Vector Autoregression in Forecasting COVID-19 Under-Reporting–Nepal as a Case Study

Jyoti U. Devkota

Department of Mathematics, Kathmandu University, Dhulikhel, Nepal

Correspondence to: drjdevkota@ku.edu.np

Abstract: *This paper aims to understand and predict the dynamics of spread of COVID-19. It is based on government data on COVID-19 from February 1, 2021 to August 31, 2021. First, Vector Autoregression (VAR) model is used here to model the interrelationships between time series data of daily tested, infected, dead and discharged. The impact of under-reporting on interrelated variables is quantified. The behavior of the parameters of these VAR model is also analyzed. The entire time period of study is divided into three phases, according to the intensity of vaccination drive. The impact of vaccination in controlling the spread of the pandemic is measured by studying the behavior of the coefficients of VAR model for these three time periods. Then, Granger causality is also measured. At 10% level of significance, it is found that if the number of infected is under-reported today, this is due to the significant influence of number of infected until previous two days. The number of discharged one day ago and three days ago also significantly influence this number. Number of tests conducted two days ago also significantly contributes to this underreporting. The impact of latent variables on the spread of COVID-19 is measured here.*

Keywords: Under-reporting, Granger causality, Multivariate statistics, Time series, Latent variables

DOI: `https://doi.org/10.3126/jnms.v5i2.50016`

1 Introduction

Nepal has the capacity to conduct PCR tests on 23,000 samples on average daily. But tests are conducted only on 15000 samples on average daily [1]. The government of Nepal has announced free COVID tests and treatment for its citizens. But the services are very fast and efficient in the private sectors. A PCR test costs around 20 USD in the private sector. This is beyond the reach of an average citizen of such countries. Sometimes, these tests give negative results during the first time and positive results during the second time. Thus less people turn up for testing. This has resulted in less contact tracing and rapid increase in infections. Thus there are under-reported cases of infections and deaths from COVID–19.

In Nepal, the first case of COVID-19 was reported on 23 January 2020. It was a 32 years old man who had recently returned from Huwei, China. The patient recovered and contacts were also asymptomatic [5]. The Government of Nepal enforced a strict lockdown from 24 March 2020. This lockdown was completely relaxed on 19 September 2020.

The second wave of COVID-19 started in April 2021. A lockdown was enforced on 29 April 2021 to control the spread of pandemic. This lockdown was partially relaxed from 22 June 2021[3]. From 22 June, vehicles could move on the road according to the odd and even numbers of the number plate. The shops and departmental stores were allowed to be open till 11:00 A.M. only. This lockdown was completely relaxed on 1 September 2021. Only certain restrictions were kept in place [17].

The second wave was due to Nepalese migrant workers coming back to Nepal, after the announcement of lockdowns and curfews in different parts of India [8]. Nepal shares an open border with its southern neighbor India. People from either side of the border can freely cross and work without work permits [2]. Initially the positive cases of COVID 19 were mostly either Indian nationals working in Nepal or Nepali workers who had recently returned from India. Thousands of workers have returned to Nepal without proper screening. This was also the scenario in the first wave. Many workers were stranded in different parts of India due to lockdowns in both countries. Many workers were stranded in the border [4, 9, 10, 11, 12].

Nationwide vaccination campaign was started at the end of January 2021 with operational committees

and task forces formed at all levels. At first health and social sector and front-liners (first priority group) were vaccinated achieving 86% coverage. Following this, people over 65 years throughout the country and those over 55 years in high mountainous terrain were vaccinated achieving 77% coverage. This priority group received second dose in the first week of August [16]. Nepal aims to vaccinate at least 72% of the population and vaccines used are Serum Institute India's Covishield, AstraZeneca's COVID-19, China's Vero Cell, Johnson and Johnson's Janssen and Pfizer-BioNtech [13].

This paper is arranged in the following manner. This section is followed by the section Methods. Here data and methodology used in this paper are explained in detail. This section followed by Results and then Discussion. The conclusions of this paper are mentioned in the last section of this paper titled Conclusions.

2 Methods

Data: This study is based on daily COVID-19 updates published by the Ministry of Health and Population. Thus it is a secondary data. It is published in the official website of Ministry of Health and Population, Government of Nepal developed for COVID 19 updates [6]. It gives daily data of tested, infected, dead, recovered, active, and total infected, total recovered and total dead cases. The data from 1 February 2021 to 31 August 2021 are taken for this study. During this period the vaccination was in full swing, in Nepal. Thus to study the impact of vaccination on the spread of the pandemic, this time period was found suitable.

VAR models: Vector autoregressive (VAR) models are multivariate time series models [15]. A multivariate time series z_t follows VAR model of order p, VAR(p) if

$$z_t = \phi_0 + \sum_{i=1}^{p} \phi_i z_{t-i} + a_t, \tag{1}$$

where ϕ_0 is a k dimensional constant vector and ϕ_i are $k \times k$ matrices for $i > 0$, $\phi_p \neq 0$, and a_t is a sequence of independently and identically distributed (iid) random vectors with mean 0 and covariance matrix Σ_a, which is positive definite. To begin with, let's consider a model of order 1, VAR(1) model. We put $p = 1$, in equation (1) and get

$$z_t = \phi_0 + \phi_1 z_{t-1} + a_t \tag{2}$$

This bi-variate model can also be explicitly written as

$$\begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{10} \\ \phi_{20} \end{bmatrix} + \begin{bmatrix} \phi_{1,11} & \phi_{1,12} \\ \phi_{1,21} & \phi_{1,22} \end{bmatrix} \begin{bmatrix} z_{1,t-1} \\ z_{2,t-1} \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} \tag{3}$$

For a k dimensional VAR(1) model with a backward shift operator B, Equation (2) can be reduced to

$$\phi(B) z_t = \phi(0) + a_t \tag{4}$$

where

$$\phi(B) = I_k - \phi_i B^i$$

For a k dimensional VAR(p), model (1) reduces to

$$\phi(B) z_t = \phi(0) + a_t$$

where $\phi(B) = I_k - \sum_{i=1}^{p} \phi_i B^i$ is a matrix polynomial of degree p.

Granger causality: It introduces the concept causality with reference to the VAR models [15]. For example with reference to equation (3) in bivariate VAR(1) model, if $\phi_{1,12} = 0$ but $\phi_{1,21} \neq 0$, we see that $z_{2,t}$ depends on $z_{1,t-1}$ so that knowing $z_{1,t-1}$ is helpful in forecasting $z_{2,t}$ but not vice versa. Thus $z_{1,t-1}$ causes $z_{2,t}$. Under Granger's framework $z_{1,t}$ causes $z_{2,t}$, if past information of $z_{1,t}$ improves the performance of $z_{2,t}$.

Let's consider the following vector auto-regressive moving average VARMA (p, q) model

$$Y_t = \alpha + \beta t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \beta_1 X_{t-1} + \cdots + \beta_q X_{t-q} + \epsilon_t. \tag{5}$$

Here $t = 1, 2, \cdots, T$. The equation (5) is an unrestricted model (UR). We say that X does not Granger cause Y, if $\beta_i = 0$, $\forall i$

$$Y_t = \alpha + \beta t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \epsilon_t. \tag{6}$$

This equation (6) is a restricted model (R).

We do not reject the null hypothesis $H_0 : \beta_1 = \beta_2 = \beta_q$ or H_0: There is no Granger causality, if model (5) and (6) are more or less same, that is, if $SSR_{UR} \approx SSR_R$, that means the residual sum of squares of unrestricted model is the same as the residual sum of squares of the restricted model. If the F test statistic

$$F = \frac{\frac{SSR_R - SSR_{UR}}{q}}{\frac{SSR_{UR}}{T-q-(p+2)}}$$

is greater than 0.95 quantile of F- distribution with $(q, T-q-(p+2))$ degrees of freedom, we say X granger causes Y.

