Journal of Data Science and Modeling, Vol. 2, No. 1, 79-93, December 2023

Research Manuscript

The vector autoregressive model with asymmetric shocks for the new cases and the new deaths of COVID-19 in Iran

Manijeh Mahmoodi [1], Mohammad Reza Salehi Rad*[2], Farzad Eskandari[3]

[1]Ph.D student, Department of Statistics, Allameh Tabataba'i University, Tehran, Iran.

[2]Department of Statistics, Allameh Tabataba'i University, Tehran, Iran.

[3]Department of Statistics,Allameh Tabataba'i University, Tehran, Iran.

Received: 16/03/2024 Accepted: 12/07/2024

Abstract: The novel coronavirus (COVID-19) spread quickly from person to person. One of the basic aspects of country management has been to prevent the spread of this disease. Therefore, its prediction is very important. In such matters, the estimation of new cases and new deaths of COVID-19 has been widely considered by researchers. The new cases and new deaths of COVID-19 affect each other. For this reason, the autoregressive (AR) model doesn't provide the necessary efficiency, and vector autoregressive (VAR) models are used instead. Data were taken from the World Health Organization (WHO) from March 2020 until March 2023 in Iran. VAR models are usually considered with Gaussian shocks. However, in this study, the shocks (noises) have an asymmetric distribution, which has not been considered in previous research. In this article, we estimate the parameters of the VAR model of order 1 with the multivariate skew normal (MSN) distribution for the asymmetric shocks using the expectation conditional maximization (ECM) algorithm. We then apply the proposed model to the weekly data of the new cases and new deaths of COVID-19. We show that the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) in the proposed model with the MSN distribution for shocks are lower than those in the model with the multivariate normal (MN) distribution for shocks. We predict the new cases and new deaths for four weeks and calculate the RMSE and MAPE of this model.

Keywords: COVID19, ECM algorithm, Maximum likelihood estimation, Multivariate skew normal, Skewness, Vector autoregressive

Mathematics Subject Classification (2010): 62M10, 62H12.

*Corresponding Author: salelehirad@ atu.ac.ir

1. Introduction

The coronavirus was first detected in Wuhan, China, on December 19, 2019, and spread from bats to humans (Jahangir and Muheem , 2020). The World Health Organization (WHO) later declared this disease a pandemic (Arfan and et al. , 2021). The coronavirus spread significantly with new strains worldwide. Although restrictions were applied to control it, predicting infected patients, deaths, and recoveries remains very important. The emergence of new virus strains indicates that this disease is always with us, and its pandemic represents an unprecedented event in global health. Therefore, modeling has become important for related variables in this disease, and time series modeling has been applied. According to WHO guidelines, new deaths refer to the number of people who died with COVID-19, while new cases refer to the number of people who were infected with COVID-19 and tested positive.

Box and Jenkins (1970) introduced classical time series models, but due to the influence of many factors on the main variable, these classical models are often not used. Tsay (2002), Lütkepohl (2005), and Kilian and Lütkepohl (2017) discussed vector time series models, with one of the most important being the vector autoregressive model. A lot of research has been conducted in predicting COVID-19, with varying results based on the models and data used.

Time series analysis and forecasting of COVID-19 in India using genetic programming were discussed by Salgotra and et al. (2020). They utilized genetic programming (GP), an enhanced version of the genetic algorithm (GA), in which new solutions are generated as computer-based programs rather than simple binary strings. Two major time series, including confirmed cases (CC) and death count (DC), were analyzed to predict the possible impact of COVID-19. Adiga and et al. (2020) presented mathematical models for COVID-19 and showed that these models differ in their use and mathematical form. Modeling and forecasting the spread and death rate of coronavirus globally using ARMA models based on two-piece scale mixtures of the normal distribution for errors were discussed by Maleki and et al. (2020). Sahai and et al. (2020) presented the ARIMA model for forecasting the epidemic spread in five countries. Talkhi and et al. (2021) applied nine models, including Hybrid, NNETAR, ARIMA, Holt-Winter, etc., to find the best model for forecasting the number of confirmed and death cases in Iran. Petropoulos and et al. (2022) forecasted the confirmed cases and deaths of COVID-19 using a simple time series model.

Busari and Sasmon (2022) compared forecasts of new COVID-19 cases using regression, ARIMA, and machine learning models in Nigeria. The results revealed that, in terms of forecasting performance, the inverse regression model outperformed other regression methods, and ARIMA (4,1,4) outperformed other

ARIMA models. Abonazel and Darwish (2022) predicted recovered cases and the number of deaths in Egypt after virus mutation using the ARIMA model. Ebrahimoghli and et al. (2023) conducted interrupted time series analysis using a generalized least squares regression model to estimate excess mortality after the COVID-19 pandemic. Awajan and et al. (2024) discussed the enhanced hybrid empirical mode decomposition with the autoregressive integrated moving average method, applying it to forecast daily new COVID-19 reported cases in Jordan.

