



Available online at <http://scik.org>

Commun. Math. Biol. Neurosci. 2023, 2023:29

<https://doi.org/10.28919/cmbn/7887>

ISSN: 2052-2541

## SVIRD EPIDEMIC MODEL WITH DISCRETE-TIME HYBRID MARKOV/SEMI-MARKOV ASSUMPTIONS

FAIHATUZ ZUHAIROH<sup>1,2</sup>, DEDI ROSADI<sup>1,\*</sup>, ADHITYA RONNIE EFFENDIE<sup>1</sup>

<sup>1</sup>Department of Mathematics, Universitas Gadjah Mada, Daerah Istimewa Yogyakarta 55281, Indonesia

<sup>2</sup>Department of Mathematics Education, STKIP YPUP Makassar, Makassar 90223, Indonesia

Copyright © 2023 the author(s). This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Abstract.** Multi-state models with discrete-time Markov assumptions are widely used in epidemiology. However, this model has a memoryless property, making it less suitable for specific applications. One solution to this problem is to use additional assumptions by paying attention to the sojourn time in a particular state, which brings this model into a semi-Markov form. This paper aims to model the spread of infectious diseases by combining Markov and semi-Markov assumptions in one multi-state model known as a hybrid Markov/semi-Markov model. The first step is to check whether the SVIRD epidemic model satisfies the Markov assumptions. If not, then the SVIRD epidemic model uses a hybrid Markov/semi-Markov assumption. The second step is to test the semi-Markov hypothesis for each transition in the SVIRD model. Usually, the distribution is Geometric for semi-Markov sojourn times, but if the hypothesis is rejected, it uses the discrete Weibull distribution or negative Binomial. The hybrid Markov/semi-Markov model aims to reduce the complexity of the model in terms of the number of parameters to be estimated by only taking into account the sojourn time for transitions that do not meet the Markov assumptions. The final step is to make predictions by modifying the cohort state transition model by generating the number of individuals infected with an infectious disease at time  $t + 1$ .

**Keywords:** SVIRD epidemic model; hybrid Markov/semi-Markov; sojourn time; discrete-time; prediction.

**2020 AMS Subject Classification:** 60J10, 60J20.

---

\*Corresponding author

E-mail address: [dedirosadi@ugm.ac.id](mailto:dedirosadi@ugm.ac.id)

Received January 14, 2023

## 1. INTRODUCTION

The spread of infectious diseases can be modeled using deterministic and stochastic models according to Andersson and Britton [1]. The main purpose of communicable disease modeling is to analyze the spread, as well as provide solutions to the government about what actions can be taken to control the spread of the disease. As happened at the end of 2019, there was an outbreak of the coronavirus that spread throughout the world to cause a pandemic. The number of people infected has resulted in inadequate hospital capacity and excessive demand for healthcare workers. Due to the fact that the number of infected individuals is tied to the description and prediction of a group of susceptible individuals in a specific region, the spread of which is unexpected at the individual level, statistical patterns may be formed that lead to the usage of stochastic models.

One assumption of the Markov model is that the length of time a person spends in state  $i$  before moving to state  $j$  depends only on state  $i$ . Additionally, the sojourn time distribution exhibits a memoryless property, implying that it does not account for the length of probable occupancy in a specific condition. In the Markov model, the distribution of sojourn time is Geometrically distributed. However, in reality, such assumptions might result in implausible constraints. An alternative model that frames this concern is the semi-Markov model, which can be viewed as an extension of the Markov model. Discrete-time semi-Markov models have gained traction in recent years due to their usefulness in a number of different contexts, such as survival and reliability analysis by Barbu and Limnios [2], disability insurance by Stenberg, Manca, and Silvestrov [3], and credit risk by Vassiliou and Vasileiou [4] and Amico, Janssen, and Manca [5].

Most mathematical models for reliability assume that time is continuous. But in reality, the system has a discrete life span. Such as systems that work on demand, that work on a cycle basis, or that are only monitored at certain times (such as once a month or every day). In other words, all of this life is intrinsically discrete. The same is true of the spread of COVID-19, where the number of infected people is reported on a daily basis which can be seen on Worldometer (<https://www.worldometers.info/coronavirus/>). Previously conducted research on COVID-19 includes the determination of reproductive numbers using transition intensity

by Zuhairoh, Rosadi, and Effendie [6] and the prediction of COVID-19 distribution with the Richards curve model by Zuhairoh and Rosadi [7]. Bracquemond and Gaudoin [8] provides a good overview of the discrete probability distributions utilized in reliability theory. Additionally, insights from discrete-time semi-Markov models contribute to the application of time continuous semi-Markov models by Wu, Zheng, and Chen [9].

In recent years, discrete-time Markov models have become more widely used. As in the research Emmert and Allen [10] which investigated the spread of disease in a structured population using a discrete-time model. In comparison to semi-Markov processes, continuous-time and associated inference problems are of greater interest to many individuals. In this research, a discrete-time semi-Markov model will be employed. The research paper Limnios [11] provides an introduction to the discrete-time update procedure.

There are various advantages and disadvantages to using Markov and semi-Markov models, the main one being that Markov models are simpler and more visible. This simplifies the interpretation and understanding of conventional Markov models when used to simulate the spread of infectious diseases, for example. Simultaneously, the semi-Markov model allows quantification of duration under certain circumstances, due to a wider distribution of transit times. This provides a justification for developing a hybrid model that combines the two techniques as Verbeke and Guerry [12] has done in the case of manpower planning.