3 Results

The progression of the time series data for daily tested, infected, deaths and discharged is shown in Figure 1. The progression of its first order differences is shown in Figure 2. This is based on 212 observations from 1 February to 31 August 2021. The parameters of VAR model for the COVID-19 daily data on tested, infected, deaths and discharged are shown in Table 1. Here T_t stands for tested for COVID-19 on day t, I_t represents infected on day t, D_t represents deaths from COVID-19 on day t and Di_t stands for discharged on day t. As seen from Table 1, the roots of characteristic polynomial for both the models are less than 1. This implies that the system is stable. Here VAR(3) model is fitted to the COVID-19 daily data on tested, infected, deaths and discharged. VAR(4) model is fitted to the first order difference of COVID-19 daily data on tested, infected, deaths and discharged. Figure 3 and Figure 4 show the observed verses predicted data for VAR (3) and VAR (4) model. The closeness of fit between observed and predicted values shown in Figure 3 and Figure 4 highlight the accuracy of these models. The predicted values of daily tested and daily infected using VAR(3) and VAR(4) models are shown in Figure 5. The goodness of fit of these two models is further stressed in the Figure 6 - Figure 9. These are the plots of autocorrelation functions (ACF) and partial ACF of the residuals of VAR(3) and VAR(4) models. As seen from Figure 6 and Figure 7, the auto-correlation of lag 1 and higher are not statistically significant. The auto correlation of lag 0 is always 1. Similarly most of the partial autocorrelation values, as seen from Figure 8 and Figure 9 are not statistically significant.

The VAR(3) model for the COVID-19 daily data given in Table 1. Similarly, the behavior of the first differences of daily data of COVID-19 for tested, infected, deaths and discharged is also shown in Table 1. This is represented by VAR(4) model for differences. These results are based on 212 observations from 1 February to 31 August 2021. These can be interpreted in the following manner.

1.

$$\begin{aligned}
\hat{T_t} &= 1069.505^* + 18.607^* t + 0.424^* T_{t-1} + 0.705^* I_{t-1} + 2.013 D_{t-1} \\
&- 0.687^* Di_{t-1} - 0.1178 T_{t-2} + 0.365 I_{t-2} + 2.068 D_{t-2} + 0.016 Di_{t-2} \\
&+ 0.192^* T_{t-3} - 0.149 I_{t-3} - 2.253 D_{t-3} + 0.612^* Di_{t-3}, R^2 = 0.936
\end{aligned}$$

The data on the number of daily deaths and the number of discharged is not under-reported. As the number of tests conducted on the previous day increases by 1 unit, the number of tests conducted on this day increases by 0.424. This implies that if ten COVID-19 cases were not tested yesterday, that is they are under-reported, then, there will be an increment of four cases today. If 100 cases of COVID-19 would remain untested yesterday, this will result in an increment of 40 cases today. Similarly, as the number of infected increases by 1 in the previous day, the increase in the number of tested on this day increases by 0.705. If the 100 infected cases are unreported yesterday this implies an increment of 70 cases today. This gives us an idea of the extent to which less number of tests

Table 1: Details of VAR models

	Model	R^2	AIC
VAR(3)	1. $\hat{T_t} = 1069.505^* + 18.607^*t + 0.424^*T_{t-1} + 0.705^*I_{t-1} + 2.013D_{t-1}$ $-0.687^*Di_{t-1} - 0.1178T_{t-2} + 0.365I_{t-2} + 2.068D_{t-2}+$ $0.016Di_{t-2} + 0.192^*T_{t-3} - 0.149I_{t-3} - 2.253D_{t-3} + 0.612^*Di_{t-3}$	0.936	12255.79
	$2.\hat{I_t} = 199.669^* + 1.668t + 0.058T_{t-1} + 0.629^*I_{t-1} - 0.188D_{t-1}$ $-0.201^*Di_{t-1} - 0.134^*T_{t-2} + 0.474^*I_{t-2} - 0.185D_{t-2}$ $-0.034Di_{t-2} + 0.023T_{t-3} + 0.051I_{t-3} - 0.494D_{t-3} + 0.156^*Di_{t-3}$	0.960	
	$3.\hat{D_t} = 22.856^* - 0.018t - 0.00082T_{t-1} + 0.0036I_{t-1}$ $+0.175^*D_{t-1} + 0.0037Di_{t-1} - 0.0017T_{t-2} + 0.0043I_{t-2} + 0.044D_{t-2}$ $+0.0037Di_{t-2} - 0.002T_{t-3} + 0.008I_{t-3} + 0.025D_{t-3} - 0.00018Di_{t-3}$	0.465	
	4. $\hat{Di_t} = -97.781 - 2.363t + 0.053T_{t-1} + 0.028I_{t-1} - 0.092D_{t-1}$ $+0.357^*Di_{t-1} - 0.012T_{t-2} - 0.107I_{t-2} + 0.710D_{t-2}$ $+0.125^*Di_{t-2} + 0.005T_{t-3} + 0.155I_{t-3} + 0.632D_{t-3} + 0.335^*Di_{t-3}$ Roots of Characteristic Polynomial 0.9799, 0.9799, 0.6986, 0.6863, 0.6863, 0.4328, 0.4328, 0.4086, 0.2843, 0.2416, 0.2416, 0.1739	0.95	
VAR(4)	$1.\hat{\Delta T_t} = -0.4885^*\Delta T_{t-1} + 0.63^*\Delta I_{t-1} + 2.2572\Delta D_{t-1}$ $-0.6092^*\Delta Di_{t-1} - 0.524^*\Delta T_{t-2} + 0.938^*\Delta I_{t-2} + 4.23\Delta D_{t-2}$ $-0.415^*\Delta Di_{t-2} - 0.279^*\Delta T_{t-3} + 0.563\Delta I_{t-3} + 1.5583\Delta D_{t-3}$ $+0.251\Delta Di_{t-3} - 0.114\Delta T_{t-4} + 0.175\Delta I_{t-4} + 0.619\Delta D_{t-4} + 0.414^*\Delta Di_{t-4}$	0.261	12234.18
	$2.\hat{\Delta I_t} = 0.0344\Delta T_{t-1} - 0.275^*\Delta I_{t-1} + 0.092\Delta D_{t-1} - 0.159^*\Delta Di_{t-1}$ $-0.105^*\Delta T_{t-2} + 0.252^*\Delta I_{t-2} + 0.219\Delta D_{t-2} - 0.105\Delta Di_{t-2} - 0.084^*\Delta T_{t-3}$ $+0.195^*\Delta I_{t-3} + 0.029\Delta D_{t-3} + 0.089\Delta Di_{t-3} - 0.033\Delta T_{t-4}$ $+0.025\Delta I_{t-4} + 0.409\Delta D_{t-4} + 0.230^*\Delta Di_{t-4}$	0.231	
	3. $\Delta\hat{D}_t = -0.00049\Delta T_{t-1} + 0.0066\Delta I_{t-1} - 0.655^*\Delta D_{t-1} + 0.0042\Delta Di_{t-1}$ $-0.0017\Delta T_{t-2} + 0.0062\Delta I_{t-2} - 0.477^*\Delta D_{t-2} + 0.0092\Delta Di_{t-2}$ $-0.0018\Delta T_{t-3} + 0.0072\Delta I_{t-3} - 0.334^*\Delta D_{t-3} + 0.012\Delta Di_{t-3}$ $-0.0065^*\Delta T_{t-4} + 0.015\Delta I_{t-4} - 0.166^*\Delta D_{t-4} + 0.0065\Delta Di_{t-4}$	0.328	
	$4.\Delta\hat{Di}_t = 0.0505\Delta T_{t-1} + 0.096\Delta I_{t-1} - 0.284\Delta D_{t-1} - 0.530^*\Delta Di_{t-1}$ $0.039\Delta T_{t-2} - 0.044\Delta I_{t-2} + 0.170\Delta D_{t-2} - 0.298^*\Delta Di_{t-2}$ $+0.055\Delta T_{t-3} + 0.106\Delta I_{t-3} + 0.296\Delta D_{t-3} + 0.008\Delta Di_{t-3}$ $0.022\Delta T_{t-4} + 0.171\Delta I_{t-4} + 0.391\Delta D_{t-4} + 0.094\Delta Di_{t-4}$ Roots of Characteristics Polynomial 0.8071, 0.8071, 0.7118, 0.7118, 0.6772, 0.6772, 0.6684 0.6567, 0.6352, 0.6352, 0.6046, 0.6046, 0.5941, 0.5941, 0.383, 0.383	0.326	

* indicates that the value is significant at 10% level of significance.

conducted can have an impact of the number of tests and number of infections on the whole. Similarly for increment of 1 tested case 3 days ago, there is an increment of 0.192 cases tested this day. So for every increment of 3 cases, that is 1 for tested and infected in the previous day and 1 for tested three days ago, there will be an increment of $0.424 + 0.705 + 0.192 = 1.32$ tested cases. The coefficient of determination R^2 is 0.936. This implies that this model explains 93.6% variability of the data.