Khan and et al. (2020) forecasted new cases, new deaths, and recoveries of COVID-19 using the VAR model in Pakistan. Sujath and et al. (2020) presented a model for predicting COVID-19 based on the number of recoveries and deaths in India using machine learning and the VAR model. Nguyen and et al. (2021) developed a multivariate time series model based on local infection incidence, showing that the VECM model had very good 7-day-ahead forecast performance and outperformed the traditional ARIMA model for COVID-19. Meimela and et al. (2021) used a vector autoregressive integrated moving average (VARIMA) model for the new cases and new deaths due to COVID-19 in Indonesia. Rajab and et al. (2022) proposed an approach to forecasting the spread of the pandemic based on the vector autoregressive model and applied it to forecast the number of new cases and deaths in the UAE, Saudi Arabia, and Kuwait.

The number of new cases and deaths in COVID-19 are related (Rajab et al. 2022). As described above, it is crucial to model the relationship between these variables using the vector autoregressive (VAR) model in multivariate time series. Usually, the multivariate normal distribution is used for the shocks in VAR models. However, in economic, financial, and medical problems, factors can cause skewness (asymmetry) in shocks, leading to non-symmetric distributions. In such cases, the normal distribution is not suitable for shocks, posing a significant modeling challenge. For this reason, the multivariate skew distribution should be used for shocks, which has not been considered in previous research. The family of multivariate skew distributions is large and complex, making it challenging to work with. Therefore, this article considers the VAR model of order one with the multivariate skew normal (MSN) distribution for shocks. Our aim is to estimate the parameters of the VAR(1) model for the weekly number of new cases and new deaths of COVID-19 in Iran, assuming that shocks are asymmetric. The parameters of this model will be estimated using the MSN distribution for shocks, and the datasets will be forecasted. We will demonstrate that the proposed model is more suitable than the VAR model with the multivariate normal distribution for shocks. Our goal is to introduce a model that can be used for forecasting COVID-19 when shocks are asymmetric.

The paper is organized as follows: In Section 2, we review the vector autoregressive model and the estimation of parameters with the multivariate skew normal distribution for shocks using the expectation conditional maximization (ECM) algorithm. In Section 3, we consider the number of new cases and new deaths of COVID-19 in Iran, fit the proposed model to the data, evaluate the fitted model, and forecast the data. In Section 4, we present the conclusion.

2. Methods

In this section, we review the vector autoregressive model of order p, as described by Lütkepohl (2005):

$$VAR(p) : \mathcal{Y}_t = A_1\mathcal{Y}_{t-1} + A_2\mathcal{Y}_{t-2} + \cdots + A_P\mathcal{Y}_{t-p} + \mathcal{E}_t \tag{2.1}$$

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{y}_{1t} \\ \mathbf{y}_{2t} \\ \vdots \\ \mathbf{y}_{kt} \end{bmatrix}, \quad A_i = \begin{bmatrix} a_{11,i} & \cdots & a_{1k,i} \\ \vdots & \ldots & \vdots \\ a_{k1,i} & \cdots & a_{kk,i} \end{bmatrix}, \mathcal{E}_t = \begin{bmatrix} \mathcal{E}_{1t} \\ \vdots \\ \mathcal{E}_{kt} \end{bmatrix}, \quad i = 1, 2, \cdots, p$$

where $\mathcal{E}_t$ is typically assumed to follow a multivariate normal (MN) distribution with a nonsingular covariance matrix $\mathbf{\Sigma}$. Lütkepohl (2005) discussed the maximum likelihood (ML) estimator of the parameters under the MN distribution for the shocks. In this paper, we consider VAR(1), and the log-likelihood function in VAR(1) with $\mathcal{E}_t \sim MN(\boldsymbol{o}, \mathbf{\Sigma})$ is defined below:

$$\ell\left(\mathbf{\Theta}|\boldsymbol{y}\right) = Ln\ L\left(\mathbf{\Theta}|\boldsymbol{y}\right) \propto -\frac{T}{2}Ln|\mathbf{\Sigma}| - \frac{1}{2}\sum_{t=1}^{T}\left(\mathcal{Y}_t - A_1\mathcal{Y}_{t-1}\right)'\mathbf{\Sigma}^{-1}\left(\mathcal{Y}_t - A_1\mathcal{Y}_{t-1}\right) \tag{2.2}$$

where $\mathbf{\Theta} = (A_1, \mathbf{\Sigma})$. The ML estimators of the parameters are obtained by setting the first-order partial derivatives to zero.

$$\frac{\partial \ell(\mathbf{\Theta}|\boldsymbol{y})}{\partial A_1} = o \quad , \quad \frac{\partial \ell(\mathbf{\Theta}|\boldsymbol{y})}{\partial \mathbf{\Sigma}} = 0$$