The contribution of this study is to build a multi-state model in the spread of infectious diseases and then estimate the model parameters formed by first seeing whether a multi-state model meets the Markov assumptions or not. Furthermore, if the model does not meet the Markov assumptions, a semi-Markov test is carried out to see the distribution of sojourn time at each transition that occurs so that a hybrid Markov/semi-Markov model can be applied. This study combines Markov and semi-Markov models applied to the epidemic model of infectious diseases, which uses six states: susceptible, vaccinated, infected vaccinated, infected unvaccinated, recovered, and deceased. In conclusion, a theorem is established to anticipate the number of COVID-19 instances during the subsequent period by modifying the cohort state transition model.

The strengths and weaknesses of Markov and semi-Markov models for modeling the spread of infectious diseases are discussed in Section 1. Section 2 introduces a hybrid Markov/semi-Markov model, a combination of Markov and semi-Markov models with Geometric, negative Binomial, and discrete Weibull sojourn time distributions. One of the advantages of a hybrid Markov/semi-Markov model is that it captures a fixed effect duration, which is useful when trying to estimate as few parameters as possible. Consequently, the hybrid Markov/semi-Markov model facilitates the advancement of the semi-Markov model even when there is a lack of data. In section 2, we also describe a predictive model for the spread of infectious diseases. In Section 3, we apply the SVIRD epidemic model to COVID-19 data and then test whether the model meets the Markov assumptions. If not, the transition probability will be calculated using a statistical test assuming a hybrid Markov/semi-Markov. Furthermore, in Section 4, predicting COVID-19 cases in the short term with the prediction equation of the SVIRD epidemic model. Finally, Section 5 contains the conclusions from the discussion presented previously.

## 2. HYBRID MARKOV/SEMI-MARKOV MODEL

A hybrid Markov/semi-Markov model is presented for each pair of  $(i, j)$  whether the transition from state  $i$  to state  $j$  can be considered a Markov or semi-Markov transition. The first step in building a hybrid Markov/semi-Markov model is to test the multi-state model to determine whether it meets the Markov assumptions. The Markov process introduced by a mathematician named Andrei A. Markov in 1906 is a stochastic process from the previous time that has no influence on the future time if the present time is known. Suppose  $X_t$  is a stochastic process that has a discrete-state space  $\mathcal{S} = \{1, 2, \dots, m\}$ . In general, according to Haberman and Pitacco [13] for any sequence of time points  $t_1 < t_2 < \dots < t_{n-1} < t_n$  which corresponds to the state set  $i_1, i_2, \dots, i_{n-1}, i_n$ , then the probability conditional must fulfill.

$$(1) \quad \Pr(X_{t_n} = i_n | X_{t_1} = i_1, \dots, X_{t_{n-1}} = i_{n-1}) = \Pr(X_{t_n} = i_n | X_{t_{n-1}} = i_{n-1})$$

If  $i_{n-1} = i, i_n = j$  and  $i_{n+1} = k$  then the first-order or 1-step transition probability can be written as follows.

$$(2) \quad \Pr(X_{t_n} = j | X_{t_1} = i_1, X_{t_2} = i_2, \dots, X_{t_{n-1}} = i) = \Pr(X_{t_n} = j | X_{t_{n-1}} = i) = p_{ij}$$

A second-order or higher-order Markov chain is a Markov chain that depends on two or more of the previous values. According to Shamshad et. al. [14], the probability of a second-order transition or a 2-step transition can be written as follows.

$$(3) \quad \Pr(X_{t_{n+1}} = k | X_{t_1} = i_1, \dots, X_{t_{n-1}} = i) = \Pr(X_{t_{n+1}} = k | X_{t_n} = j, X_{t_{n-1}} = i) = p_{ijk}$$

Markov property tests need to be performed to verify whether the transition probability is  $\Pr\{X_{t+1} = k | X_t = j, X_{t-1} = i\}$  from the current state to the next does not depend on the previous state. Written in the notation  $p_{ijk}$  which is the probability to state  $k$  with the previous condition that there has been a transition from state  $i$  to  $j$  ( $i \leq j \leq k$ ). If the Markov property applies, then  $p_{ijk} = p_{jk}$ . Three tests can be used to test the properties of Markov, namely:

- (1) Test based on a contingency table,
- (2) Tests to verify whether a chain of a given order,
- (3) Test to verify if the transition probability is constant over time.

A second test is used in this paper because it uses theoretical assumptions. According to Anderson and Goodman [15], the test to verify whether a chain is of second-order or not is defined as follows:

$$H_0 : p_{1jk} = p_{2jk} = \dots = p_{mjk} = p_{jk}, \quad j, k = 1, 2, \dots, m.$$

$$H_1 : \text{that it is of second-order}$$

The chi-square test statistic for the null hypothesis ( $H_0$ ) is

$$(4) \quad \chi_j^2 = \sum_{i,k} n_{ij}^* \frac{(\hat{p}_{ijk} - \hat{p}_{jk})^2}{\hat{p}_{jk}}$$

where

$$(5) \quad n_{ij}^* = \sum_k n_{ijk} = \sum_{t=1}^{T-1} n_{ij}(t)$$

$\chi_j^2$  has the normal limiting distribution with  $(m-1)^2$  degrees of freedom if  $H_0$  is true.