2.

$$\begin{aligned}\hat{I_t} &= 199.669^* + 1.668t + 0.058T_{t-1} + 0.629^*I_{t-1} - 0.188D_{t-1} \\ &- 0.201^*Di_{t-1} - 0.134^*T_{t-2} + 0.474^*I_{t-2} - 0.185D_{t-2} - 0.034Di_{t-2} \\ &+ 0.023T_{t-3} + 0.051I_{t-3} - 0.494D_{t-3} + 0.156^*Di_{t-3}, R^2 = 0.960\end{aligned}$$

If the number of infected in the previous day increases by 1 unit, the number of infected on this day increases by 0.629. If the number of people testing two days ago increases by 1 unit, the number of

infected on this day decreases by 0.134. As the number of infected two days ago increases by 1 unit, the number of infected on this day increases by 0.474. So for every increment of 3 cases, 1 for infected, 2 for tested and infected 2 days ago, there will be an increment of $0.629 - 0.134 + 0.474 = 0.969$ infected cases. The coefficient of determination R^2 is 0.96. This implies that this model explains 96% variability of the data.

3.

$$\begin{aligned}\hat{D}_t &= 22.856^* - 0.018t - 0.00082T_{t-1} + 0.0036I_{t-1} + 0.175^*D_{t-1} + 0.0037Di_{t-1}\\ &- 0.0017T_{t-2} + 0.0043I_{t-2} + 0.044D_{t-2} + 0.0037Di_{t-2} - 0.002T_{t-3}\\ &+ 0.008I_{t-3} + 0.025D_{t-3} - 0.00018Di_{t-3}, R^2 = 0.465.\end{aligned}$$

The number of deaths on this day is only significantly associated with the number of deaths taking place on previous day. The intercept is also significant at 10% level of significance. This implies that the inherent value of the number of deaths due to COVID-19 on a particular day from 1 February to 31 August 2021 is 22.856. The coefficient of determination R^2 is 0.465. This implies that this model explains 46.5% variability of the data.

4.

$$\begin{aligned}\hat{Di}_t &= -97.781 - 2.363t + 0.053T_{t-1} + 0.028I_{t-1} - 0.092D_{t-1} + 0.357^*Di_{t-1}\\ &- 0.012T_{t-2} - 0.107I_{t-2} + 0.710D_{t-2} + 0.125^Di_{t-2} + 0.005T_{t-3}\\ &+ 0.155I_{t-3} + 0.632D_{t-3} + 0.335^*Di_{t-3}, R^2 = 0.95\end{aligned}$$

The number of discharged on this day is significantly associated to the number of discharged in the previous three days. Here the coefficient of determination R^2 is 0.95. This implies that the model explains 95% variability of the data.

5.

$$\begin{aligned}\hat{\Delta T_t} &= -0.4885^*\Delta T_{t-1} + 0.63^*\Delta I_{t-1} + 2.2572\Delta D_{t-1} - 0.6092^*\Delta Di_{t-1}\\ &- 0.524^*\Delta T_{t-2} + 0.938^*\Delta I_{t-2} + 4.23\Delta D_{t-2} - 0.415^*\Delta Di_{t-2}\\ &- 0.279^*\Delta T_{t-3} + 0.563\Delta I_{t-3} + 1.5583\Delta D_{t-3} + 0.251\Delta Di_{t-3}\\ &- 0.114\Delta T_{t-4} + 0.175\Delta I_{t-4} + 0.619\Delta D_{t-4} + 0.414^*\Delta Di_{t-4}, R^2 = 0.261\end{aligned}$$

We see that the first differences of the number of tested on this day is significantly related to the first difference of the number of tested and infected in the previous day and the previous two days. It is also significantly associated with the first difference of number of discharged on the previous day, previous two days and previous four days. The coefficient of determination of this model is, R^2=0.261. This model is 26.1% accurate.

6.

$$\begin{aligned}\hat{\Delta I_t} &= 0.0344\Delta T_{t-1} - 0.275^*\Delta I_{t-1} + 0.092\Delta D_{t-1} - 0.159^*\Delta Di_{t-1}\\ &- 0.105^*\Delta T_{t-2} + 0.252^*\Delta I_{t-2} + 0.219\Delta D_{t-2} - 0.105\Delta Di_{t-2}\\ &- 0.084^*\Delta T_{t-3} + 0.195^*\Delta I_{t-3} + 0.029\Delta D_{t-3} + 0.089\Delta Di_{t-3}\\ &- 0.033\Delta T_{t-4} + 0.025\Delta I_{t-4} + 0.409\Delta D_{t-4} + 0.230^*\Delta Di_{t-4}, R^2 = 0.231\end{aligned}$$

We see that the first differences of the number of infected on this day is significantly related to the first difference of the number of infected on the previous day. It is also significantly related to the first difference of the number of tested and number of infected two days ago. It is also significantly related to the first difference of the number of tested and number of infected three days ago. This is also significantly associated with the first difference of number of discharged on the first day and fourth day. The coefficient of determination R^2 is 0.231. This implies that this model is 23.1% accurate.

7.

$$\begin{aligned}\Delta D_t &= -0.00049\Delta T_{t-1} + 0.0066\Delta I_{t-1} - 0.655^*\Delta D_{t-1} + 0.0042\Delta Di_{t-1} \\ &- 0.0017\Delta T_{t-2} + 0.0062\Delta I_{t-2} - 0.477^*\Delta D_{t-2} + 0.0092\Delta Di_{t-2} \\ &- 0.0018\Delta T_{t-3} + 0.0072\Delta I_{t-3} - 0.334^*\Delta D_{t-3} + 0.012\Delta Di_{t-3} \\ &- 0.0065^*\Delta T_{t-4} + 0.015\Delta I_{t-4} - 0.166^*\Delta D_{t-4} + 0.0065\Delta Di_{t-4}, R^2 = 0.328\end{aligned}$$

The first difference of the number of deaths on this day is significantly related the first difference of deaths on previous day, previous two days, three days and previous four days. It is also associated with the first difference of the number of tested on the previous four days. The coefficient of determinations R^2 is 0.328. This implies that this model is 32.8% accurate

8.

$$\begin{aligned}\Delta \hat{Di}_t &= 0.0505\Delta T_{t-1} + 0.096\Delta I_{t-1} - 0.284\Delta D_{t-1} - 0.530^*\Delta Di_{t-1} \\ &+ 0.039\Delta T_{t-2} - 0.044\Delta I_{t-2} + 0.170\Delta D_{t-2} - 0.298^*\Delta Di_{t-2} \\ &+ 0.055\Delta T_{t-3} + 0.106\Delta I_{t-3} + 0.296\Delta D_{t-3} + 0.008\Delta Di_{t-3} \\ &+ 0.022\Delta T_{t-4} + 0.171\Delta I_{t-4} + 0.391\Delta D_{t-4} + 0.094\Delta Di_{t-4}, R^2 = 0.326\end{aligned}$$

The first difference of number of discharge on this day is associated significantly to the first difference of discharge on the previous day and the previous two days. The coefficient of determinations R^2 is 0.326. This implies that this model is 32.6% accurate.

The values of the statistically significant regression coefficients of the VAR(3) model are also given in Table 2. Here also we are assuming that only data on number of tests conducted and number of infected is under-reported. Here * indicates that the value is significant at 10% level of significance.

Table 2: Regression coefficients significant at $\alpha = 0.1$for COVID-19 data

	Independent variables VAR(3) Model											
Dependent Variable	T_{t-1}	I_{t-1}	D_{t-1}	Di_{t-1}	T_{t-2}	I_{t-2}	D_{t-2}	Di_{t-2}	T_{t-3}	I_{t-3}	D_{t-3}	Di_{t-3}
Tested -T_t	0.424	0.705		-0.687					0.192			0.612
Infected - I_t		0.629		-0.201	-0.134	0.474						0.156
Deaths - D_t			0.175									
Discharged - Di_t				0.357				0.125				0.335

The rate of vaccination in Nepal is given in Table 3. The data on vaccination was provided in a concise form in the source mentioned below the table. Although Ministry of Health was the source of COVID-19 data for this study, the data on vaccination was not provided in detail here. So the other source was chosen. As seen from Table 3, the rate of vaccination is slow in February- March 2021 as only 5.9% are have received the first dose of vaccine by 5 April 2021. By the end of the period April–May 2021, only 7.4% have received the first dose and 2.4% have received the second dose. But the rate of vaccination has been very high during the period June–August 2021. By the end of this period, 18.7% have received the first dose and 15.6% have received both the doses of vaccine.