Sometimes, real-world shocks exhibit skewness, necessitating the use of asymmetric distributions. Since the family of multivariate skew distributions is extensive, we consider the multivariate skew normal (MSN) distribution for the shocks and denote $\mathcal{E}_t \sim MSN(\mathbf{0}, \mathbf{\Sigma}, \boldsymbol{S})$, where $\mathbf{0}$ is the location vector,$\mathbf{\Sigma}$ is the scatter matrix (scale parameter), and $\boldsymbol{S}$ is the diagonal matrix (skew parameter). Its density function is given by:

$$f(\mathcal{E}_t) = 2^k\phi_k(o, \Omega)\Phi(S'\Omega^{-1}\mathcal{E}_t; o, \Delta) \tag{2.3}$$

where $\Omega = \Sigma + SS'$, $\Delta = (I + S'\Sigma^{-1}S)^{-1} = I - S'\Omega^{-1}S$, $S = diag(s_1, s_2, \cdots, s_k)$ and $\phi_k(.,.))$ denotes density function of multivariate normal, and $\Phi(.,.)$ denotes the standard cumulative distribution function (Gupta and Chang (2002); Sahu and et al. (2003); Tsung (2009)). If S is a zero matrix, then $\mathcal{E}_t$ has the multivariate normal distribution. log-likelihood function using T independent observation is:

$$\begin{aligned}\ell(A_1, \cdots, A_p, \Omega, S) &= kT\ Ln2\pi + \sum_{t=1}^{T} \phi_k(\mathcal{Y}_t - A_1\mathcal{Y}_{t-1}; \Omega) \\ &+ \sum_{t=1}^{T} Ln\Phi(S'\Omega^{-1}(\mathcal{Y}_t - A_1\mathcal{Y}_{t-1})) \end{aligned} \tag{2.4}$$

To find the maximum likelihood estimation for the parameters, derivatives of (2.4) are needed, which do not have closed forms. Therefore, they must be approximated by numerical methods. The ML estimators of the parameters are obtained using the Expectation Conditional Maximization (ECM) algorithm. Based on the hierarchical form of the multivariate skew normal distribution (Arellano-valle and Genton (2005); Tsung (2009)), we have:

$$\begin{aligned}\mathcal{E}_t|H_t &= h_t \sim MN_k(Sh_t, \Sigma) \\ H_t &\sim TN_k(o, I)\ I(\Re^k_+) \\ H_t|\mathcal{E}_t &\sim TN_k(S'\Omega^{-1}\mathcal{E}_t; \Delta, \Re^k_+),\ \Re^k_+ = (0, +\infty)\end{aligned}$$

where H_t has a truncated normal (TN) distribution. The logarithm of the conditional likelihood function in the $VAR(1)$ model can be written as:

$$\begin{aligned}Ln\ L(\Theta|\mathcal{E}, H) &= \ell(\Theta|\mathcal{E}, H) \\ &\propto -\frac{T}{2}Ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T}(\mathcal{E}_t - Sh_t)'\Sigma^{-1}(\mathcal{E}_t - Sh_t) - \frac{1}{2}\sum_{t=1}^{T} h_t'h_t \\ &\propto -\frac{1}{2}\sum_{t=1}^{T}(\mathcal{Y}_t - A_1\mathcal{Y}_{t-1} - Sh_t)'\Sigma^{-1}(\mathcal{Y}_t - A_1\mathcal{Y}_{t-1} - Sh_t) \\ &- \frac{T}{2}Ln|\Sigma| - \frac{1}{2}\sum_{t=1}^{T} h_t'h_t \end{aligned} \tag{2.5}$$

where $\Theta = (A_1, \Sigma, S)$.
The expectation of the logarithm of the conditional likelihood is denoted by $Q\left(\mathbf{\Theta}|\ \widehat{\mathbf{\Theta}}\right) = E\left(\ell\left(\mathbf{\Theta}|y\right)\right) \equiv E\left(\ell\left(\mathbf{\Theta}|\mathcal{E}\right)\right)$, and the steps for the maximizing $Q\left(\mathbf{\Theta}|\ \widehat{\mathbf{\Theta}}\right)$ are as below:

- Step One: Assuming no skewness, estimate the initial values for the coefficients and scale parameters.