Consider now the joint hypothesis that  $p_{ijk} = p_{jk}$  where  $i, j, k = 1, 2, \dots, m$ . By computing the sum, a test of this joint hypothesis can be obtained.

$$(6) \quad \chi^2 = \sum_{j=1}^m \chi_j^2 = \sum_{j,i,k} n_{ij}^* \frac{(\hat{p}_{ijk} - \hat{p}_{jk})^2}{\hat{p}_{jk}}$$

The standard limiting distribution has  $m(m-1)^2$  degrees of freedom.

Due to its simplicity, the homogeneous time Markov model is used in various domains and applications. In addition, the estimates made have relatively few parameters and are not very demanding of data availability. However, Markov chains cannot consider the duration of stay; it is also less flexible because there is a memoryless property which means sojourn time is Geometric distribution. This issue is solved by the semi-Markov model.

If it does not meet the Markov assumptions, then proceed with second step, which is to estimate the sojourn time distribution for each transition between states. The second step is known as the semi-Markov hypothesis test. Similar to the Markov process, a semi-Markov process can transition from one state to another. To be more precise, the length of time spent in each state before moving to the next is a random variable that is independent of the next state of the new process.

A stochastic process  $X_t := J_{N(t)}$  is called a semi-Markov process if it considers the sojourn time in state  $i$  before transitioning to state  $j$ , where this sojourn time is a random variable with a cumulative distribution function  $F_{ij}(t)$ .

$$(7) \quad F_{ij}(t) = \Pr(H_n \leq t | J_n = i, J_{n+1} = j), \quad t \geq 0$$

where  $H_n = T_{n+1} - T_n$ .

The main difference for the Markov model is that the sojourn time distribution  $f$  can be any discrete probability distribution, combining the possible lengths of the sojourn. The Markov model with the  $P = (p_{ij} : i, j \in \mathcal{S})$  transition matrix can be called a semi-Markov model with a Geometric distribution of sojourn times where

$$(8) \quad q_{ij}(k) = \begin{cases} p_{ij}p_{ii}^{k-1}, & \text{if } i \neq j \text{ and } k \in \mathbb{N} \\ 0, & \text{elsewhere.} \end{cases}$$

Then the semi-Markov hypothesis is tested at the level of the sojourn time distribution  $h_{ij}$ . The transition from state  $i$  to  $j$  satisfies the Markov assumption if the sojourn time is geometrically distributed. Under the geometric hypothesis, the equation  $h_{ij}(2) = h_{ij}(1)(1 - h_{ij}(1))$  and a significant deviation from  $h_{ij}(1)(1 - h_{ij}(1)) - h_{ij}(2)$  should be seen as evidence to the contrary, that is, evidence in favor of a more general distribution of sojourn times. The statistical test

used according to Stenberg, Manca, and Silvestrov [3] is as follows.

$$(9) \quad \widehat{S}_{ij} = \frac{\sqrt{n_{ij}}(\widehat{h}_{ij}(1)(1 - \widehat{h}_{ij}(1)) - \widehat{h}_{ij}(2))}{\sqrt{\widehat{h}_{ij}(1)(1 - \widehat{h}_{ij}(1))^2(2 - \widehat{h}_{ij}(1))}}$$

$\widehat{h}_{ij}(t)$  is the maximum likelihood estimator of the probability  $h_{ij}(t)$

$$(10) \quad \widehat{h}_{ij}(t) = \frac{n_{ij}(t)}{n_{ij}}$$

where  $n_{ij}(t)$  is the total number of individuals who transition from state  $i$  to state  $j$  at time  $t$ .

Based on the Geometric hypothesis  $H_0$ , the statistical test  $\widehat{S}_{ij}$  has an asymptotic normal distribution. With a significance level of  $\alpha = 0.05$ , so will reject  $H_0$  if and only if  $|\widehat{S}_{ij}| > 1.96$ . Because this test makes it possible to make decisions about the  $\forall f_{ij}$  sojourn time distribution that allows the application of a hybrid Markov/semi-Markov model.

A hybrid Markov/semi-Markov model is a semi-Markov model that combines the sojourn time distribution from the classical Markov model for the pair  $(i, j)$  with a negative Binomial or discrete Weibull distribution from the semi-Markov model. The reason for choosing negative Binomial or discrete Weibull is because it is a generalization of the Geometric distribution.

The probability mass function of each distribution used is written in Nakagawa and Yoda [16]. Here is the relationship between the 3 distributions selected as the distribution of sojourn time of each transition that occurs between states.

(1) Discrete Weibull distribution ( $X \sim dwei(x; p, \beta)$ )

$$f(x) = (1 - p)^{x^\beta} - (1 - p)^{(x+1)^\beta} \quad x = 0, 1, 2, \dots$$

If the value of  $\beta = 1$  then it will be a Geometric distribution.

(2) Negative Binomial distribution ( $X \sim nbin(x; p, n)$ )

$$f(x) = \binom{n+x-1}{x} p^n (1-p)^x, \quad x = 0, 1, 2, \dots$$

If the value of  $n = 1$  then it will be a Geometric distribution.