We observed the behavior of parameters of VAR(3) model during these three periods in Figure 10 – Figure 13. We see that the progression of values of the parameters changes in the period June to August. During this period, the vaccination rate is maximum. The time period February–August 2021 gives the behavior of the parameter for the VAR(3) fitted to the time series data of the entire time period. That is from 1 February to 31 August 2021. The behavior of statistically significant coefficients of VAR (3) model fitted to the data from 1 February to 31 August 2021 are studied over four different time periods in Figure 10 – Figure 13. The behavior of the VAR model for daily tested is shown in Figure 10. We see that there is a drastic change in the trend in the period of maximum vaccination June to August 2021. This change is extremely remarkable for intercept, slope, coefficients of I_{t-1}, Di_{t-1} and T_{t-3}. Similarly, Figure 11 shows

the behavior of statistically significant coefficients of the VAR model of daily infected. The behavior of the coefficients changes remarkably during this period of maximum vaccination, especially in the case of intercept, slope, I_{t-1}, D_{t-1} and I_{t-2}. As seen from Figure 12, there is a drastic change in the values of these coefficients for VAR model of deaths, for intercept and D_{t-1}. The coefficients of Di_{t-1}, Di_{t-2}, and Di_{t-3} change drastically in the period of maximum vaccination, as seen from Figure 13. This figure visually displays the coefficients of VAR model for daily discharged.

The interdependence and causality between the daily data of tested, infected, deaths and discharged are measured by granger causality in Table 4. The conclusions of causality are made here at 5% level of significance. The daily data of number of tests conducted on previous day, previous two days and previous three days is the granger cause of daily deaths due to COVID-19 and number of cases discharged or recovered from COVID-19. Thus it can be concluded that the number of tests conducted till previous three days granger causes deaths and recovery on this day. But the daily data on number of tests conducted on previous day is not a granger cause of number of infected on this day. But the number of tests conducted on previous two days and previous three days is a granger cause of number of infected on this day. This shows that, if fewer tests are conducted because of poor public health care system, then the tests are under-reported. This results in under-reported infected, deaths and discharged cases. Similarly as seen from Table 4, number of infected on previous day is a granger cause of number of tests conducted on this day. The number of infected till the previous three days is the granger cause of number of deaths and discharged on this day. Similarly, the number of deaths on previous day, previous two days and previous three days is the granger cause of number discharged (recovered) on this day. Thus it can be concluded that the number of deaths till previous three days granger causes the number of discharged on this day. The number of discharged on previous two days and previous three days granger causes the number of tests conducted on this day. Number of discharged on previous day, previous two days and previous three days granger causes number of infected on this day. Thus it can be concluded that the number of discharged (recovered) till previous three days granger causes number of infected on this day. It can also be seen that the number of discharged in the previous three days granger causes the number of deaths on this day.

Table 3: Percent of population vaccinated in Nepal

Sr. No.	Date	First Dose	Second Dose
1	1 Feb. 2021	0.4	0
2	5 Apr. 2021	5.9	0
3	30 May 2021	7.4	2.4
4	30 June 2021	9.1	2.6
5	31 Aug 2021	18.7	15.6

Source: Our world in data https://ourworldindata.org/covid-vaccinations?country=NPL

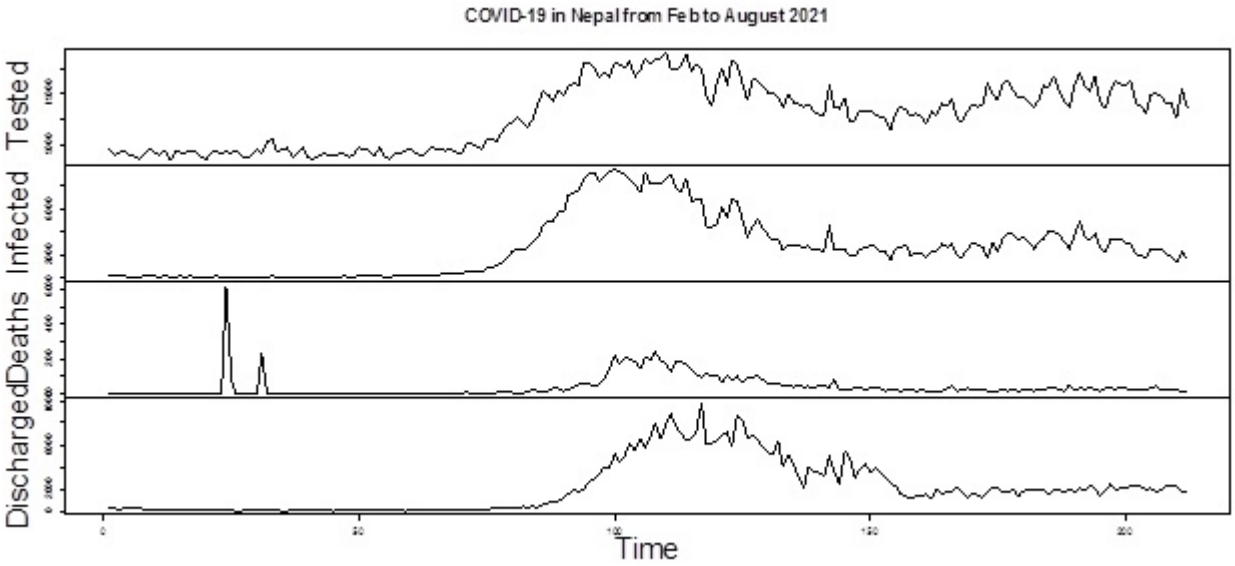

Figure 1: Time series daily data of tested, infected, deaths and discharged cases.

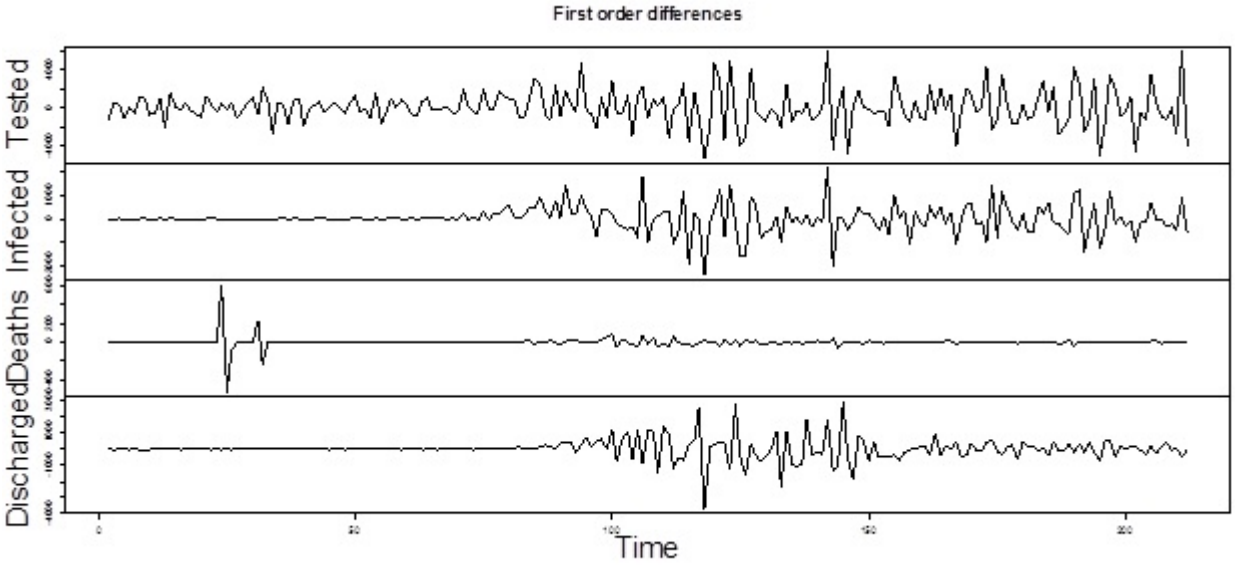

Figure 2: Time series data of first order differences of daily tested, infected, deaths and discharged cases.