- Step Two: Estimate the coefficients of the model:

$$\hat{A}_1^{(k+1)} = [\sum_{t=1}^{T} \left(\mathcal{Y}_t - \hat{S}^{(k)} \hat{E}\left(H_t|\mathcal{E}_t\right)^{(k)} \right) \mathcal{Y}'_{t-1}][\sum_{t=1}^{T} \mathcal{Y}_{t-1}\mathcal{Y}'_{t-1}] \tag{2.6}$$

where:

$$\begin{aligned} h_t^{(k)}|\mathcal{E}_t &\sim TN\left(\widehat{\boldsymbol{S}}^{(k)}\widehat{\boldsymbol{\Omega}}^{(k)}\mathcal{E}_t, \widehat{\boldsymbol{\Delta}}^{(k)}\right) \\ \widehat{\boldsymbol{\Omega}}^{(k)} &= \widehat{\boldsymbol{\Sigma}}^{(k)} + \left(\widehat{\boldsymbol{S}}^{(k)}\right)\left(\widehat{\boldsymbol{S}}^{(k)}\right)' \\ \widehat{\boldsymbol{\Delta}}^{(k)} &= I - \left(\widehat{\boldsymbol{S}}^{(k)}\right)'\left(\widehat{\boldsymbol{\Omega}}^{(k)}\right)^{-1}\widehat{\boldsymbol{S}}^{(k)} \end{aligned}$$

- Step three: Estimating of the skew parameter

$$\widehat{S}^{(k+1)} = diag\left(\left[\sum_{t=1}^{T}\left(\mathcal{Y}_t - \hat{A}_1^{(k)}\mathcal{Y}_{t-1}\right)\left(\hat{E}\left(H_t|\mathcal{E}_t\right)^{(k)}\right)'\right]\left[\sum_{t=1}^{T}\hat{E}\left(H_tH'_t|\mathcal{E}_t\right)^{(k)}\right]^{-1}\right) \tag{2.7}$$

- Step four: Estimating of the scale parameter

$$\begin{aligned} \widehat{\boldsymbol{\Sigma}}^{(k+1)} &= \frac{1}{T}\left[\sum_{t=1}^{T}\left(\mathcal{Y}_t - \hat{A}_1^{(k)}\mathcal{Y}_{t-1}\right)\left(\mathcal{Y}_t - \hat{A}_1^{(k)}\mathcal{Y}_{t-1}\right)'\right. \\ &- \sum_{t=1}^{T}\left(\mathcal{Y}_t - \hat{A}_1^{(k)}\mathcal{Y}_{t-1}\right)\left(\hat{E}'\left(H_t|\mathcal{E}_t\right)^{(k)}\right)\widehat{\boldsymbol{S}'}^{(k)} \\ &- \left(\sum_{t=1}^{T}\left(\mathcal{Y}_t - \hat{A}_1^{(k)}\mathcal{Y}_{t-1}\right)\hat{E}\left(H_t|\mathcal{E}_t\right)^{(k)}\widehat{\boldsymbol{S}'}^{(k)}\right)' \\ &+ \left.\sum_{t=1}^{T}\widehat{\boldsymbol{S}}^{(k)}\hat{E}\left(H_tH'_t|\mathcal{E}_t\right)^{(k)}\widehat{\boldsymbol{S}'}^{(k)}\right] \end{aligned} \tag{2.8}$$

- Step five: Repeat steps 2 to 5 until the convergence condition of the algorithm is established, i.e.:

$$\left|\frac{\ell\left(\widehat{\boldsymbol{\Theta}}^{(k+1)}|\mathbf{y}\right)}{\ell\left(\widehat{\boldsymbol{\Theta}}^{(k)}|\mathbf{y}\right)} - 1\right| \leq \mathcal{E}$$

3. Data and Results

In this paper, we considered the data of new cases and new deaths in COVID-19. It appears that new cases and new deaths are related in COVID-19 (Rajab and et al. , 2022). The weekly data on new cases and new deaths were obtained from the World Health Organization (WHO), https://www.who.int/countries/iran. The selected period in this study is from 21 March 2020 to 31 March 2023 (146 weeks). The plots of the time series of new cases and new deaths for these data are shown in Figure 1, using R software. Based on this figure and Table 1, it is known that the fourth week in August had the maximum new cases and new deaths, with numbers 344,228 and 6,341, respectively.

Table 1: The variable descriptive statistics

Variable	Min	Max	Median	Standard deviation
The new cases	289	344228	22311	6541.69
The new deaths	7	6341	625.5	1096.066

Table 2 shows the estimation of the parameters in the first-order vector autoregressive model with MN distribution for shocks.

Table 2: The estimated parameters of VAR(1) with MN distribution for shocks (signif.* 0.05)

Estimate	std.error	t value	pr	
$\hat{a}_{11}$	1.017133915	0.05360655	18.97406	6.893762e-41 $*$
$\hat{a}_{12}$	-4.8086139	3.02479973	-1.58973	1.141042e-01
$\hat{a}_{21}$	0.002954897	0.0008693051	3.399148	8.761237e-04 $*$
$\hat{a}_{22}$	0.8036608	0.0490513485	16.384070	1.254416e-34 $*$

The plots of the fit and the residuals for new cases and new deaths with ACF plots are shown in Figure 2.

