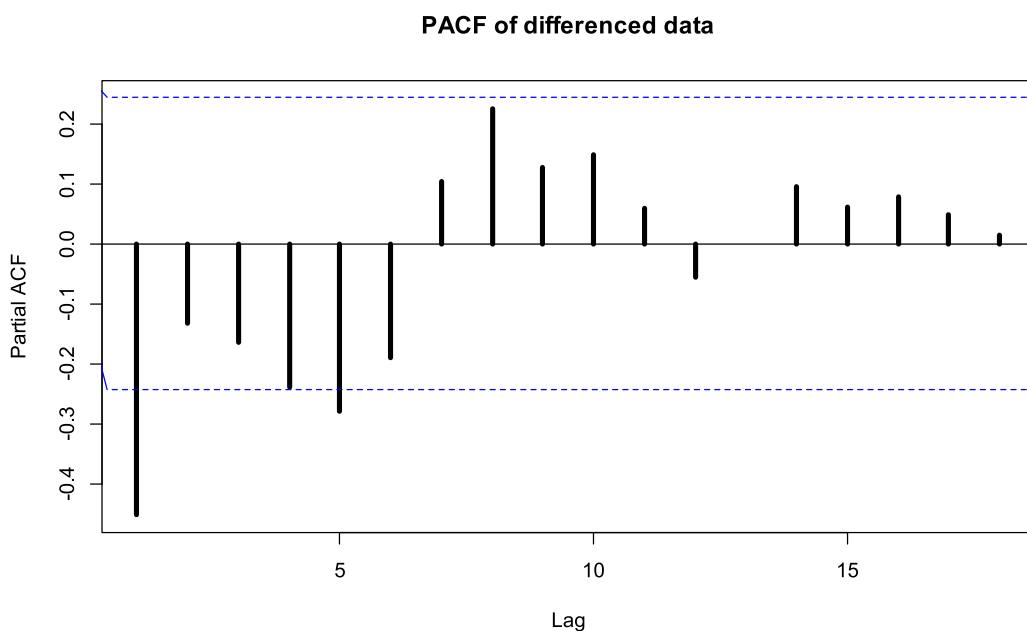


Fig. 4 PACF of the differenced data (stationary).

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
<i>P</i> -values	0.9882	0.9625	0.9613	0.8821

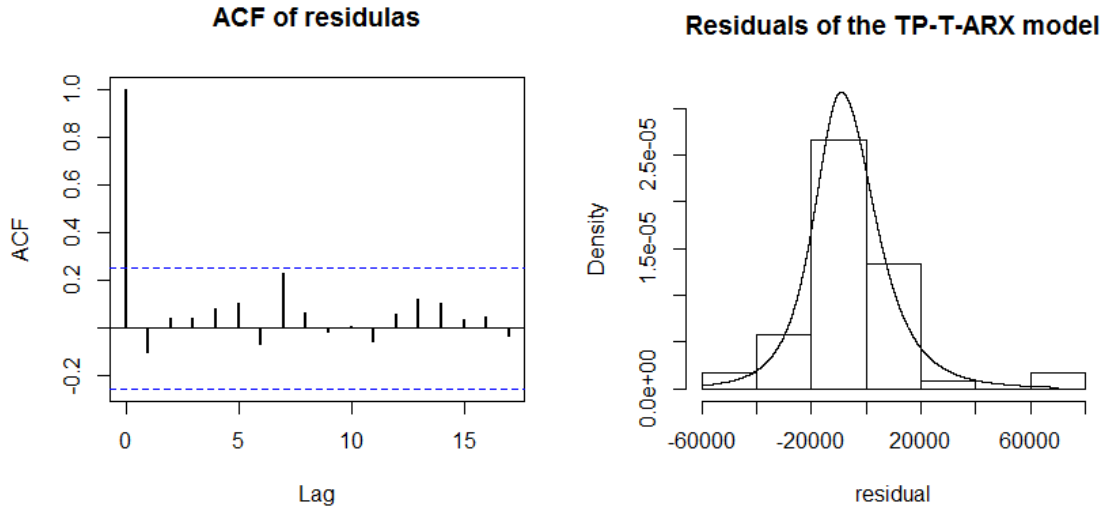


Fig. 5 Sample ACF diagram of the estimated innovations of the fitted model (Left) Histogram of the estimated innovations of the fitted model via superimposed estimated TP-T density (Right).