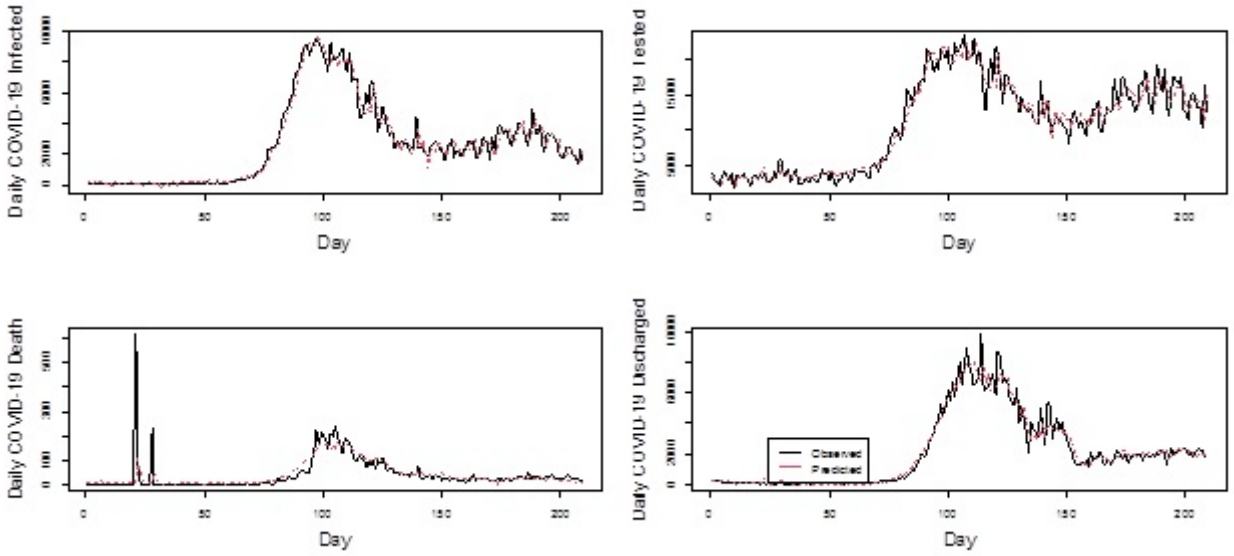

Figure 3: Observed versus fitted for VAR(3) of daily tested, infected, dead and discharged cases.

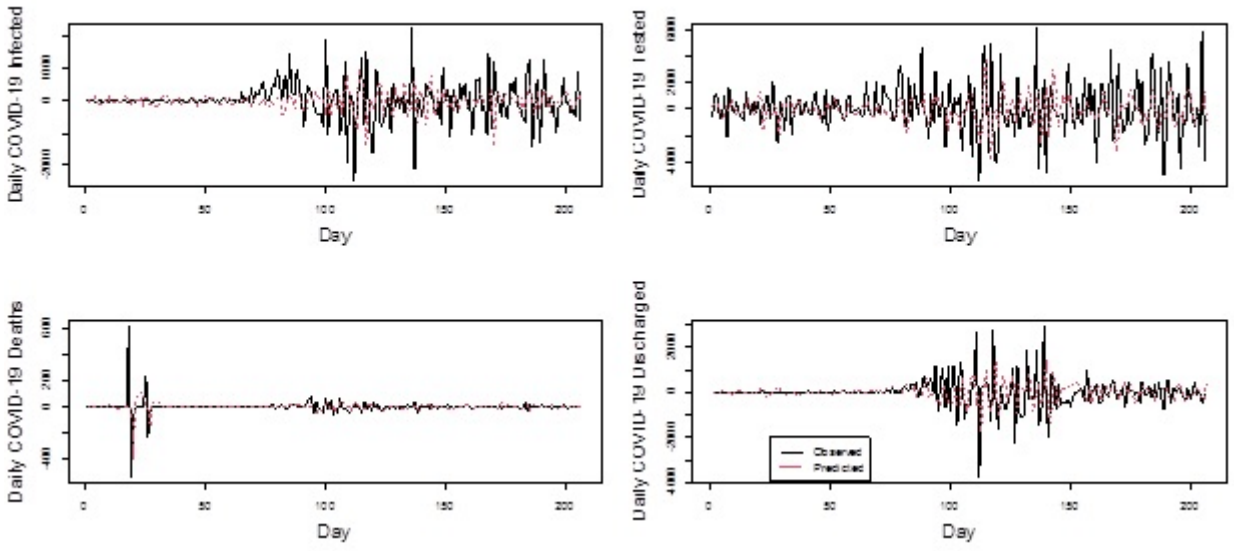

Figure 4: Observed versus fitted for VAR(4) for first order differences of daily tested, infected, deaths and discharged cases.

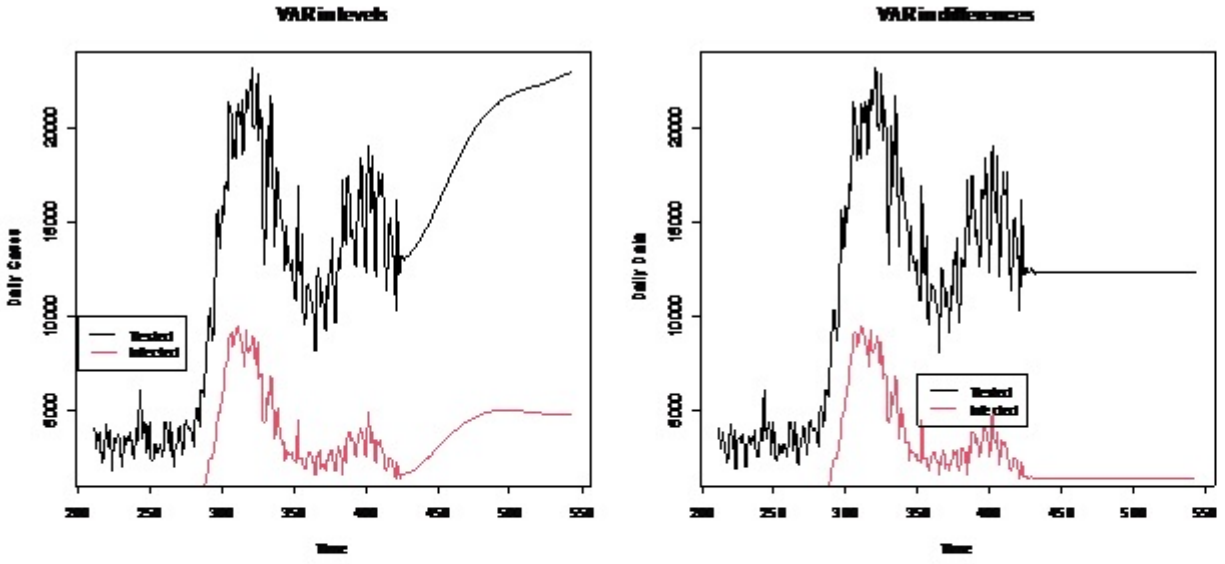

Figure 5: Predicted values of daily data of tested and infected using VAR and VAR in differences.

