Asian Economic and Financial Review
ISSN(e): 2222-6737
ISSN(p): 2305-2147
DOI: 10.18488/journal.aefr.2020.108.906.919
Vol. 10, No. 8, 906-919.

URL: www.aessweb.com

A SIMULATION STUDY: OBTAINING A SUFFICIENT SAMPLE SIZE OF DISCRETE-TIME MARKOV CHAINS OF INVESTMENT IN A SHORT FREQUENCY OF TIME

Wajeeh Mustafa Sarsour[1+]
Shamsul Rijal Muhammad Sabri[2]

[1,2]*School of Mathematical Sciences, Universiti Sains Malaysia, Penang, Malaysia.*
Email: wajeeh.sa88@gmail.com Tel: +970598165812

(+ Corresponding author)

Article History
Received: 18 May 2020
Revised: 26 June 2020
Accepted: 28 July 2020
Published: 12 August 2020

Keywords
Simulation
Markov chain
Investment
Sufficient sample size
Comparison for matrices.

JEL Classification:
G11; G17; E22.

ABSTRACT

This paper deals with the problem of determining the sufficient sample size needed to estimate the transition matrix in the Markov chain. In particular, this paper focuses on systems with insufficient data or a short frequency of time caused by the difficulty of acquiring data. This study developed a Markov chain simulation technique that achieves a sufficient sample and can be used to estimate the size of the transition probability, despite having a short frequency of time. It also shows how this technique can be used in the short-, medium-, and long-term, and how a sufficient sample size can be found in these three situations. More specifically, this study illustrates the proposed simulation Markov chain model that estimates the transition probability matrix of the return of assets (ROA) in the industrial sector in Malaysia between 2007 and 2018. In this study, we present a method of determining an adequate sample size using a Markov chain simulation model. This model uses data from a number of companies in the industrial sector in Malaysia in order to study the performance of ROA and assist investors in making investment decisions. However, each company only has yearly ROA values. In other words, the frequency of the values is low, which makes studying the performance of ROA in the industrial sector more difficult. This could be the case because companies don't publish financial yearly reports, or because they are emerging companies that don't have adequate financial reports to calculate their ROA. This study was able to compensate for the lack of data through the number of companies used.

Contribution/Originality: This study is one of very few studies that has investigated how to determine an adequate sample size using a Markov chain simulation model. It presents a selection of companies, in order to study the performance of ROA in Malaysia's industrial sector, and to assist investors in their decision making.

1. INTRODUCTION

Over the past decades, there has been a considerable amount of debate regarding the extent to which the past can be used to forecast the future. Markov chain models, developed by Russian scientist Andrey Markov in 1906, have been used extensively to forecast the future. A Markov chain, which is a type of Markov process, is a stochastic model that describes a sequence of potential events in which the probability of every event depends entirely on the state accomplished in the previous event (Asmussen, 2008).

In stochastic analysis, the appeal of the Markov chain model is not new. Many stochastic processes used to model biological, physical, financial, engineering systems are Markovian, which means that it is easy to simulate,

compared to other models. For example, Anderson and Goodman (1957) estimated the transition probability matrix in a Markov chain model using maximum likelihoods, asymptotic distributions, and test hypotheses on model parameters. Moreover, Fama (1965) discussed the theory of random walks and provided strong evidence to support the stochastic nature of stock prices. Furthermore, Goldfeld (1973) applied a Markov model to switching regressions in order to study growth dynamics that rely on an extended period. Some studies have dealt with small sample sizes, such as Abidin and Jaffar (2014), who used a Geometric Brownian motion to forecast share prices in Bursa Malaysia.

The majority of Markov models are based on a sufficient data set, either in terms of sample size or frequency of time, as demonstrated by Fama (1965); Zhang and Zhang (2009); Mettle, Quaye, and Laryea (2014); Sarsour and Sabri (2020). However, these models would be difficult to calibrate for data that are characterized by a short frequency of time, which results in an unreliable estimation of the transition probability matrix. Increasing the sample size in a system of transition could help to overcome this shortcoming.

The present study proposes a new Markov chain simulation method to determine the required sample size, in order to obtain a reliable estimate of the transition probability matrix in cases with only a short frequency of time. From an applicative point of view, the main feature of the proposed simulation method is that it allows the estimation of the transition probability matrix in the short-, medium-, and long-term. This provides analysts with all the required information about the system of transitions, in order to implement risk analyses and evaluations.

Investors are interested in gaining profit from their investments, but they face many challenges when making decisions due to price fluctuations and unstable financial situations. Financial analysts study the performance of prices and investments using net present value (NPV), internal rate of return (IRR), and return of asset (ROA). Recently, Sabri and Sarsour (2019) studied a new strategy for modeling stock investment valuations by developing the modified internal rate of return. In this strategy, they divided and shared issuance, through split shares and consolidation, as the financial analysts understood the great importance of reducing investment risks and making better informed decisions on future prices. Raheem and Ezepue (2016) predicted the movements of asset returns of a Nigerian Bank by dividing asset returns into three states—positive, moderate, and negative—using the Markov chain model. In the months of May and October, they revealed the fact that their maximum trading cycle was 18 days and their minimum trading cycle was 7 days in February. Additionally, many studies have examined investment behavior (Helms, Salm, & Wüstenhagen, 2020; Qolbi, Karisma, & Rosyadi, 2020).