(3) Geometric distribution ( $X \sim geo(x; p)$ )

$$f(x) = p(1-p)^x \quad x = 0, 1, 2, \dots$$

After determining the transition probabilities between states, the last step is to make predictions of infected cases at  $t + 1$  using Theorem 1. The prediction equation for disease vectors in the semi-Markov model is just an extension of the disease vector prediction equation in Markov sets. In the last case  $P(t)$  equals for every  $t$ . Thus, in order to get the disease vector prediction equation for the hybrid Markov/semi-Markov model, we may develop the Theorem 1 where  $p_{ij}(t)$  will depend on  $t$  because of the sojourn time distribution associated with  $(i, j)$  where the Markov hypothesis does not apply.

**Theorem 1.** *Suppose the known probability of transition from state  $i$  to  $j$  is written with  $p_{ij}$ . If there is an infected vector at time  $t$  with  $\mathbf{I}(t)$ , susceptible vector at time  $t$  with  $\mathbf{S}(t)$ , vaccinated vector at time  $t$  with  $\mathbf{V}(t)$ , and  $\mathbf{R}(t)$  for the recovered vector as well as  $\mathbf{D}(t)$  for the vector deceased at time  $t$ , then for the semi-Markov system the prediction equation for individuals infected with infectious diseases at time  $t + 1$  is obtained as follows.*

$$(11) \quad \begin{aligned} \mathbf{I}(t+1) = & \sum_{i=1}^m n_{i\mathbf{I}_s}(t)p_{i\mathbf{I}_s} + \sum_{j=1}^m (n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s} - n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s j}) + \\ & \sum_{i=1}^m n_{i\mathbf{I}_v}(t)p_{i\mathbf{I}_v} + \sum_{j=1}^m (n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v} - n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v j}) \end{aligned}$$

*Proof.* It is known that the SVIRD epidemic model consists of five states, namely susceptible, vaccinated, infected, recovered, and deceased. So it can be written as follows.

$$\mathbf{S}(t) + \mathbf{V}(t) + \mathbf{I}_s(t) + \mathbf{I}_v(t) + \mathbf{R}(t) + \mathbf{D}(t) = \mathbf{N}$$

The infected state will increase if there is a transition from a susceptible or vaccinated state. Meanwhile, it will decrease if there is a transition to a recovered or deceased state. Both statements can be written as

$$\mathbf{I}(t+1) = \mathbf{S}(t) - \mathbf{S}(t+1) + \mathbf{I}(t) - \mathbf{R}(t+1) - \mathbf{D}(t+1)$$

Meanwhile, the value of  $S(t+1)$  is obtained from

$$\begin{aligned} p_{ij}(t) &= \Pr(J_{n+1} = j, T_{n+1} - T_n = t | J_n = i, T_{n+1} - T_n > t - 1) \\ &= \frac{\Pr(J_{n+1} = j, T_{n+1} - T_n = t | J_n = i)}{\Pr(T_{n+1} - T_n > t - 1 | J_n = i)} \end{aligned}$$



$$\begin{aligned}
&= \frac{\Pr(J_{n+1} = j, T_{n+1} - T_n = t | J_n = i)}{1 - \Pr(T_{n+1} - T_n \leq t - 1 | J_n = i)} \\
&= \frac{q_{ij}(t)}{1 - \sum_{i \in \mathcal{I}} q_{ii}(t)}
\end{aligned}$$

If we know the number of individuals in each state at time  $t$ , then  $n_{ij}(t)$  is a binomial random variable with parameters  $n_i(t)$  and  $p_{ij}$ . So that the following expected value is obtained.

$$\bar{n}_{ij}(t) = \bar{n}_i(t-1)p_{ij}$$

Likewise, the number of people who leave the infected state due to recovery or death is denoted by  $n_{i,k+1}(t)$  with the following expected value.

$$\bar{n}_{i,k+1}(t) = \bar{n}_i(t-1)w_{ij} = \bar{n}_i(t-1)(r_{ij} + d_{ij})$$

Meanwhile, the number of newly infected people who come from vulnerable states is denoted  $n_{0j}(t+1)$  with an expected value of  $S(t+1)s_j$ . So the following prediction equation is obtained.

$$(12) \quad n_j(t+1) = \sum_{i=1}^m n_{ij}(t) + n_{0j}(t+1) - w_{ij}(t+1)$$

Given  $R(t+1), D(t+1), n_{0j}(t+1)$  with expected values  $R(t+1)r_j$  and  $D(t+1)d_j$ . The expected value from the equation (12) is obtained.

$$\bar{n}_j(t+1) = \sum_{i=1}^m \bar{n}_i(t)p_{ij} + S(t+1)s_j - R(t+1)r_j - D(t+1)d_j$$

So that the SVIRD epidemic model is obtained

$$n_{\mathbf{I}_s}(t+1) = \sum_{i=1}^m n_{i\mathbf{I}_s}(t)p_{i\mathbf{I}_s} + \sum_{j=1}^m (n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s} - n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s j})$$

The same proof holds for  $n_{\mathbf{I}_v}(t+1)$ , until obtained

$$n_{\mathbf{I}_v}(t+1) = \sum_{i=1}^m n_{i\mathbf{I}_v}(t)p_{i\mathbf{I}_v} + \sum_{j=1}^m (n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v} - n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v j})$$

□

### 3. SVIRD EPIDEMIC MODEL

The SVIRD epidemic model has six states, Susceptible, Vaccinated, and Infected, divided into two states: infected without vaccination and infected after being vaccinated, Recovered, and Deceased. This model has random variables  $S(t), V(t), I_s(t), I_v(t), R(t)$ , and  $D(t)$ , whose meanings are explained in Table 1. An illustrative representation of the mathematical model can be seen in Figure 1. It can also be seen in the transitions that occur between states. Here, it is assumed that vaccination effectively prevents an individual from dying, meaning that death comes from individuals who are not vaccinated only. In addition, this model assumes that there is immunity in individuals infected with the disease, so there is no transition out of the recovered state.