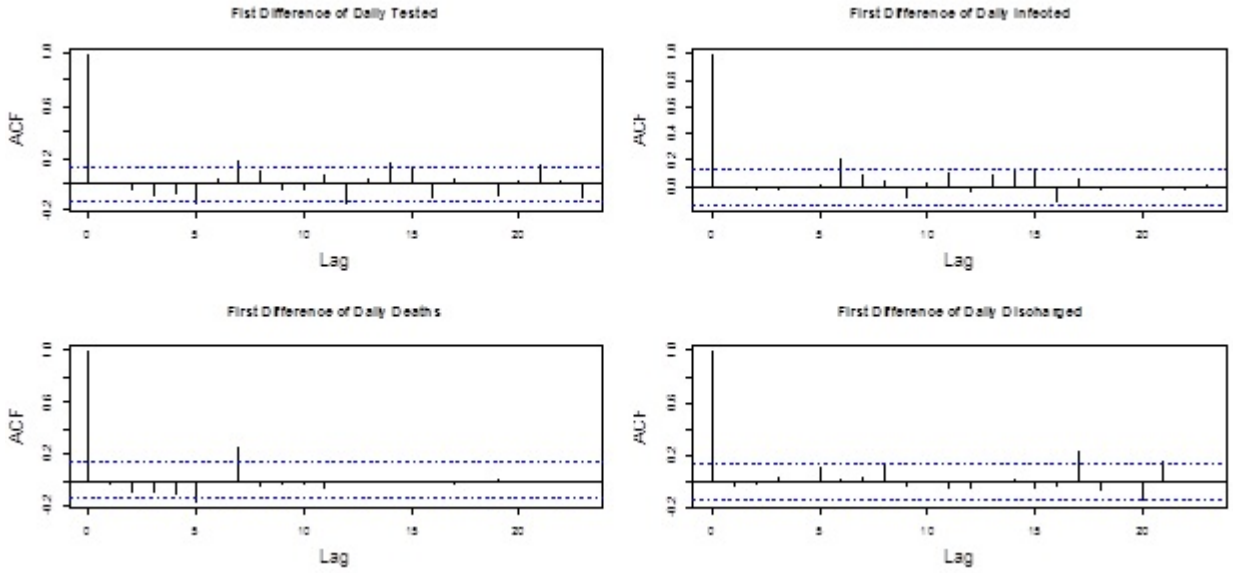

Figure 6: Auto correlation function (ACF) of residuals of VAR for first order difference.

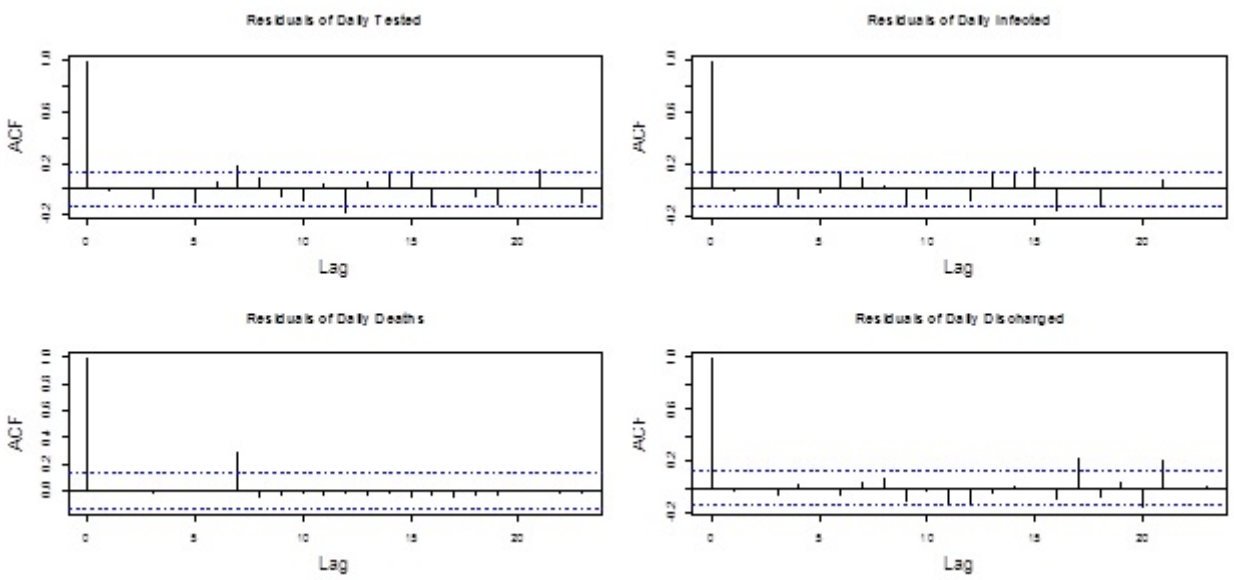

Figure 7: Auto correlation function of residuals of VAR.

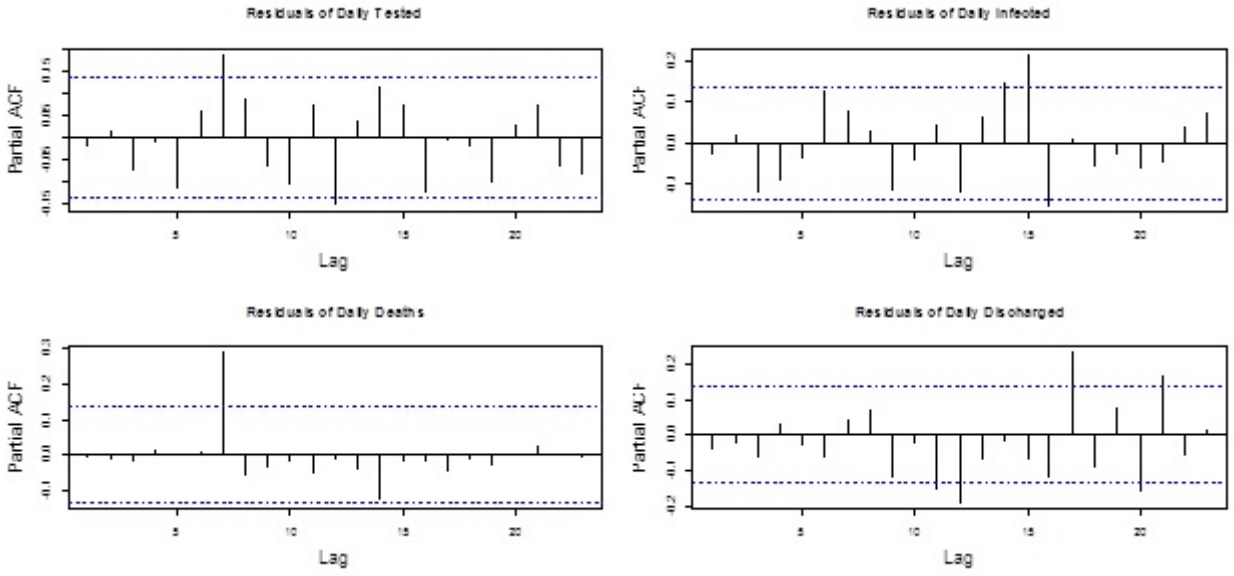

Figure 8: Partial auto correlation function of residuals of VAR.

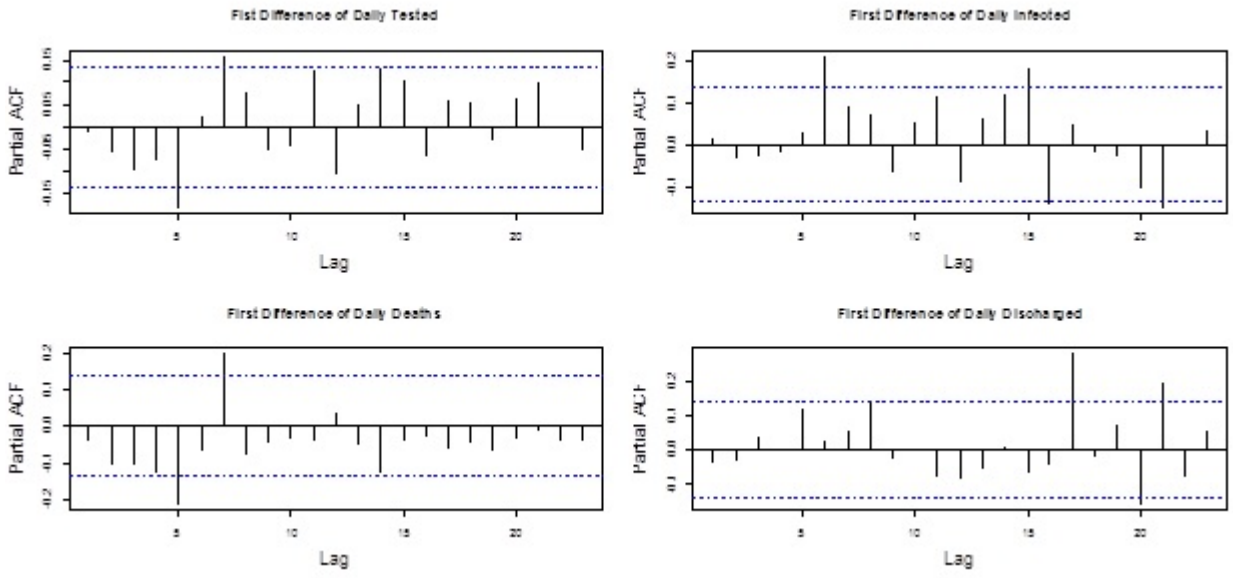

Figure 9: Partial auto correlation function of residuals of VAR for first order differences.