Furthermore, a considerable number of studies, such as Vázquez-Quintero et al. (2016), Rimal et al. (2018), Ahmed, Kamruzzaman, Zhu, Rahman, and Choi (2013), have used simulations such as forecasting changes in land cover through a simulation technique based on the Markov chain. Kumar, Trehan, and Joorel (2018) used simulation studies to estimate the population mean using stratified random sampling and two auxiliary variables. It can also be used in studies where it is difficult to collect data, such as the relationship between an earthquake and human activities, as in Albano et al.'s study (2017), as well as medical studies that use the Markov chain Monte Carlo, such as Karami et al. (2019), Hamdy, El-Azab, Al-Saeed, Hassan, and Solouma (2017), and Ricci et al. (2019). It can also be used when multiple data patterns exist, and so a simulation comparison of imputation methods for quantitative data is required (Solaro, Barbiero, Manzi, & Ferrari, 2018). The finance industry can also benefit from such a simulation, as seen in Ye, Zhu, Wu, and Miao's study (2016), which employed the Markov financial contagion detection and regime switch quantile with a regression model.

This study attempts to determine the number of companies that are required in order to examine the behaviour of ROA in the entire Malaysian industrial sector using a simulation Markov chain model. Previous similar studies have considered ROA for each company separately; however, as ROA values are yearly, their frequency is too low to study. In other words, if we assume that we want to study ROA performance in the industrial sector between 2007 and 2018, the number of ROA values would be twelve for each company, which is too few to study. Therefore, in

this study, we established the number of companies required to obtain adequate ROA values, which, based on the time period of our study, is 50. Therefore, the number of ROA values is 600.

The proposed Markov chain simulation method was then applied to determine the number of companies that should be involved in order to perform an accurate forecasting analysis on ROAs in the Malaysian industrial sector. The results revealed that, in order to obtain reliable estimates of the parameters in the transition probability matrix, the number of companies should at least be 69, 37 or 24 companies with more than 3, 6, or 10 years of frequency of time, respectively. MATLAB software was used to implement the proposed simulation.

2. METHODOLOGY

2.1. Markov Chain Model

The Markov chain model is widely used in many fields. It is a type of stochastic process that was introduced by Andrey Markov in the 1900s and developed by Kolmogorov in 1936. A fundamental part of the stochastic process is the Markov chains model, in which the occurrence of every event depends only on the past event. The state space is the set of values that the Markov process takes, which may be a discrete or continuous value.

If a sequence $\{X_t,\ t = 1,2, \dots\}$ satisfies the Markov property, it can be expressed as:

$$P\{X_t = j|X_0 = i_0, X_1 = i_1, \dots, X_{t-1} = i\} = P\{X_t = j|X_{t-1} = i\}$$

2.1.1. Transition Matrix and Transition Probability Matrix

The number of parameters observed is $f_{ij}(t)$ with the state i at the $(t\ -\ 1)^{st}$ year, observed in state j at the t^{th} year. Additionally, the transition count matrices of A and L_A, for the combined years, can be obtained using:

$L_A = [f_{ij}]$, where $f_{ij} = \sum_{t=1}^{T} f_{ij}(t)$ (1)

and $f_i(t-1) = \sum_{t=1}^{r} f_{ik}(t),\ k = 1,2, \dots, r$, where r is the number of states.

A Markov process is known as the stationary transition probabilities, if $p_{ij}(t)$ is independent of time t, which is $p_{ij}(t) = p_{ij}$. The maximum likelihood method can be used to estimate the multinomial trials with probabilities of $p_{ij}\ (i,j = 1,2, \dots, r)$, which can be written as:

$$\hat{p}_{ij} = \frac{\sum_{t=1}^{T} f_{ij}(t)}{\sum_{k=1}^{r}\sum_{t=1}^{T} f_{ik}(t)} = \frac{\sum_{t=1}^{T} f_{ij}(t)}{\sum_{t=1}^{T} f_i(t-1)} \quad (2)$$

After obtaining the estimates of the parameters, the transition probability matrix will be expressed as:

$P = [\hat{p}_{ij}]$ where $\sum_{k=1}^{r} \hat{p}_{ij} = 1$ (3)

2.2. Comparison for the Transition Probability Matrices

In this section, we will display some of the measures taken when comparing the transition probability matrices, in order to determine the nearest closed matrix. Some researchers have presented measures for the comparison of credit migration matrices based on eigenvalues, eigenvectors, singular values, or Manhattan and Euclidean metrics

(Jafry and Schuermann, 2004; Trueck and Rachev, 2009; Bangia, Diebold, Kronimus, Schagen, and Schuermann, 2002).

a. Eigenvector Distance Metric

The present study determines whether or not the matrices are equivalent, based on their eigenvectors and regardless of their eigenvalues, which computes a ratio of the matrix norm (Arvanitis, Gregory, & Laurent, 1999). The following equation will be used:

$$\Delta\omega_{AGL}(P_e, P_u) = \frac{\|P_e P_u - P_u P_e\|}{\|P_e\|.\|P_u\|} \quad (4)$$

Where:

P_e = the exact matrix.

P_u = the numerical matrix obtained from simulation results.

If $\Delta\omega \leq 0.08$, the eigenvectors of the matrices are almost similar, whereas the difference in eigenvectors would increase if $\Delta\omega > 0.08$. Furthermore, the eigenvectors are the same when ω equals zero. Although the values of ω range from 0 to 2, the authors do not give a reason why they have chosen a value of 0.08.

b. Metrics Based on Singular Values

We can also compare transition probability matrices using the average of all singular values of the mobility matrix ($\tilde{P}$), as follows:

$$\Delta\omega_{SVD} = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\lambda_i(\tilde{P}_z'\tilde{P}_z)} \qquad z = e, u \quad (5)$$

When $\tilde{P}$ is the mobility matrix, we obtain it using $\tilde{P}$ = transition matrix (P) – identity matrix (I), λ_i is the eigenvalue, and N is the dimension of matrix P.