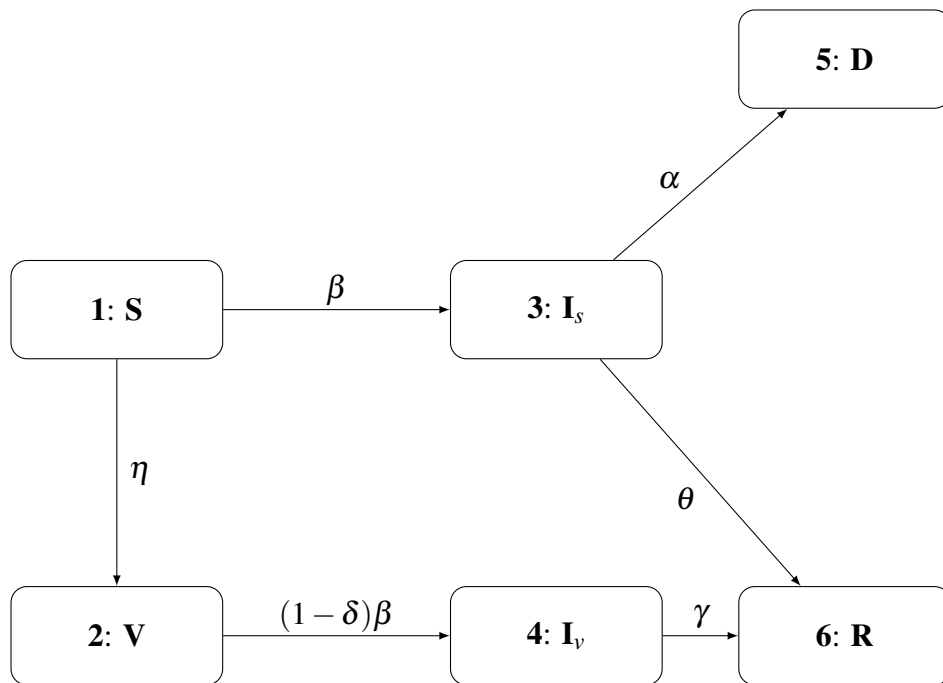


FIGURE 1. SVIRD epidemic model

TABLE 1. Basic notation

Notation	Description
$S(t)$	Number of susceptible at time $t$
$V(t)$	Number of vaccinated at time $t$
$I_s(t)$	Number of infected from the susceptible state at time $t$
$I_v(t)$	Number of infected from the vaccinated state at time $t$
$R(t)$	Number of recovered at time $t$
$D(t)$	Number of deceased at time $t$
$\beta$	Infection rate
$\eta$	Vaccination rates
$\delta$	Efficacy rates
$\gamma$	Recovery rate after vaccination
$\theta$	Recovery rate without vaccination
$\alpha$	Mortality rate

If using the Markov assumption, the following parameter estimates are obtained based on the results of previous research on Zuhairoh, Rosadi, and Effendie [17]. Meanwhile, if it does not meet the Markov assumption, then use the semi-Markov assumption by finding the distribution of sojourn time for each transition using the equation (9).

**Theorem 2.** *The estimated value of each parameter from the discrete-time SVIRD epidemic model can be obtained by the maximum likelihood method, which is written in Zuhairoh, Rosadi, and Effendie [17].*

(1) Vaccination rate

$$\hat{\eta} = \frac{N_{\eta_{kl}}}{N_{\eta_{kk}}}$$

(2) Infection rate

$$\hat{\beta} = \frac{N_{\beta_{kl}}}{N_{\beta_{kk}}}$$

(3) Efficacy rate

$$\hat{\delta} = \frac{N_{\delta_{kl}}}{N_{\delta_{kk}}}$$

(4) *Mortality rate*

$$\hat{\alpha} = \frac{N_{\alpha_{kl}}}{N_{\alpha_{kk}}}$$

(5) *Recovery rate without vaccination*

$$\hat{\theta} = \frac{N_{\theta_{kl}}}{N_{\theta_{kk}}}$$

(6) *Recovery rate after vaccination*

$$\hat{\gamma} = \frac{N_{\gamma_{kl}}}{N_{\gamma_{kk}}}$$

where  $N_{\eta_{kl}}, N_{\beta_{kl}}, N_{\delta_{kl}}, N_{\alpha_{kl}}, N_{\theta_{kl}}, N_{\gamma_{kl}}$  represents the total individuals that transitioned from state  $k$  to  $l$  and  $N_{\eta_{kk}}, N_{\beta_{kk}}, N_{\delta_{kk}}, N_{\alpha_{kk}}, N_{\theta_{kk}}, N_{\gamma_{kk}}$  represents the total individuals living in state  $k$  divided by the total population minus the number of individuals transitioning between states.