The stationary condition of the VAR process is $|I - A_1 x| \neq 0$ for $|x| \geq 1$. Thus, all the eigenvalues of matrix A1 have modulus less than 1 (Lütkepohl, 2005). The eigenvalues of A1 in VAR(1) are $z_1 = 0.9103973 + 0.0530684i$ and $z_2 = 0.9103973 - 0.0530684i$, and the modulus of both is $|z_1| = |z_2| = 0.83147$, which is less than 1, indicating stationarity. The Mardia measures of the multivariate skewness and kurtosis for shocks are computed ((Mardia , 1974), (Balakrishnan and Scarpa , 2012)). The Mardia's multivariate skewness statistic is 5.2927 with a zero p-value, indicating skewness in the shocks. The kurtosis statistic is 28.261 with a 0.96 p-value, showing no kurtosis in the shocks. Therefore, we estimate the parameters of the VAR(1) model with the MSN distribution for the shocks in Table 3.

According to the Akaike information criterion, defined as $AIC = 2k - 2\ell(\widehat{\Theta}|Y)$ (Akaike , 1974), and the Bayesian information criterion,defined as $BIC = k\, log(n) -$

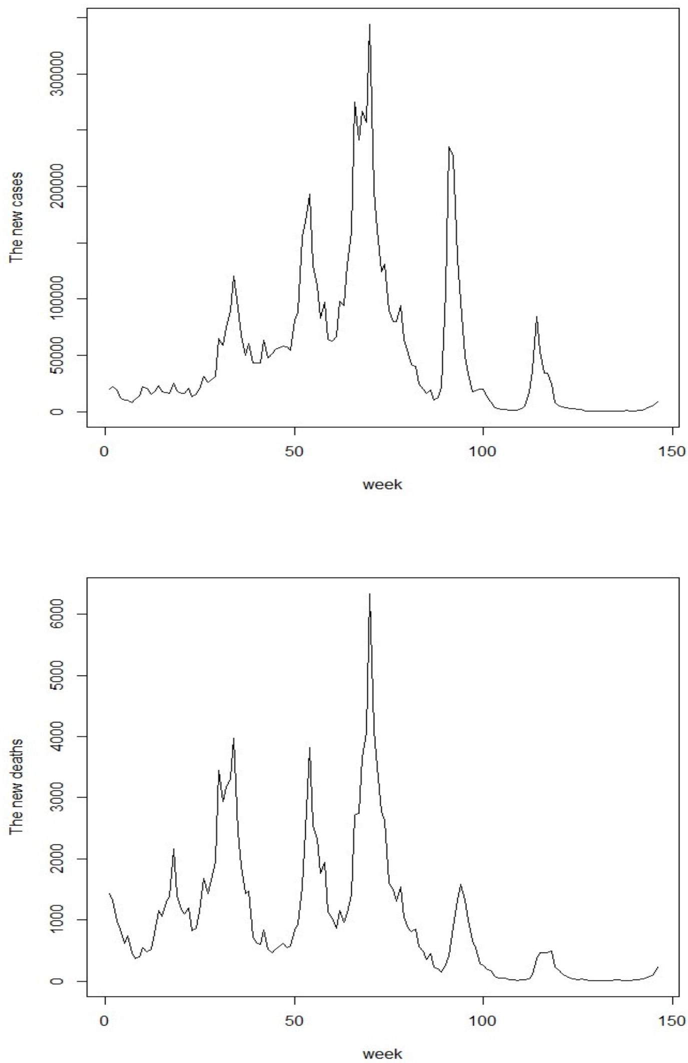

Figure 1: Time series plots of the new cases in and the new deaths,respectively.