c. Metrics Based on Eigenvalues

We will illustrate a variety of measures that have been taken to make comparisons between matrices, which depend on eigenvalues (Geweke, Marshall, & Zarkin, 1986). For example:

$$\omega_1(P_z) = 1 - |\lambda_2(P_z)| \quad (6)$$

$$\omega_2(P_z) = \frac{1}{N-1}\left(N - \sum_{i=1}^{N}|\lambda_i(P_z)|\right) \quad (7)$$

$$\omega_3(P_z) = 1 - |det(P_z)| \tag{8}$$

$$\omega_4(P_z) = \frac{1}{N-1}\left(N - trace(P_z)\right) \tag{9}$$

Where $\lambda_2(P_z)$ is the second-largest eigenvalue of P_z, $det(P_z)$ denotes the determinant of P_z, $trace(P_z)$ denotes the trace of matrix P_z, and $z = e, u$.

d. Manhattan and Euclidean Distance Metrics

These measures are popular approaches to comparing two matrices using the cell by cell distance technique. The Euclidean metric calculates the average root mean square difference, whereas the Manhattan metric calculates the average of the absolute difference between the corresponding elements of the matrices. Specifically,

$$\omega_{Euclidean}(P^z) = \frac{\sqrt{N-1}}{N}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(P_{i,j}^z - I_{i,j})^2} \tag{10}$$

$$\omega_{Manhattan}(P^z) = \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|P_{i,j}^z - I_{i,j}\right| \tag{11}$$

In order to see the closed or nearest of the two matrices, we need to take the above measures to calculate the difference between P_e and P_u.

$$\Delta\omega_l(P_e, P_u) = \omega_l(P_e) - \omega_l(P_u) \tag{12}$$

Where l denotes the type of metric.

3. SIMULATION MODEL

The model must contain a sufficient sample size with a high enough frequency of time to ensure the correct estimation of the model's parameters. The present study, however, has a shorter frequency time. Increasing the sample size would be one solution to this problem. However, to what extent should the sample size be increased in order to obtain reliable estimates of the parameters of the transition matrix? In order to suitably increase the sample size, a novel simulation technique was proposed and implemented, which is summarized in the following steps:

1. Generate a random number group: W for q times, where q is the sample size and W is uniformly distributed $W \sim U(0,1)$.

$$W = \{s_x\} \quad , x = 1,2,\ldots,q \quad where\ s_x \in (0,1) \tag{13}$$

where q is the sample size for the W group.

2. Let I be the initial matrix, where $I = [I_1 \ I_2 \dots I_r]$ and r is the number of states which $I_1 = I_2 = \cdots = I_r$. Based on I, we can define W for each state, based on the following criterion: if $W < I_1$, it stays in state Y_1; if $I_1 \leq W < I_1 + I_2$, it stays in state Y_2, and so on, until it reaches the final state, Y_r; (Ross, 2019).

$$Y_i = \begin{cases} Y_1, & s_x < I_1 \\ Y_2, & I_1 \leq s_x < I_1 + I_2 \\ \vdots & \\ Y_r, & \sum_{x=1}^{r-1} I_x \leq s_x < \sum_{x=1}^{r} I_x \end{cases} \quad (14)$$

3. Generate a new discrete random number group, G, for each state from 1 to r, according to the number of elements in $Y_1, Y_2, \dots, Y_r$, respectively, which is then uniformly distributed over the interval (0,1), as follows:

$$G_i = \{o_x\} \quad , i = 1,2, \dots, r \text{ and } x = 1,2, \dots, u \quad (15)$$

where u is the number of elements in $Y_i, i = 1,2, \dots, r$, $s_x \in (0,1)$.

4. After G_m is obtained, let P be the transition probability matrix where $P = \begin{bmatrix} p_{11} & \cdots & p_{1r} \\ \vdots & \ddots & \vdots \\ p_{r1} & \cdots & p_{rr} \end{bmatrix}$. Based on the first row of P, the movement from state i to another state or remaining in state i, can be determined as follows:

$$Y_{ij} = \begin{cases} Y_{i1}, & o_x < p_{i1} \\ Y_{i2}, & p_{i1} \leq o_x < p_{i1} + p_{i2} \\ \vdots & \\ Y_{ir}, & \sum_{i=1}^{r-1} p_{ir} \leq o_x < \sum_{i=1}^{r} p_{ir} \end{cases} \quad (16)$$

$where \ i = j = 1,2, \dots, r.$

5. Following on from the previous step, obtain the number of elements in each state by gathering the elements of Equation 16, which displays the number of movements to the same state:

$$H_j = Y_{1j} + Y_{2j} + \cdots + Y_{rj} \quad (17)$$

For simplicity, Equation 17 can be expressed as:

$$H_j = \sum_{i=1}^{r} Y_{ij} \quad , \quad j = i = 1,2, \dots, r \quad (18)$$

6. Generate a new discrete random variable for each state, based on a new number of elements from Equation 18 using a uniform distribution in the interval (0,1), i.e. the number of a discrete random variable of state r is H_r.

7. Repeat steps 3, 4, 5, and 6 k-times after updating the number of elements in each state, where k is the frequency time of the study.

8. The performance measures from $\Delta\omega$ will be used to identify whether or not the sample size is sufficient. Accordingly, if $\Delta\omega$ is less than 0.08, it means that the exact matrix is equivalent to the numerical matrix, which means that the sample size is sufficient. However, if $\Delta\omega$ is higher than 0.08, it means that the matrices are not equal. Therefore, the sample size needs to be increased until $\Delta\omega$ is lower than 0.08, based on the eigenvectors. If it is based on other metrics, we will assume a critical value $(\alpha = 0.1)$.

Figure 1 shows the simulation of the Markov chain model. This model generates r independent sample paths beginning with W, and $Y_{t_k}^i$ denotes a generic node at time t_k in the ith path (Raychaudhuri, 2008).