The subjects in this study were people infected with COVID-19 disease in Indonesia. The data is sourced from <https://covid19.go.id/peta-sebaran>, where the data used are the number of people who tested positive for COVID-19, the number who recovered, the number who died, and the number of people vaccinated each day. State vaccinated in this model uses data of individuals who have been vaccinated up to the second dose. Here we will see the average value of the efficacy of vaccines that have been used in Indonesia. This is in line with the WHO recommendation on giving two doses of vaccine to get better efficacy. There are five types of vaccines used in this model, namely Sinovac, Sinopharm, Pfizer, AstraZeneca, and modern, which are then searched for the average effectiveness by looking at the proportion of each vaccine used in Indonesia.

This research uses COVID-19 cases, so not all states are connected to each other. The transitions in the SVIRD epidemic model can be seen in Figure 1 which consists of only six transitions. The aim is to estimate the number of infected individuals in the short term. The limitations of the problem in this study have been previously stated that we assume that people who have received at least 2 doses of vaccination will not die from COVID-19.

The first step is to test whether the SVIRD epidemic model meets Markov properties or not using the equation (4), with the following hypothesis.

$$H_0 : p_{1jk} = p_{2jk} = \dots = p_{6jk} = p_{jk}, \quad j, k = 1, 2, \dots, 6.$$

$$H_1 : \text{is a second order Markov chain}$$

Then define a significance level of  $\alpha = 0.05$ .  $H_0$  states that the Markov property is satisfied. This model uses six states, so the test statistic  $\chi_{ijk}^2$  is distributed  $\chi^2$  with degrees of freedom  $6(5)^2$  under the null hypothesis. From the calculation of the value of  $\chi_{ijk}^2$  obtained as follows.

$$\begin{aligned} \chi_{ijk}^2 &= \sum_{i,j,k} n_{ij} \frac{(\hat{p}_{ijk} - \hat{p}_{jk})^2}{\hat{p}_{jk}} \\ &= n_{12} \frac{(\hat{p}_{124} - \hat{p}_{24})^2}{\hat{p}_{24}} + n_{13} \frac{(\hat{p}_{135} - \hat{p}_{35})^2}{\hat{p}_{35}} + n_{13} \frac{(\hat{p}_{136} - \hat{p}_{36})^2}{\hat{p}_{36}} + \\ &\quad n_{24} \frac{(\hat{p}_{246} - \hat{p}_{46})^2}{\hat{p}_{46}} \\ &= 299.73 \end{aligned}$$

where  $\chi_{tab}^2 = \chi_{(0.05;150)}^2 = 179.58$ . Because the value of  $\chi_{ijk}^2 > \chi_{tab}^2$  then  $H_0$  is rejected. So it is obtained that the SVIRD epidemic model does not meet the Markov characteristics.

The second step is testing the semi-Markov hypothesis at the  $\hat{h}_i$  sojourn time distribution using COVID-19 data, using the equation (10) and applying the  $\hat{S}_{ij}$  in the equation (9) for each SVIRD epidemic model transition. The results are summarized in Table 2.

$$\begin{aligned} \hat{S}_{12} &= \frac{\sqrt{n_{12}}(\hat{h}_{12}(1)(1 - \hat{h}_{12}(1)) - \hat{h}_{12}(2))}{\sqrt{\hat{h}_{12}(1)(1 - \hat{h}_{12}(1))^2(2 - \hat{h}_{12}(1))}} \\ &= \frac{\sqrt{667861} \left( \frac{667861}{1298911} \left( 1 - \frac{667861}{1298911} \right) - \left( \frac{667861}{1298911} \right) \right)}{\sqrt{\frac{667861}{1298911} \left( 1 - \frac{667861}{1298911} \right)^2 \left( 2 - \frac{667861}{1298911} \right)}} \\ &= -3051.075597 \end{aligned}$$

TABLE 2. Statistical test results  $\widehat{S}_{ij}$ 

State	$S$	$V$	$I_s$	$I_v$	$R$	$D$
$S$	-	-3051.08	-18.73	0	0	0
$V$	0	-	0	-6.25	0	0
$I_s$	0	0	-	0	0.13	-196.55
$I_v$	0	0	0	-	-1108.91	0
$R$	0	0	0	0	-	0
$D$	0	0	0	0	0	-

Under the Geometric  $H_0$  hypothesis, the  $S_{ij}$  test statistic is asymptotically Normal distribution. At the  $\alpha = 0.05$  significance level, the null hypothesis ( $H_0$ ) is rejected if  $|\widehat{S}_{ij}| > 1.96$ . Based on the results in Table 2, it is found that only  $h_{35}$  fulfills the Markov assumption while the other transitions use semi-Markov assumptions. The more common sojourn time distributions than the Geometric distribution are the negative Binomial distribution and the discrete Weibull distribution. So in this study, divided into 3 cases as follows.

- (1)  $h_{35}$  has a Geometric distribution, while the other transition sojourn times have a negative Binomial distribution.
- (2)  $h_{35}$  has a Geometric distribution, while the other transition sojourn times have a discrete Weibull distribution.
- (3)  $h_{35}$  has a Geometric distribution, while the other transition sojourn times have a negative Binomial or discrete Weibull distribution.