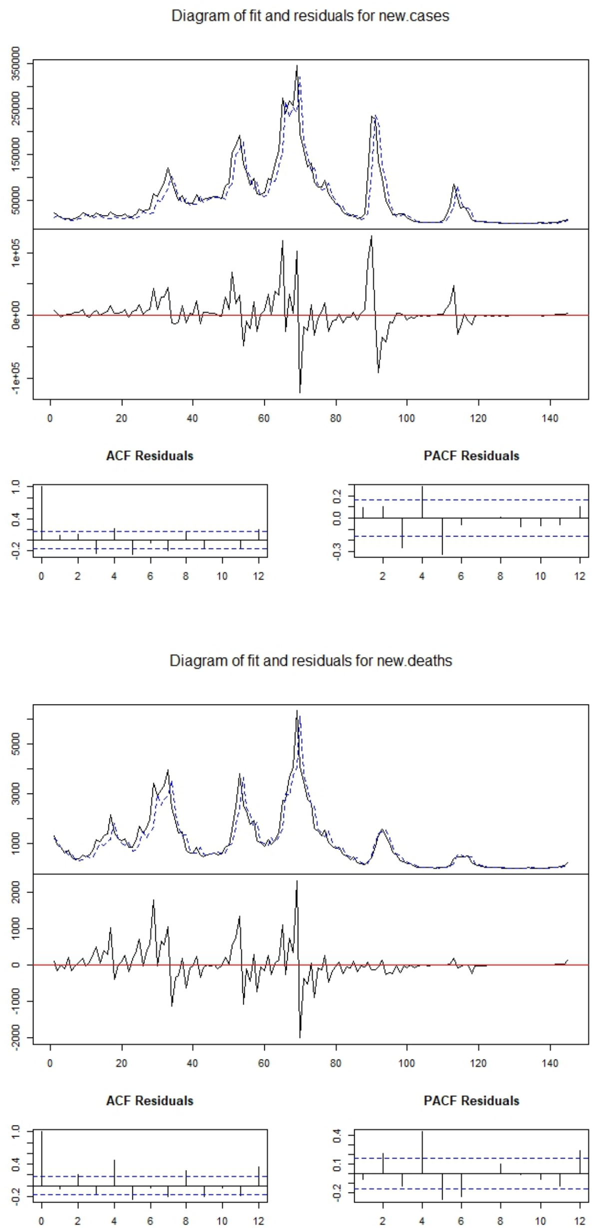

Figure 2: The plots of the fit and the residuals for the new cases and for the new deaths,respectively.

Table 3: The estimated parameters of VAR(1) with MN and MSN distribution for shocks

Estimation	$\mathcal{E}_t \sim MN2$	$\mathcal{E}_t \sim MSN2$
$\hat{a}_{11}$	1.017133915	1.000891335
$\hat{a}_{12}$	-4.8086139	-7.3870142
$\hat{a}_{21}$	0.002954897	0.002818799
$\hat{a}_{22}$	0.8036608	0.7820562
$\hat{s}_1$	-	4651.33417
$\hat{s}_1$	-	38.94252
$\hat{\sigma}_{11}$	728996760	697742157
$\hat{\sigma}_{22}$	194431.9	192267.2
$\hat{\sigma}_{12}$	8087314.9	7827069.2

$2\ell(\widehat{\Theta}|Y)$ (Schwarz , 1978), as well as the log-likelihood for comparing the VAR(1) model with the assumptions of the MN and MSN distributions for the shocks in Table 2, the model with the MSN distribution for the shocks is more appropriate than the model with the MN distribution for the shocks.

Table 4: The AIC, BIC and Log-likelihood of models

distributions	AIC	BIC	-LogLike
$\mathcal{E}_t \sim MN2$	5467.122	5473.075	2731.561
$\mathcal{E}_t \sim MSN2$	4951.271	4957.224	2473.635

The first-order vector autoregressive model was fitted with the assumption of the multivariate skew normal (MSN) distribution for the shocks, and the fitted model can be written as below:

$$\begin{aligned} y_{1t} &= 1.000891335 y_{1,t-1} - 7.3870142 y_{2,t-1} + \epsilon_{1t} \\ y_{2t} &= 0.002818799 y_{2,t-1} + 0.7820562 y_{2,t-1} + \epsilon_{2t} \end{aligned} \tag{3.9}$$

Here y_1 and y_2 represent the new cases and new deaths of COVID-19, respectively. Forecasting is one of the main objectives of multivariate time series analysis. In the context of VAR models, forecasts that minimize the mean square errors (MSE) are the most widely used (Lütkepohl, 2005). The h-step predictor of the VAR(p) model is:

$$\begin{aligned} E_t\left(\mathcal{Y}_{t+h}\right) &= E\left(\left(A_1\mathcal{Y}_{t+h-1} + A_2\mathcal{Y}_{t+h-2} + \cdots + A_p\mathcal{Y}_{t+h-p} + \mathcal{E}_{t+h}\right) \mid \{\mathcal{Y}_s | s \leq t\}\right) \\ &= A_1 E_t\left(\mathcal{Y}_{t+h-1}\right) + A_2 E_t\left(\mathcal{Y}_{t+h-2}\right) + \cdots + A_p E_t\left(\mathcal{Y}_{t+h-p}\right) + E_t\left(\mathcal{E}_{t+h}\right) \end{aligned} \tag{3.10}$$

The four-week predictor results of the new cases and new deaths in March 2023 are presented in Figures 3 and 4, respectively.

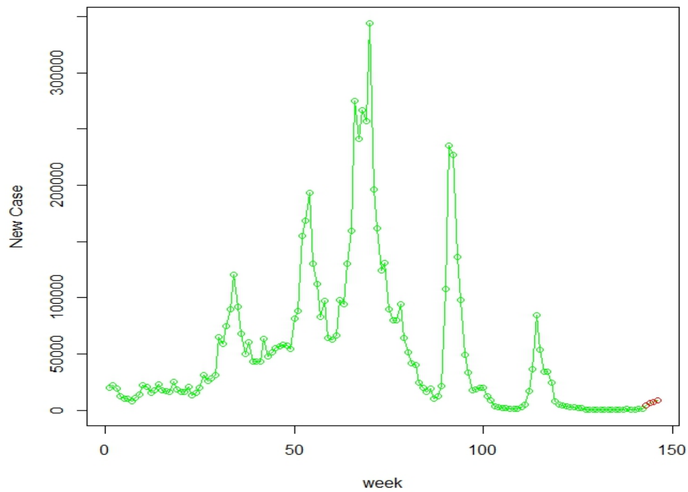

Figure 3: The time series plots of the actual new cases with green points and the predicted cases with red points.