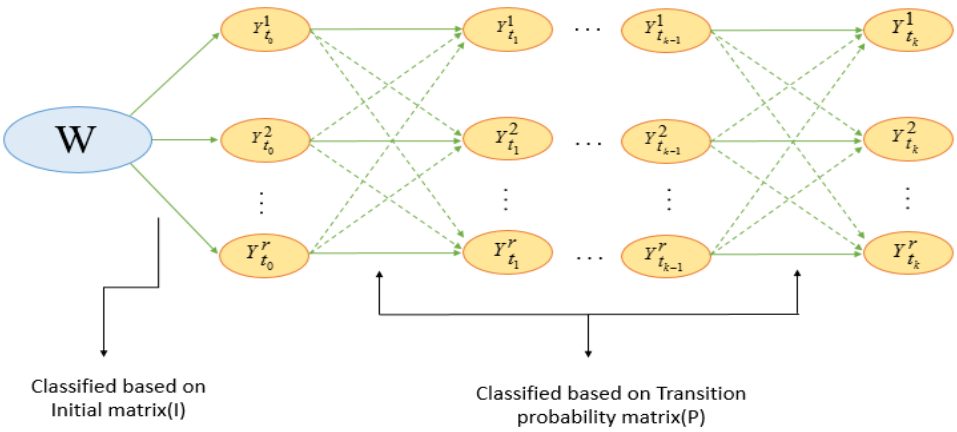

Figure 1. A simulation of the Markov chain model.

4. RESULTS OF SIMULATION AND DISCUSSION

4.1. Eigenvectors

To obtain a sufficient sample size for two states $S = \{1,2\}$ in three varying situations (short-, medium-, and long-term) we shall assume that the initial and transition probability matrices I and P of our Markov chain are $I = [0.5 \quad 0.5]$, and $P = \begin{bmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{bmatrix}$, respectively.

Figure 2 shows the simulation results of the short-term experiment reaching a sufficient sample size to estimate the transition matrices. Typically, the experiment has two states, a frequency time of three years, and is replicated 1000 times. This experiment was performed using different sample sizes. About 63% of the replications are achieved at a $\Delta\omega$ less than 0.08 when the sample size is 15, which indicates that the overall matrices have a high level of differences, meaning that the sample size is insufficient. Using a sample size of 24, 30, 39 and 50 failed to achieve more than 95% of the trials, and therefore, we were unable to acquire a sufficient sample size. However, when the sample size was 69 and above, more than 95% of the trials had performance measures that were less than 0.08, thereby achieving a sufficient sample size. This suggests that studies that involve a three-year frequency time should have a sample size of at least 69. Hence, once the years are merged, the number of movements becomes 207.

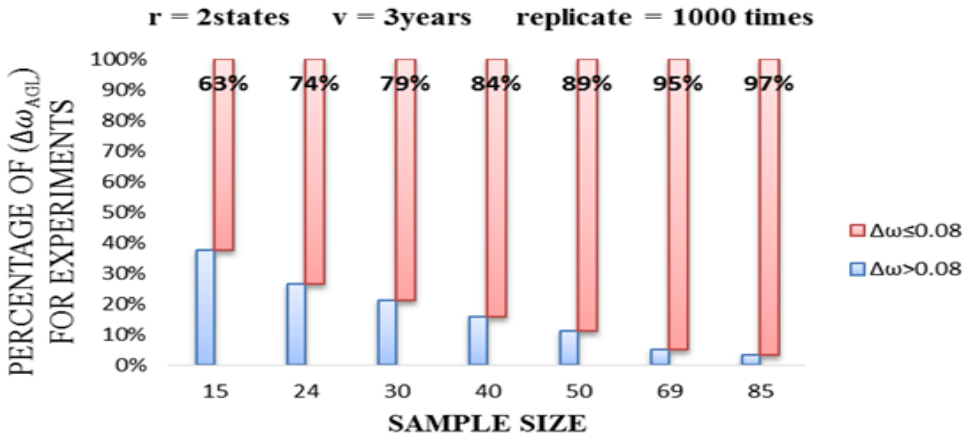

Figure 2. Simulation for the medium-term (three years).

The experiments were performed for the medium-term frequency, using two states that have six years of frequency time and were replicated 1000 times. As seen in Figure 3, about 78%, 88% and 91% of the replications were achieved at a $\Delta\omega$ that was less than 0.08 when the sample size was 15, 24 and 30, respectively. This indicates that the overall matrices have a high level of difference, meaning that these experiments failed to reach the required sample size. On the other hand, having a sample size of 37 or more resulted in more than 95% of the trials having performance measures that were less than 0.08. Therefore, one should select a sample size of at least 37 when the studies involved have data that spans six years. Once all the years were merged, the number of movements was 222.

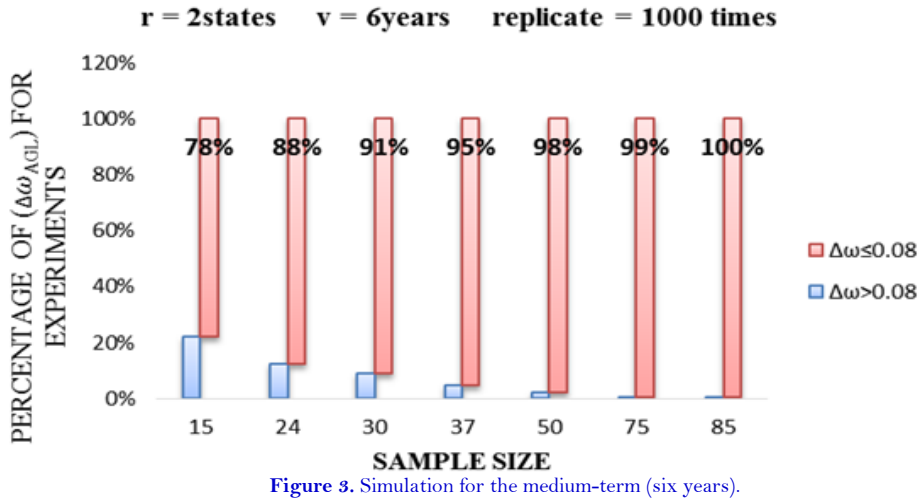

Figure 3. Simulation for the medium-term (six years).