TABLE 3. AIC value of each discrete-time model

Model	Distribution	AIC
Markov	Geom	386.6896
semi-Markov	nbinom	435.2654
	dweibull	297.8675
	Geom-nbinom	373.6303
Hybrid Markov/semi-Markov	Geom-dweibull	311.9556
	Geom-nbinom-dweibull	188.5243

The Table 3 shows the best model used for making short-term predictions, from the Akaike information criterion (AIC) value, the best model is the hybrid Markov/semi-Markov model, that is, if it does not meet the geometric assumptions, then it can be selected. whether to use the negative Binomial distribution or the discrete Weibull.

#### 4. PREDICTION MODEL WITH HYBRID MARKOV/SEMI-MARKOV ASSUMPTION

According to the Theorem 1, predictions can be made for individuals infected with COVID-19 in the short term using the transition probabilities in Table 4. For the SVIRD epidemic model, the infected state is influenced by four conditions, namely the transition from the susceptible and vaccinated state and the transition to the recovered and deceased state. While the transition using only Markov assumptions is the transition from infected to cured. So for discrete time the following prediction results are obtained for  $t = 27$ , namely March 1, 2022 and  $t = 28$ , namely March 2, 2022. The prediction results for the next 10 days can be seen in the Table 5.

$$\begin{aligned}
\mathbf{I}(t+1) &= \sum_{i=1}^m n_{i\mathbf{I}_s}(t)p_{i\mathbf{I}_s} + \sum_{j=1}^m (n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s} - n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s j}) + \\
&\quad \sum_{i=1}^m n_{i\mathbf{I}_v}(t)p_{i\mathbf{I}_v} + \sum_{j=1}^m (n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v} - n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v j}) \\
\mathbf{I}(27) &= n_{13}(26)p_{13} + n_{24}(26)p_{24} + n_{35}(27)p_{33} - n_{35}(27)p_{35} + \\
&\quad n_{36}(27)p_{33} - n_{36}(27)p_{36} - n_{46}(27)p_{46} \\
&= 16556(0.354) + 8498(1.000) + 26357(0.628) - 26357(0.225) + \\
&\quad 13530(0.628) - 13530(0.775) - 325(1.000) \\
&= 22667
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{I}(t+1) &= \sum_{i=1}^m n_{i\mathbf{I}_s}(t)p_{i\mathbf{I}_s} + \sum_{j=1}^m (n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s} - n_{\mathbf{I}_s j}(t+1)p_{\mathbf{I}_s j}) + \\
&\quad \sum_{i=1}^m n_{i\mathbf{I}_v}(t)p_{i\mathbf{I}_v} + \sum_{j=1}^m (n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v} - n_{\mathbf{I}_v j}(t+1)p_{\mathbf{I}_v j}) \\
\mathbf{I}(28) &= n_{13}(27)p_{13} + n_{24}(27)p_{24} + n_{35}(28)p_{33} - n_{35}(28)p_{35} + \\
&\quad n_{36}(28)p_{33} - n_{36}(28)p_{36} - n_{46}(28)p_{46}
\end{aligned}$$

$$\begin{aligned}
&= 16340(0.354) + 8388(1.000) + 28371(0.628) - 28371(0.225) + \\
&\quad 14564(0.628) - 14564(0.775) - 376(1.000) \\
&= 23089
\end{aligned}$$

TABLE 4. Transition parameter estimation of each discrete-time model

Model	Transition	Distribution	$\hat{p}_{ij}$
Markov	1 → 2	Geometric	0.814
	1 → 3	Geometric	0.186
	2 → 4	Geometric	1.000
	3 → 5	Geometric	0.216
	3 → 6	Geometric	0.784
	4 → 6	Geometric	1.000
semi-Markov	1 → 2	dweibull	0.784
	1 → 3	dweibull	0.215
	2 → 4	dweibull	1.000
	3 → 5	dweibull	0.270
	3 → 6	dweibull	0.730
	4 → 6	dweibull	1.000
Hybrid Markov/semi-Markov	1 → 2	dweibull	0.646
	1 → 3	dweibull	0.354
	2 → 4	dweibull	1.000
	3 → 5	Geometric	0.225
	3 → 6	nbinomial	0.775
	4 → 6	dweibull	1.000

TABLE 5. Hasil Prediction results of the SVIRD epidemic model with discrete-time hybrid Markov/semi-Markov assumptions

Date	Actual	Prediction	MAPE (%)
1/03/2022	24,728	22,667	19.13
2/03/2022	40,920	23,089	
3/03/2022	37,259	31,911	
4/03/2022	26,347	29,783	
5/03/2022	30,156	24,879	
6/03/2022	24,867	27,652	
7/03/2022	21,380	24,556	
8/03/2022	30,148	23,784	
9/03/2022	26,336	23,837	
10/03/2022	21,311	23,127	



## 5. CONCLUSION

This study used discrete-time on a hybrid Markov/semi-Markov model to perform short-term predictions in people who confirmed positive for COVID-19. Hybrid Markov/semi-Markov combines Markov and semi-Markov assumptions on one multi-state model. Before determining the assumptions that will be used in the multi-state model, it is necessary to perform a chi-square test to see if the model meets the Markov assumptions or not. If not, then an additional assumption is required that is semi-Markov. The semi-Markov assumption is identical to sojourn time. Next, tests were performed to see if each transition met Markov or semi-Markov assumptions where we get the transitions  $p_{35}$  that satisfy the Markov assumption. In contrast, the other transitions use the semi-Markov assumption.