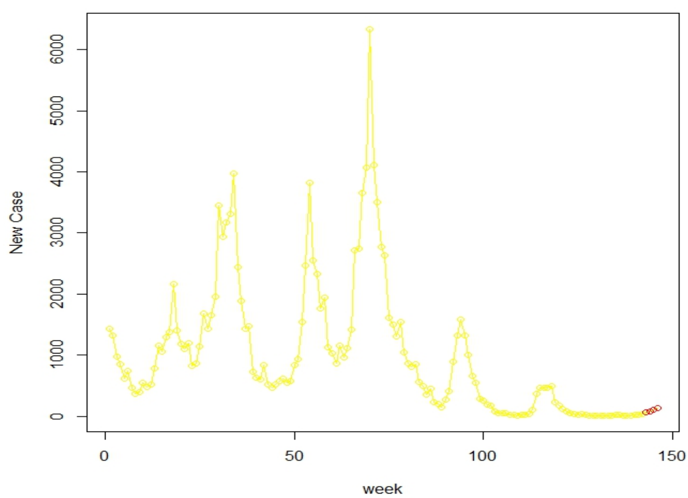

Figure 4: The time series plots of the actual new deaths with yellow points and the predicted deaths with red points.

The values of root mean square prediction error (RMSPE) and mean absolute prediction error (MAPE) are shown in Table 3.

$$RMSPE = \sqrt{\frac{\sum_{t=1}^{n}(Y_t - P_t)^2}{n}} \quad , \quad MAPE = \frac{1}{n}\sum_{t=1}^{n} |\frac{Y_t - P_t}{Y_t}|$$

Where Y_t is the actual value, and P_t is the predicted value (Chang and et al , 2007) According to the MAPE for forecasting in Table 3, the MAPE of the VAR

Table 5: The RMSE and MAPE in VAR(1) model

Variable	RMSE	MAPE
New cases	1464.096	0.3163272
New deaths	47.30935	0.1428171

model with MSN for the shocks is between 10% to 50%, indicating that this model has reasonable forecasting ability (Chang and et al , 2007).

4. Conclusion

In this paper, we estimated the model for the new cases and new deaths from COVID-19 in Iran using multivariate time series analysis. Mardia's multivariate skewness statistic was 5.29 with a zero p-value, indicating that the shocks did not follow a Gaussian distribution in this context. Therefore, we considered the first-order vector autoregressive (VAR) model with the multivariate skew normal (MSN) distribution for the shocks and used the maximum likelihood method to estimate the model parameters. Since closed forms were not obtained for parameter estimation, we used the ECM algorithm method.

We compared the results of the VAR(1) model with the MN and MSN distributions for the shocks and showed that the VAR model with the MSN distribution assumption is more appropriate, with AIC = 4951.271 and BIC = 4957.224. This model can be used to predict new cases and new deaths. We then predicted the new cases and new deaths for four weeks and obtained the RMSE and MAPE for this model. The MAPE values were 0.316 and 0.143, respectively, suggesting that the proposed model can be useful for forecasting COVID-19 trends, aiding policymakers in the health sector, as new strains of COVID-19 continue to spread, and healthcare efforts must be maintained.

Future research could explore VAR models of order $p > 1$ and other models, such as Lasso VAR with MSN or other asymmetric distributions for shocks. The proposed approach has great potential to be a valuable tool in addressing problems in medicine and economics.

References

Abonazel, M.R., and Darwish, M.N. (2022), Forecasting confirmed and recovered COVID-19 cases and deaths in Egypt after the genetic mutation of the virus. *Commun. Math. Biol. and Neurosci.* https:// doi.org /10.28919/cmbn/6888

Adiga, A., Dubhashi, D., Lewis, B., Marathe, M., Venkatramanan, S., and Vulikanti, A. (2020), Mathemtical models for COVID-19 pandemic. *A comparative Analysis. Review Article*, **100**, 793-807.

Akaike, H. (1974). A new look at the statistical model identification.*IEEE Trans Automat Contr*, **19**, 716–723.

Arellano-Valle, R.B., andGenton, M.G. (2005). Fundamental skew distributions. *Journal of Multivariate Analysis.* **96**, 93– 116.

Arfan, M., Shah, K., Abdeljawad, T., Mlaiki, N., and Ullah, A. (2021) A Caputo power law model predicting the spread of the COVID-19 outbreak in Pakistan Alex. *Eng. J*, **60**, 447-456.

Awajan, A., Al- Hasanain, R., Elkaroui, E., Al Edamat, A., Al- Gounmeein, R., Al- Jawarneh, A., Ayyoub, H., and Al- Farajat, E. (2024), Time Series Forecasting of New Cases for COVID-19 Pandemic in Jordan Using Enhanced Hybrid EMD-ARIMA. *J Stat App Pro*, **13**, 261-271.