The experiments were also performed to examine long-term frequency, with two states, a frequency time of ten years, and 1000 replications. The simulation results are shown in Figure 4, and it is indicated that about 88% of replications have been achieved with a $\Delta\omega$ that is less than 0.08 when the sample size is 15, which indicates that it failed to reach a sufficient sample size. However, when the sample size is 24 or more, 95% of the trial performances measured less than 0.08, meaning that they have achieving a sufficient sample size. This means that studies that involve three years of frequency time should have a sample size of at least 24. After all the years are merged, the number of movements becomes 240.

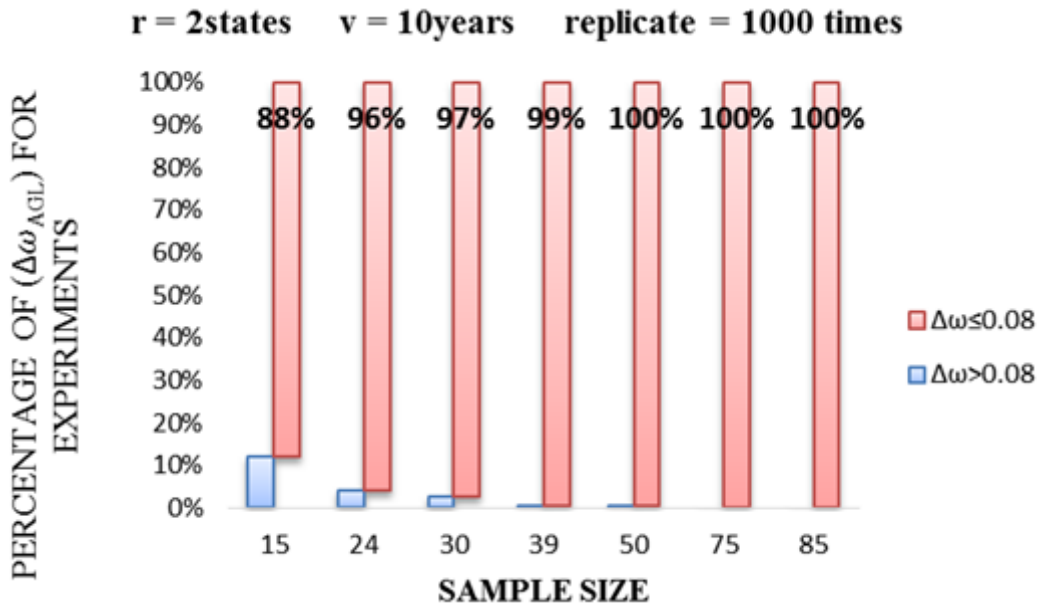

Figure 4. Simulation for a long-term (ten years).

4.2. Result Based on Singular, Manhattan and Euclidean Distance Metrics

In order to evaluate the differences between matrices P_e and P_u, we chose critical value α $(say\ \alpha = 10\%)$ to calculate the error in the estimate transition matrix using the above metrics, which allowed the error to be between ± 0.1.

In this section, we will only discuss singular metrics because the result of a singular metric, similar to the Manhattan and Euclidean distance metrics, will determine a sufficient sample size for the transition matrix estimation.

Figure 5 shows the simulation results over three years, in order to evaluate the amount of error in the difference between P_e and P_u through a singular metric that reaches a sufficient sample size and is able to estimate the transition matrices. Typically, this experiment uses two states, has a frequency of time of three years, and is replicated 1000 times. This experiment was performed using different sample sizes. The result demonstrates the fact that, when sample sizes of 10, 20, 30, 40, and 50 are used, they are insufficient to estimate the transition matrix because the errors are more than 0.1; however, when a sample size of 60 or more is used, the trials had an error that was less than $\alpha = 0.1$, which is sufficient. We can also see that, when the sample size increases, the box plot will approach zero.

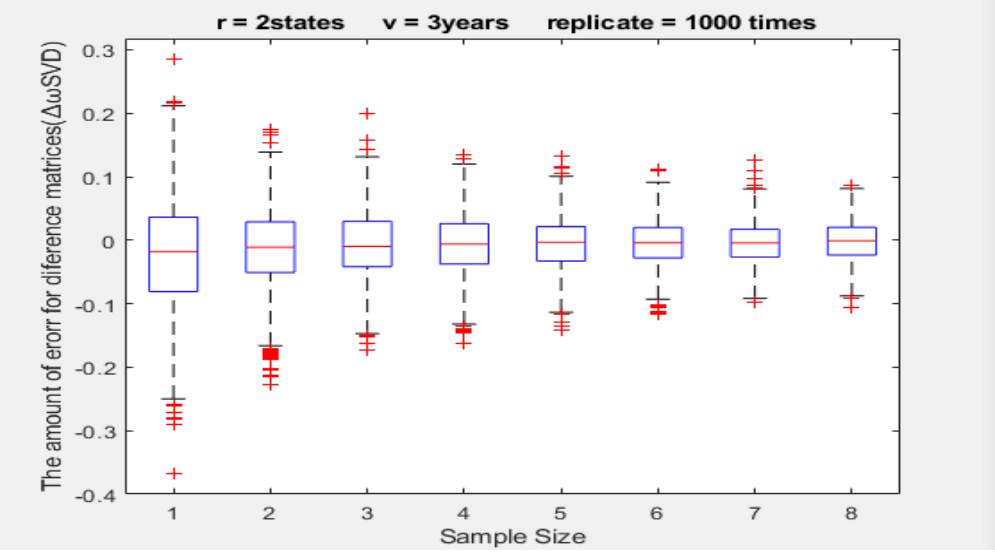

Figure 5. Simulation for the short-term (three years).