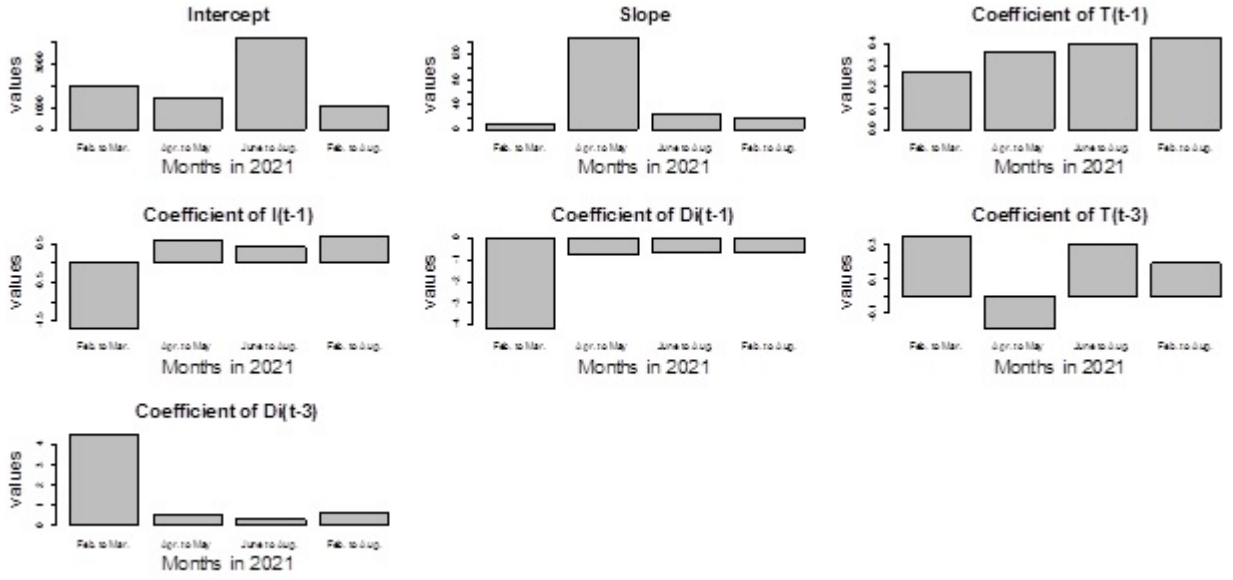

Figure 10: Behavior of coefficients of VAR (3) model for daily tested for different time periods.

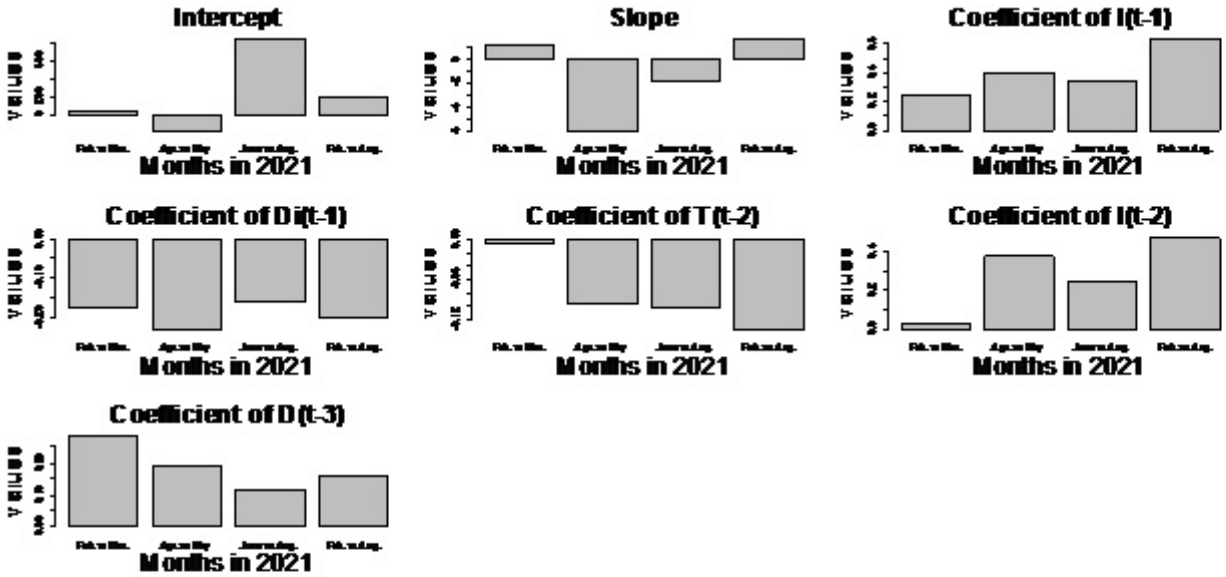

Figure 11: Behavior of coefficients of VAR (3) model for daily infected for different time periods.

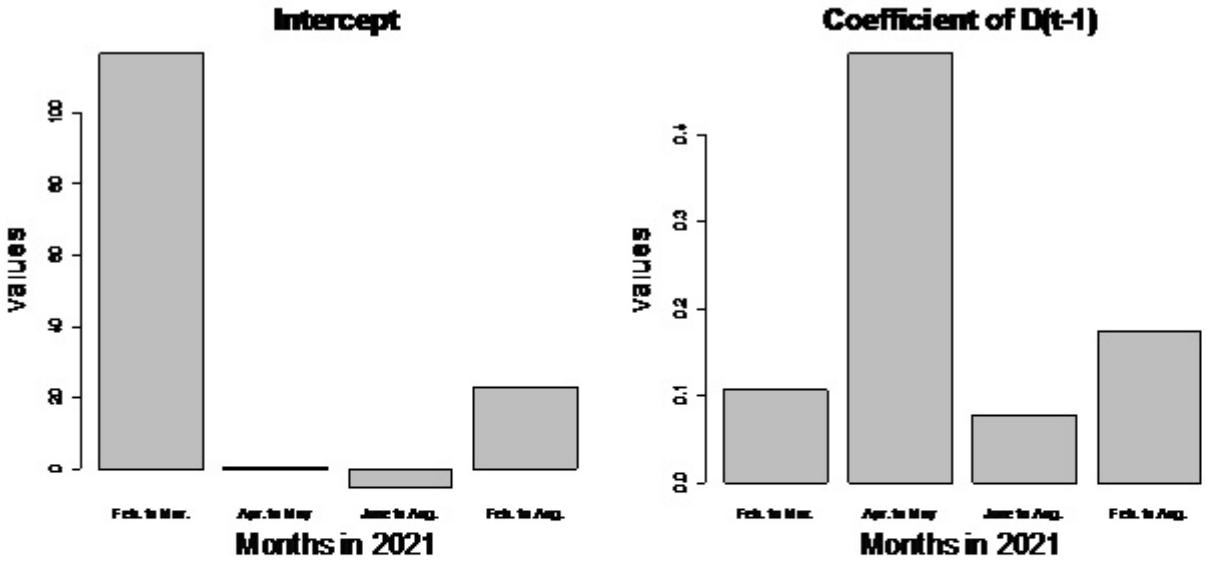

Figure 12: Behavior of coefficients of VAR (3) model for daily deaths for different time periods.