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 as given in Fig. 5 (right). Figure 5 shows the suitability of the proposed estimated model to the stationary series of infected numbers of COVID-19 related to the US.

To investigate the appropriateness 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 Table 2. 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 of the proposed fitted TP-SMN-ARX model. Also, comparing the corresponding P -values of the proposed tests with the ordinary Gaussian-ARX

model, a better performance of the TP-SMN-ARX model can be shown (see Table 2).

To show further reasonability of the fitted model, we predicted the 10 data of the infected cases (15 March 2021 up to 24 March 2021), and 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 shows the predicted values based on the modeling with and without covariate which superimposed the diagram of the real values of infected numbers of COVID-19 related to US Percent Relative Prediction Error (PRPE) of the predicted values via the model with covariate is 0.0762%, while it is 0.1469% for the model without covariate. So, the results and Fig. 6 show the importance of the covariate (considering the

Table 3 Real Values of the Infected Numbers of COVID-19 Related to the US Since 15 March 2021 to 24 March 2021 With and Without Vaccinated Numbers Covariate and Predictions with 95% Confidence Interval.

COVID-19 in the US	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
	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

Predicted and real values

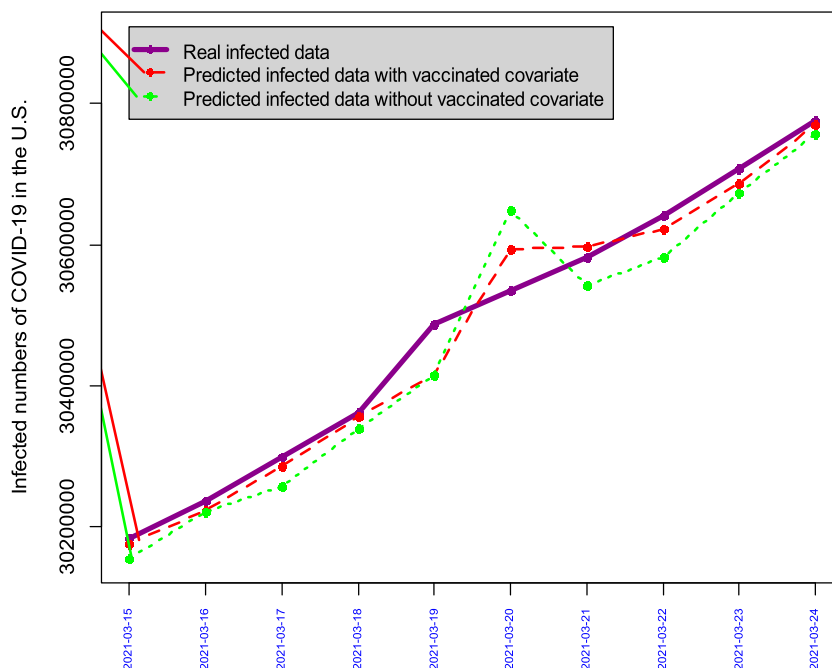


Fig. 6 The experimental values and predicted infected numbers of COVID-19 related to the US time series plots since 15 March 2021 up to 24 March 2021 with and without vaccinated numbers covariate.

vaccination) on the modeling and predictions of the infected cases.

5. CONCLUSION

Coronavirus is a wide category that affects gastrointestinal, hepatic, neurological and respiratory systems. Extended vaccination might reduce the prevalence of COVID-19 diseases. In the present methodology, considering a robust class of time series models with covariate was intended to model the infected numbers of COVID-19 in the US along with the vaccinated COVID-19 cases as the covariate. The performance of the proposed fitted model shows that it would be better to consider the vaccinated numbers of COVID-19 as the covariate in order to predict infected numbers of COVID-19. For future works, the proposed model can be extended to the mixture models to classify several regions and also add some new effective covariates to improve the model such as the mobility data.