The simulation results in Figure 6 present an experiment that uses two states, has a frequency of time of six years, and is replicated 1000 times to evaluate the error. This differentiates between matrices using the value of a singular metric in order to reach a sufficient sample size, so that the transition matrices can be estimated. Typically, the experiment is performed using different sample sizes. However, 10 and 20 were considered to be insufficient sample sizes to determine the transition matrix because the error would still be more than $\alpha = 0.1$. On the other hand, using a sample size of 31 or more will result in the trials having errors that are less than $\alpha = 0.1$. Hence, having 31 samples or more is considered to be a sufficient sample size.

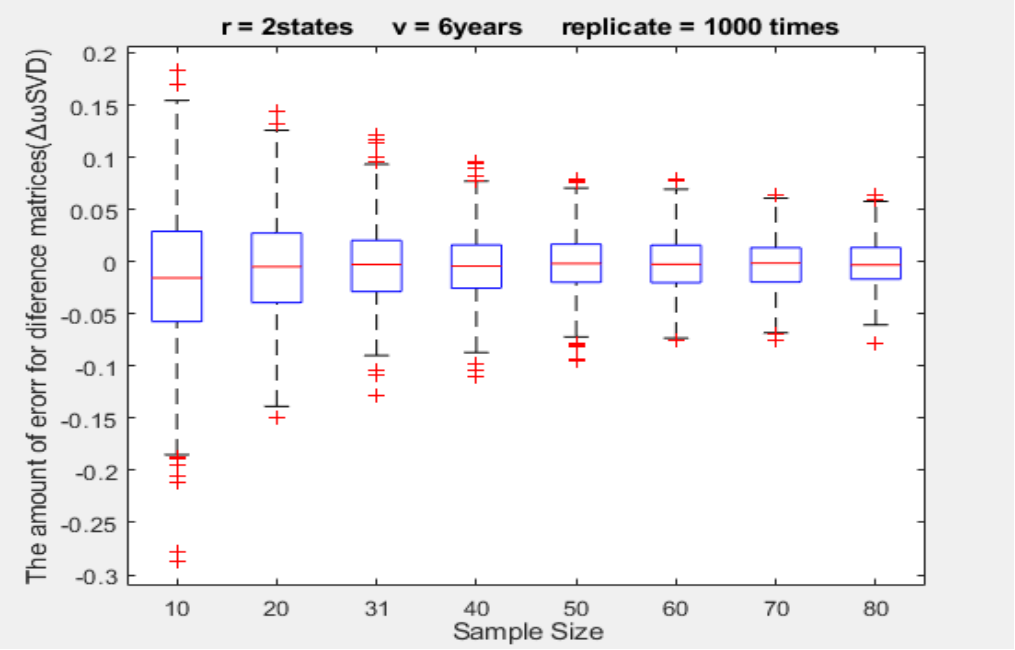

Figure 6. Simulation for the medium-term (six years).

The experiments were performed for the medium-term using two states, a frequency of ten years, and being replicated 1000 times. As seen in Figure 7, the sample size is ten or less, which indicates that the overall matrices have a high level of differences. That means that these experiments have failed to reach the required sample size. On

the other hand, having a sample size of 20 or more resulted in less than 10% error. Therefore, one should select a sample size of at least 20 when the study involves data that spans ten years.

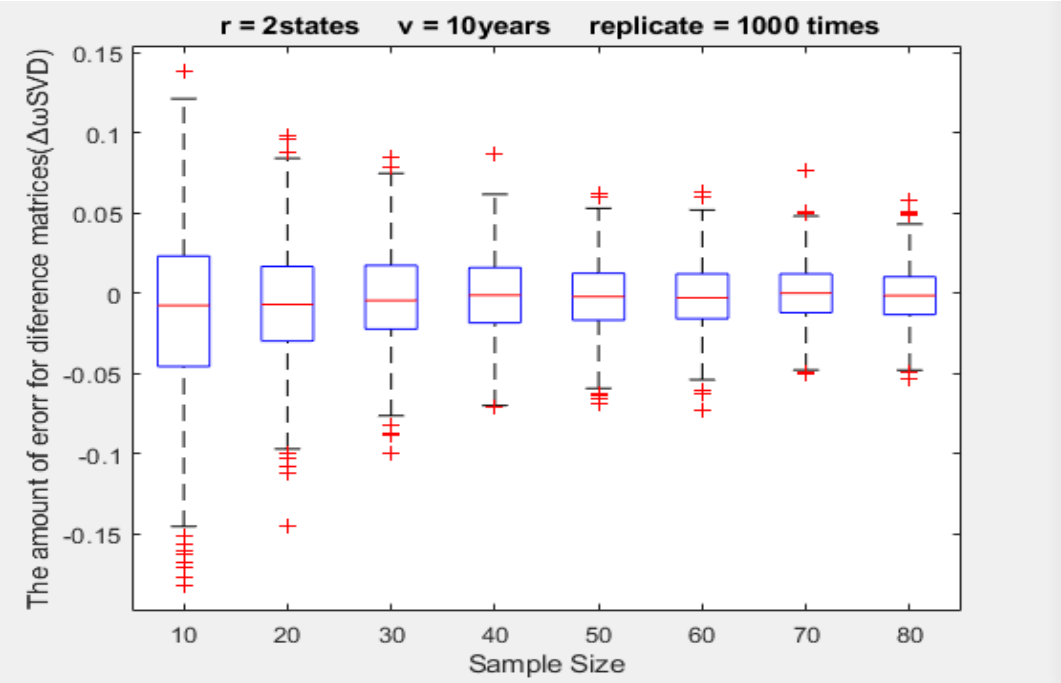

Figure 7. Simulation for a long-term (ten years).

We can conclude that a sufficient sample size can be achieved rapidly as we increase the frequency of time and decrease the number of states. The results of the simulation of the adequate sample size can be determined using the above metrics. Typically, simulations with a higher frequency of time result in adequate sample sizes with lower values.