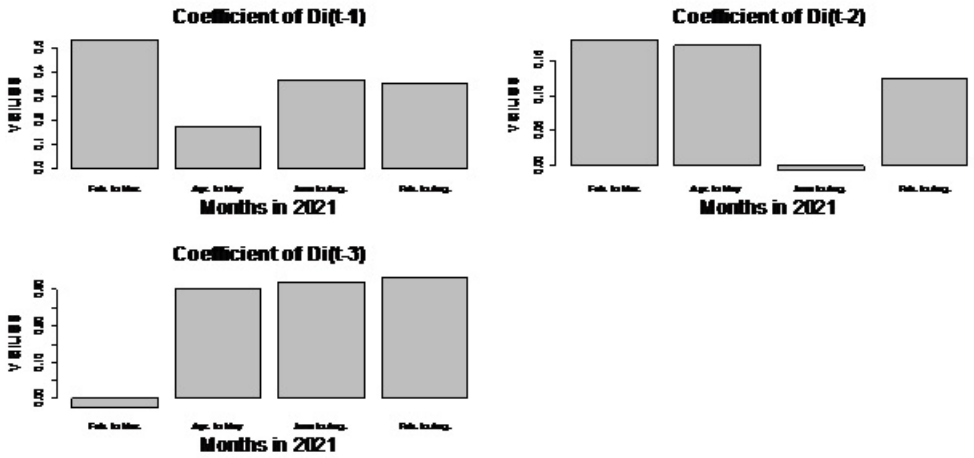

Figure 13: Behavior of coefficients of VAR (3) model for daily discharged for different time periods.

Table 4: Granger Causality of COVID-19 data

Sr. No.	Granger Cause	Effect Variable	Lag	P value	Remarks
1	Tested	Infected	1	0.448	granger cause
2	Tested	Infected	2	0.00074	granger cause
3	Tested	Infected	3	0.00063	granger cause
4	Tested	Deaths	1	0.0000063	granger cause
5	Tested	Deaths	2	0.0013	granger cause
6	Tested	Deaths	3	0.0205	granger cause
7	Tested	Discharged	1	0.000093	granger cause
8	Tested	Discharged	2	0.000025	granger cause
9	Tested	Discharged	2	0.000035	granger cause
10	Infected	Tested	1	0.015	granger cause
11	Infected	Tested	2	0.076	granger cause
12	Infected	Tested	3	0.133	granger cause
13	Infected	Deaths	1	< 0.01	granger cause
14	Infected	Deaths	2	< 0.01	granger cause
15	Infected	Deaths	3	< 0.01	granger cause
16	Infected	Discharged	1	< 0.01	granger cause
17	Infected	Discharged	2	< 0.01	granger cause
18	Infected	Discharged	3	< 0.01	granger cause
19	Deaths	Tested	1	0.143	granger cause
20	Deaths	Tested	2	0.456	granger cause
21	Deaths	Tested	3	0.259	granger cause
22	Deaths	Infected	1	0.948	granger cause
23	Deaths	Infected	2	0.683	granger cause
24	Deaths	Infected	3	0.952	granger cause
25	Deaths	Discharged	1	0.037	granger cause
26	Deaths	Discharged	2	0.013	granger cause
27	Deaths	Discharged	3	0.008	granger cause
28	Discharged	Tested	1	0.953	granger cause
29	Discharged	Tested	2	0.021	granger cause
30	Discharged	Tested	3	0.0064	granger cause
31	Discharged	Infected	1	0.026	granger cause
32	Discharged	Infected	2	0.002	granger cause
33	Discharged	Infected	3	0.0000204	granger cause
34	Discharged	Deaths	1	0.000000019	granger cause
35	Discharged	Deaths	2	0.000044	granger cause
36	Discharged	Deaths	3	0.002	granger cause

4 Discussion

This paper tries to measure under-reporting with the help of VAR models and granger causality. Here the main assumption is that numbers of daily tests conducted are less. If less people turn up for testing, there are under-reported cases of daily infections. These under-reported cases influence the daily and total data of discharged or recovered cases and deaths. This impact is quantified and measured in this paper. Under-reporting in recovered cases and deaths, if daily infections are under-reported, is also measured here. This method of understanding and measuring the interdependence helps in better management of the COVID-19 pandemic. This has special significance in developing country like Nepal. Such countries have a knowledge gap due to limited and scarce data.

5 Conclusions

The interdependence between the time series data of daily tested, infected, deaths and discharged is studied using VAR models. This is for the time period of 1 February to 31 August 2021. VAR(3) is found suitable for the COVID-19 data of daily tested, infected, deaths and discharged. VAR(4) is found suitable for its first order differences.

In an effort to study this interrelationship, several variables are considered, at different time lags. Many are found to have statistically significant values. These variables affect the predicted variable significantly. But many variables are also found to have statistically insignificant values.

Statistically significant coefficients of these two VAR models measure the degree of interdependence between the different lags of this multivariate data. So if less COVID-19 tests are conducted due to expensive testing facilities, tests will be under-reported. Under-reported tests also mean under-reported infection rate. So even if daily deaths and discharges due to COVID-19 take place in governmental facilities, it can be seen from the statistically significant coefficients of these models that under-reported test and infections have a significant impact on the daily data of deaths and discharged. Following conclusions are made from VAR models at 10% level of significance.

1. Number of tested in current day is significantly associated with the number of tested, infected and discharged previous day. It also depends on number of tested and discharged three days ago. So if fewer tests are conducted on the current day, it is due to the factors mentioned above. Here number of tests conducted can be associated with number of susceptible.

2. Number of infected on current day is significantly dependent on number of infected one day ago and two days ago. It is also affected by the number of discharged on previous day and number of discharged three days ago. Further number of tests conducted two days ago also significantly contributes to under-reporting.

3. There is a significance association between deaths reported on the current day with deaths on previous day.

4. The number discharged on current day is significantly associated with the number discharged one days ago, two days ago and three days ago.

5. It can be concluded from above that the number of susceptible and number of infected are under-reported.

Similar inferences about the first order differences of these variables have been made here. Studying and measuring such interdependence can help in better management of COVID-19 pandemic. The Delta variant is prevalent in Nepalese population with two third of the population having developed COVID-19 antibodies due to infection with the virus. Among vaccinated, those receiving second dose of vaccines have developed antibodies as high as 90%. Those receiving the first dose have developed antibodies as high as 80% [7, 14, 18]. Total deaths due to COVID-19 amounts to 11123 and the percent population receiving both doses of vaccine are 21.1 % as of date 29 September 2021. Following inferences can be made on the basis of granger causality test at 5% level of significance.

1. Number of tests conducted since previous three days granger causes deaths and discharged on this day.
2. Number of infected since the previous three days is the granger cause of number of deaths and discharged on this day.
3. Number of deaths since previous three days granger causes the number of discharged on this day.
4. Number of discharged since previous three days granger causes number of infected on this day.
5. Number of discharged in the previous three days granger causes the number of deaths on this day.

References

[1] Aljazeera, 2021, *Nepal to offer free COVID-19 tests and treatment as cases surge.*

[2] Baruah, N., and Arjal, N., 2018, Nepalese Labor Migration - A Status Report, *The Asia Foundation.*

[3] Garda World, 2021, *Nepal: Officials extends COVID-19 control measures in Kathmandu valley until at least June 3.*

[4] Organisation, I. L., 2020, Impact of COVID 19 on Nepali migrant workers: Protecting Nepali migrant workers during health and economic crisis 2020.

[5] Ministry, H. P., 2021, Health sector emergency response plan 2020, *Ministry of Health and Population.*

[6] Ministry, H. P., 2020, COVID 19 updates 2020, *Ministry of Health and Population.*

[7] myRepublica, 2021, *68.6 percent of total population of Nepal found to have developed COVID-19 antibody: Health Ministry.*

[8] Nepali Times, 2021, *1st sign of second wave in Nepal.*

[9] Relief Web, 2020, *Thousands of jobless migrant workers take risky journey home to Nepal.*

[10] The Kathmandu Post, 2020, *Government preparing to bring stranded migrant workers home.*

[11] The Kathmandu Post, 2020, *With hundreds of thousands of migrants predicted to return home, Nepal needs to brace for a crisis.*

[12] The Kathmandu Post, 2020, *Thousands of Nepalese stranded in India lockdown urge their government to rescue them.*

[13] The Kathmandu Post, 2021, *Nepal all set to receive 100,000 doses of Pfizer's COVID-19 vaccine from COVAX.*

[14] The Kathmandu Post, 2021, *COVID-19 antibodies in large chunk of population may not mean immunity.*

[15] Tsey, R. S., 2014, Multivariate time series analysis with R and financial applications, *John Wiley.*

[16] World, H. O., 2020, Nepal Story on COVID-19 vaccine deployment a good start, *World Health Organisation.*

[17] Xinhua, 2021, *Four months lockdown ends in Kathmandu valley.*

[18] Xinhua, 2021, *Over two thirds of Nepalese have COVID-19 antibodies: Survey*