ACKNOWLEDGMENTS

The authors would like to thank the editor and anonymous three reviewers for their constructive suggestions, corrections and encouragement,

which helped us to improve earlier versions of the manuscript.

REFERENCES

1. 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 Med. Infect. Disease* (2020) 101742, <https://doi.org/10.1016/j.tmaid.2020.101742>.
2. 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* (2020) 110151, <https://doi.org/10.1016/j.chaos.2020.110151>.
3. K. E. ArunKumar, Dinesh V. Kalaga, Ch. M. Sai Kumar, G. Chilkoor, M. Kawaji and T. 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), *Appl. Soft Comput.* (2021) 107161, <https://doi.org/10.1016/j.asoc.2021.107161>.
4. K. E. ArunKumar, D. V. Kalaga, Ch. Mohan Sai Kumar, M. Kawaji and T. 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* (2021) 110861, <https://doi.org/10.1016/j.chaos.2021.110861>.
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, *Int. J. Infect. Dis.* (2021), <https://doi.org/10.1016/j.ijid.2021.01.070>.
 6. M. Kalantari, Forecasting COVID-19 pandemic using optimal singular spectrum analysis, *Chaos Solitons Fractals* (2021) 110547, <https://doi.org/10.1016/j.chaos.2020.110547>.
 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 Environ. Protect.* (2021), <https://doi.org/10.1016/j.psep.2020.11.007>.
 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 Rev.* **58** (2021) 1–15.
 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* (2020) 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, *Adv. Differ. Eq.* (2021) 167, <https://doi.org/10.1186/s13662-021-03317-9>.
 11. A. S. Alshomrani, M. Z. Ullah and D. Baleanu, Caputo SIR model for COVID-19 under optimized fractional order, *Adv. Differ. Eq.* **185** (2021), <https://doi.org/10.1186/s13662-021-03345-5>.
 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 Eng. J.* **60** (2020) 3219–3232.
 13. S. Yadav, D. Kumar, J. Singh and D. Baleanu, Analysis and dynamics of fractional order Covid-19 model with memory effect, *Results Phys.* (2021) 104017, <https://doi.org/10.1016/j.rinp.2021.104017>.
 14. A. Hajrajabi and M. Maleki, Nonlinear semiparametric Auto-Regressive model with finite mixtures of scale mixtures of skew normal innovations, *J. Appl. Stat.* **46**(11) (2019) 2010–2029.
 15. A. Hoseinzadeh, M. Maleki and Z. Khodadadi, Heteroscedastic nonlinear regression models using asymmetric and heavy tailed two-piece distributions, *AStA Adv. Stat. Anal.* (2020), <https://doi.org/10.1007/s10182-020-00384-3>.
 16. 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, *J. Appl. Stat.* (2020), <https://doi.org/10.1080/02664763.2020.1854203>.
 17. M. Maleki, G. J. McLachlan and S. X. Lee, Robust clustering based on finite mixture of multivariate fragmental distributions, *Stat. Modelling* (2021), <https://doi.org/10.1177/1471082X211048660>.
 18. M. Bazrafkan, K. Zare, M. Maleki and K. Zaha, Partially linear models based on heavy-tailed and asymmetrical distributions, *Stochastic Environ. Res. Risk Assessment* (2021), <https://doi.org/10.1007/s00477-021-02101-1>.
 19. D. R. Andrews and C. L. Mallows, Scale mixture of normal distribution, *J. R. Stat. Soc. B* **36** (1974) 99–102.
 20. C. Liu and D. B. Rubin, The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence, *Biometrika* **81** (1994) 633–648.
 21. R Core Team, R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria (2021). <https://www.R-project.org/>.
 22. P. J. Brockwell and R. A. Davis, *Time Series and Forecasting* (Springer-Verlag, New York, 1996).