5. ILLUSTRATION OF A TWO-STATE MARKOV MODEL THAT FORECASTS CHANGES TO RETURN ON ASSETS (ROA)

This section provides an example to illustrate how the proposed simulation method achieves a sufficient sample size to obtain reliable estimates of the transition matrix based on the eigenvector's metric. In this example, the frequency of time is short. Empirical calculations were based on studying the performance of investments through ROAs, from the state $(t-1)$ to another state (t). These two states were studied in each company operating in the Malaysian industrial sector between 2009 and 2018, where data on *ROA* were collected on a yearly basis. Hence, the main problem in this example is the relatively small amount of data for each company, which requires us to perform the proposed simulation method: State one for a negative *ROA* and state two for a positive *ROA* or a ROA equal to zero in the long-term.

$$Z_t = \begin{cases} 1, if\ ROA < 0 \\ 2, if\ ROA \geq 0 \end{cases} \qquad t = 1,2,\dots,T \tag{19}$$

Where *ROA* is an indicator that measures the company's profitability relative to its total assets, which demonstrates the efficiency of the administration in using its assets to achieve higher profits. The *ROA*, sometimes referred to as the return on investment, is calculated as a percentage by dividing the company's annual profits by the total of its assets.

The transition matrices of successive pairs (T-1) for the *ROA* of 147 companies between 2007 and 2018 have been calculated using Equation 1 and are shown in Table 1.

Table 1. Transition count matrices for each pair between 2007 and 2018.

	2007-2008		2008-2009		2009-2010		2010-2011		2011-2012		2012-2013	
	State1	State2	State1	State1	State1	State1	State2	State2	State2	State2	State1	State2
State1	17	7	19	12	16	24	12	10	18	6	19	18
State2	14	109	21	95	6	101	12	113	19	104	16	94
	2013-2014		2014-2015		2015-2016		2016-2017		2017-2018			
	State1	State2	State1	State1	State1	State1	State2	State2	State2	State2		
State1	21	14	25	12	25	15	22	16	28	9		
State2	16	96	15	95	13	94	15	94	23	87		

After combining data from all the years between 2007 and 2018, the transition count matrix of *ROA* will be acquired using Equation 3.

$$L_{ROA} = \begin{bmatrix} 222 & 143 \\ 170 & 1082 \end{bmatrix}$$

The transition count matrix was checked using the stationarity test. We obtained the maximum likelihood estimates of transition probability P of *ROA*.

$$P_{ROA} = \begin{bmatrix} 0.6082 & 0.3918 \\ 0.1358 & 0.8642 \end{bmatrix}$$

5.1. Long-term Performance

Based on the simulation results derived from Section 4, we have to select at least $n = 24$ companies to study the long-term performance of $v = 10$ years. Therefore, a random sample of 24 companies was selected from the same sector, in order to study the behavior of *ROA* over ten years to check the accuracy of the proposed simulation.

The transition probability matrix after carrying out the stationarity test on nine pairs is as follows:

$$P_{(v=10)} = \begin{bmatrix} 0.6140 & 0.3860 \\ 0.1635 & 0.8365 \end{bmatrix}$$

Through a calculation of Equation 4, $\Delta\omega = 0.0183$ is less than 0.08, meaning that the transition matrices for the 147 companies and 24 companies were very close.

5.2. Middle-Term Performance

For the middle-term, random companies of $n = 39$ were selected with $v = 6$ years. The transition probability matrix was calculated as:

$$P_{(v=6)} = \begin{bmatrix} 0.6129 & 0.3871 \\ 0.1037 & 0.8963 \end{bmatrix}$$

After calculating the Eigenvector distance metric between the sector matrix and 39 companies, the matrix was $\Delta\omega = 0.0182 < 0.08$, which indicates that the two matrices were very close.

5.3. Short-Term Performance

In the short-term of $v = 3$ years, we randomly chose $n = 75$ companies based on the simulation results, where the transition probability matrix was calculated as:

$$P_{(v=3)} = \begin{bmatrix} 0.6522 & 0.3478 \\ 0.1339 & 0.8661 \end{bmatrix}.$$

After comparing three years of the sector matrix of *ROA* using the eigenvector distance metric $\Delta\omega = 0.0080$, the matrices were found to be close.

Therefore, we can conclude that all matrices in all three cases were very close to each other, meaning that the result of the simulation was acceptable.

6. CONCLUSION

Sometimes a system of transition may have data that have been obtained over a short frequency of time. The present study aimed to gain reliable estimates of the transition probability matrix. Therefore, a Markov chain simulation method was developed to obtain the required sample size needed to achieve the aims of the study. Three variant models were used to determine the required sample sizes (short-term, medium-term, and long-term). Finally, the application of the proposed simulation method has been demonstrated using yearly ROA data from companies operating in the Malaysian industrial sector. The results in all three cases were close.

Funding: This study received no specific financial support.
Competing Interests: The authors declare that they have no competing interests.
Acknowledgement: The authors would like to express their sincere thanks to the editor and anonymous reviewers for their time and valuable suggestions. This research was supported, in part, by the School of Mathematical Sciences at Universiti Sains Malaysia.

REFERENCES

Abidin, S. N. Z., & Jaffar, M. M. (2014). Forecasting share prices of small size companies in Bursa Malaysia using geometric Brownian motion. *Applied Mathematics & Information Sciences*, *8*(1), 107-112. Available at: https://doi.org/10.12785/amis/080112.

Ahmed, B., Kamruzzaman, M., Zhu, X., Rahman, M., & Choi, K. (2013). Simulating land cover changes and their impacts on land surface temperature in Dhaka, Bangladesh. *Remote Sensing*, *5*(11), 5969-5998. Available at: https://doi.org/10.3390/rs5115969.