Balakrishnan, N., and Scarpa, B. (2012), Multivariate measures of skewness for the skew-normal distribution.*Journal of Multivariate Analysis*, **104**, 73–87.

Box, G., and Jenkins, G. (1970), *Time Series Analysis: Forecasting and Control.* Holden-Day, San Francisco.

Busari, S I., and Samson, T, K. (2022), Modelling and forecasting new cases of COVID-19 in Nigeria: Comparison of regression, ARIMA and machine learning models. *Scientific African*, **18**, e01404.

Chang, C., Wang, W Y., and Liu, C.H. (2007), The development of a weighted evolving fuzzy neural network for PCB sales forecasting, *Journal of Expert Systems with Applications*, **32**, 86-96.

Ebrahimoghli, R., Abbasi-Ghahramanloo, A., Moradi, E., and Adham, D. (2023), The COVID-19 pandemic's toll in IRAN after two years: an interrupted time series analysis of weekly all-cause mortality data. *BMC public Health*, **23**, 442.

Gupta,A., and Chang, C. (2002). Multivariate Skew-symmetric Distributions, *Applied mathematics Letters*, **16(5)**, 634-664.

Jahangir, M.A., and Muheem, M.F. (2020),Coronavirus (COVID-19): history, current knowledge and pipeline medications. *Int. J. Pharm. Pharmacol*, **4**, 1-5.

Khan, F., Saed, A., and Ali, Sh. (2020), Modelling and forecasting of new cases, deaths and recover cases of COVID-19 by using vector autoregressive model in Pakistan. *Chaos, solitons and fractals*, **140**, 110189.

Kilian, L., and Lütkepohl, H. (2017), *Structural vector autoregressive Analysis.* (Themes in Modern Econometrics), Cambridge, United Kingdom.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer Science and Business Media. Berlin, Heidelberg.

Maleki, M., Mahmoudi, M.R., Weaith, D., and Pho, K.H. (2020), Time series modelling to forecast the confirmed and recovered cases of COVID-19. *Travel medicine and infectious disease.* **37**, 101742.

Mardia, K. (1974), Applications of some measures of multivariate skewness and kurtosis in testing normality and robustness studies. *Sankhy-a, ser. B*, **36**, 115–128.

Meimela, A., Lestari, S., Mahdy, I.F., Toharudin, T., and Ruchjana, B.N. (2021), Modeling of covid-19 in Indonesia using vector autoregressive integrated moving average. *Journal of physics, conference series*, **1722**, 012079.

Nguyen, H.M., Turk, P., and Mcwilliams, A.D. (2021), Forecasting COVID-19 hospital census: A multivariate time series model based on local infection incidence. *Jmir public health and surveillance*, **7(8)**, e28195.

Petropoulos, F., Makridakis, S., and Stylianou, N. (2022), COVID-19: Forecasting confirmed cases and deaths with a simple time series model. *International journal if forecasting*, **38**, 439-452 .

Rajab, Kh., Kamalov, F., and Cherukuri, A.K. (2022), Forecasting COVID-19: Vector Autoregression-Based Model. *Arab J Sci Eng.* **47**, 6851-6860.

Sahai, A.K., Rath, N., Sood, V., and Singh, M.P. (2020). ARIMA modelling & forecasting of COVID-19 in top five affected countries. *Diabetes Metabol. Syndr. Clin. Res. Rev*, **14(5)**, 1419–1427.

Sahu, S.K., Dipak, D., and Branco, M.(2003), A new class of multivariate distributions with applications to Bayesian regression models. *Canadian Journal of Statistics*, **129(31)**, 129-150.

Salgotra, R., Gandomi, M., and Gandomi, A.H. (2020), Time series analysis and forecast of the COVID-19 pandemic in India using genetic programming. *Chaos Solitons Fractals.* **138**, 109945.

Schwarz, G. (1978), Estimating the dimension of a model. *Ann Stat*, **6(2)**, 461–464.

Sujath, R., Chatterjee, J.M., and Hassanien, A.E. (2020), A machine learning forecasting model for COVID-19 pandemic in India. Stoch. *Env. Res. Risk Assess*, **34**, 959–972.

Talkhi, N., Akhavan Fatemi, N., Ataei, Z., and Jabari Nooghabi, M. (2021), Modeling and forecasting number of confirmed and death caused COVID-19 in IRAN: A comparison of time series forecasting methods. *Biomed signal process control*, **66**, 102494.

Tsay, R.S. (2002), *Analysis of Financial Time Series*. John Wiley & Sons, Inc. Hoboken, New Jersey.

Tsung, L. (2009), Maximum likelihood estimation for multivariate skew normal mixture models. *Journal of Multivariate Analysis*, **100(2)**, 257-265.