The sojourn time distribution used in this paper is three, namely Geometric, negative Binomial, and discrete Weibull distribution. So that in the application, it will be tested that each transition meets the sojourn time distribution, thus producing the appropriate transition probability. In the end, a prediction is used using the Theorem 1 for positive confirmed cases on the following day using the result of the subtraction of the number of COVID-19 cases in the infected state on the day  $i$  minus the number of COVID-19 cases on this day  $i - 1$  plus cases recovered and died on the  $i$  day.

## ACKNOWLEDGMENTS

The author would like to thank the Lembaga Pengelola Dana Pendidikan (LPDP) Ministry of Finance Republic of Indonesia and Penelitian Disertasi Doktor Tahun Anggaran 2022 DRTPM, Ministry of Education, Culture, Research, and Technology for taking part in providing financial support. Furthermore, thanks to all authors whose papers are cited in this paper as a valuable reference and resource for this paper.

## CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

**REFERENCES**

- [1] H. Andersson, T. Britton, *Stochastic epidemic models and their statistical analysis*, Springer New York, 2000. <https://doi.org/10.1007/978-1-4612-1158-7>.
- [2] V.S. Barbu, N. Limnios, Reliability of semi-markov systems in discrete time: modeling and estimation, in: K.B. Misra (Ed.), *Handbook of Performability Engineering*, Springer London, London, 2008: pp. 369–380. [https://doi.org/10.1007/978-1-84800-131-2\\_24](https://doi.org/10.1007/978-1-84800-131-2_24).
- [3] F. Stenberg, R. Manca, D. Silvestrov, Semi-Markov reward models for disability insurance, *Theory Stoch. Process.* 12 (2006), 239–254.
- [4] P.C.G. Vassiliou, A. Vasileiou, Asymptotic behaviour of the survival probabilities in an inhomogeneous semi-Markov model for the migration process in credit risk, *Linear Algebra Appl.* 438 (2013), 2880-2903. <https://doi.org/10.1016/j.laa.2012.11.035>.
- [5] G. D’Amico, J. Janssen, R. Manca, Initial and final backward and forward discrete time non-homogeneous semi-Markov credit risk models, *Methodol. Comput. Appl. Probab.* 12 (2009), 215-225. <https://doi.org/10.1007/s11009-009-9142-6>.
- [6] F. Zuhairoh, D. Rosadi, A.R. Effendie, Determination of basic reproduction numbers using transition intensities multi-state SIRD model for COVID-19 in Indonesia, *J. Phys.: Conf. Ser.* 1821 (2021), 012050. <https://doi.org/10.1088/1742-6596/1821/1/012050>.
- [7] F. Zuhairoh, D. Rosadi, Real-time forecasting of the COVID-19 epidemic using the Richards model in South Sulawesi, Indonesia, *Indonesian J. Sci. Technol.* 5 (2020), 456–462. <https://doi.org/10.17509/ijost.v5i3.26139>.
- [8] C. Bracquemond, O. Gaudoin, A survey on discrete lifetime distributions, *Int. J. Reliab. Qual. Safe. Eng.* 10 (2003), 69–98. <https://doi.org/10.1142/s0218539303001007>.
- [9] K. Wu, J. Zheng, J. Chen, Utilize state transition matrix model to predict the novel corona virus infection peak and patient distribution, *medRxiv.* (2022). <https://doi.org/10.1101/2020.02.16.20023614>.
- [10] K.E. Emmert, L.J.S. Allen, Population persistence and extinction in a discrete-time, stage-structured epidemic model, *J. Differ. Equ. Appl.* 10 (2004), 1177–1199. <https://doi.org/10.1080/10236190410001654151>.
- [11] N. Limnios, A functional central limit theorem for the empirical estimator of a semi-Markov kernel, *J. Non-parametric Stat.* 16 (2004), 13–18. <https://doi.org/10.1080/10485250310001622613>.
- [12] B. Verbeken, M.A. Guerry, Discrete time hybrid semi-Markov models in manpower planning, *Mathematics.* 9 (2021), 1681. <https://doi.org/10.3390/math9141681>.
- [13] S. Haberman, E. Pitacco, *Actuarial models for disability insurance*, CRC Press LLC, Boca Raton, 1999.
- [14] A. Shamshad, M. Bawadi, W. Wanhussin, et al. First and second order Markov chain models for synthetic generation of wind speed time series, *Energy.* 30 (2005), 693–708. <https://doi.org/10.1016/j.energy.2004.05.026>.

- [15] T.W. Anderson, L.A. Goodman, Statistical inference about Markov chains, *Ann. Math. Stat.* 28 (1957), 89–110. <https://www.jstor.org/stable/2237025>.
- [16] T. Nakagawa, H. Yoda, Relationships among distributions, *IEEE Trans. Rel. R-26* (1977), 352–353. <https://doi.org/10.1109/tr.1977.5220199>.
- [17] F. Zuhairoh, D. Rosadi, A.R. Effendie, Multi-state discrete-time Markov chain SVIRS model on the spread of COVID-19, *Eng. Lett.* 30 (2022), 598–608.