Albano, M., Polcari, M., Bignami, C., Moro, M., Saroli, M., & Stramondo, S. (2017). Did anthropogenic activities trigger the 3 April 2017 Mw 6.5 Botswana earthquake? *Remote Sensing*, *9*(10), 1028-1040.

Anderson, T. W., & Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics*, *28*(1), 89-110. Available at: https://doi.org/10.1214/aoms/1177707039.

Arvanitis, A., Gregory, J., & Laurent, J.-P. (1999). Building models for credit spreads. *The Journal of Derivatives*, *6*(3), 27-43.

Asmussen, S. (2008). *Applied probability and queues* (Vol. 51). New York: Springer Science & Business Media.

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., & Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, *26*(2-3), 445-474. Available at: https://doi.org/10.1016/s0378-4266(01)00229-1.

Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, *38*(1), 34-105.

Geweke, J., Marshall, R. C., & Zarkin, G. A. (1986). Mobility indices in continuous time Markov chains. *Econometrica: Journal of the Econometric Society*, *54*(6), 1407-1423. Available at: https://doi.org/10.2307/1914306.

Goldfeld, S. M. (1973). A Markov model for switching regression. *Journal of Econometrics*, *1*(1), 3-15.

Hamdy, O., El-Azab, J., Al-Saeed, T. A., Hassan, M. F., & Solouma, N. H. (2017). A method for medical diagnosis based on optical fluence rate distribution at tissue surface. *Materials*, *10*(9), 1-13. Available at: https://doi.org/10.3390/ma10091104.

Helms, T., Salm, S., & Wüstenhagen, R. (2020). Investor-specific cost of capital and renewable energy investment decisions. *Renewable Energy Finance: Funding the Future of Energy*, 85-111.

Jafry, Y., & Schuermann, T. (2004). Measurement, estimation and comparison of credit migration matrices. *Journal of Banking & Finance*, *28*(11), 2603-2639. Available at: https://doi.org/10.1016/j.jbankfin.2004.06.004.

Karami, M. A., Fakhri, Y., Rezania, S., Alinejad, A. A., Mohammadi, A. A., Yousefi, M., . . . Ahmadpour, M. (2019). Non-carcinogenic health risk assessment due to fluoride exposure from tea consumption in iran using monte carlo simulation. *International Journal of Environmental Research and Public Health*, *16*(21), 1-12. Available at: https://doi.org/10.3390/rs9101028.

Kumar, S., Trehan, M., & Joorel, J. S. (2018). A simulation study: Estimation of population mean using two auxiliary variables in stratified random sampling. *Journal of Statistical Computation and Simulation*, *88*(18), 3694-3707.

Mettle, F. O., Quaye, E. N. B., & Laryea, R. A. (2014). A methodology for stochastic analysis of share prices as Markov chains with finite states. *Springer Plus*, *1*(3), 1-11.

Qolbi, F. A., Karisma, D. P., & Rosyadi, I. (2020). Macro variable effect analysis and non-performing financing (npf) against the return on asset (roa) islamic banks in Indonesia year 2008-2017. *Journal of Islamic Economic Laws*, *3*(1), 32-47. Available at: https://doi.org/10.23917/jisel.v3i1.10170.

Raychaudhuri, S. (2008). *Introduction to monte carlo simulation*. Paper presented at the Paper Presented at the 2008 Winter Simulation Conference.

Ricci, R., Kostou, T., Chatzipapas, K., Fysikopoulos, E., Loudos, G., Montalto, L., . . . David, S. (2019). Monte carlo optical simulations of a small FoV gamma camera. Effect of scintillator thicknesses and septa materials. *Crystals*, *9*(8), 1-14. Available at: https://doi.org/10.3390/cryst9080398.

Rimal, B., Zhang, L., Keshtkar, H., Haack, B. N., Rijal, S., & Zhang, P. (2018). Land use/land cover dynamics and modeling of urban land expansion by the integration of cellular automata and markov chain. *ISPRS International Journal of Geo-Information*, *7*(4), 154-175.

Ross, S. M. (2019). *Introduction to probability models* (Vol. 12). United States of America: Academic Press. University of Southern California Los Angeles, CA.

Sabri, S. R. M., & Sarsour, W. M. (2019). Modelling on stock investment valuation for long-term strategy. *Journal of Investment and Management*, *8*(3), 60-66. Available at: https://doi.org/10.11648/j.jim.20190803.11.

Sarsour, W. M., & Sabri, S. R. M. (2020). Forecasting the long-run behavior of the stock price of some selected companies in the Malaysian Construction Sector: A Markov Chain approach.

Solaro, N., Barbiero, A., Manzi, G., & Ferrari, P. (2018). A simulation comparison of imputation methods for quantitative data in the presence of multiple data patterns. *Journal of Statistical Computation and Simulation*, *88*(18), 3588-3619.

Trueck, S., & Rachev, S. T. (2009). *Rating based modeling of credit risk: Theory and application of migration matrices*: Academic press.

Vázquez-Quintero, G., Solís-Moreno, R., Pompa-García, M., Villarreal-Guerrero, F., Pinedo-Alvarez, C., & Pinedo-Alvarez, A. (2016). Detection and projection of forest changes by using the Markov chain model and cellular automata. *Sustainability*, *8*(3), 1-13. Available at: https://doi.org/10.3390/su8030236.

Ye, W., Zhu, Y., Wu, Y., & Miao, B. (2016). Markov regime-switching quantile regression models and financial contagion detection. *Insurance: Mathematics and Economics*, *67*, 21-26. Available at: https://doi.org/10.1016/j.insmatheco.2015.11.002.

Zhang, D., & Zhang, X. (2009). Study on forecasting the stock market trend based on stochastic analysis method. *International Journal of Business and Management*, *4*(6), 163-170. Available at: https://doi.org/10.5539/ijbm.v4n6p163.