ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 42(2022) 123-154

Development of Hidden Markov Model for Stock Market Trading

Tirupathi Rao Padi, Gulbadin Farooq Dar and Sarode Rekha
[Received on September, 2022. Accepted on January, 2023]

ABSTRACT

Selling or buying a share in a stock market is now a matter of scientific decision-making mechanism. Investors are badly in need of the advice of mathematical model developers for understanding the pulse of the market. It leads to more attention from model builders to assess the market behavior. Hidden Markov models (HMMs) are inherent structures to spell out the linkage between invisible influencing factors on visible resulting states. Finance investment processes with two hidden and two visible states are formulated for obtaining the model and exploring the dynamics of gain/pain in the stock value. A generalized mechanism for a sequence of 'n' transactions with two visible states is modeled. Probability distributions for the single, of length two and three visible states are derived. Mathematical formulae for different statistical measures and various generating functions for the corresponding probability distributions are obtained in this paper. Understanding the model's notion from a layman's point of view, an empirical study was carried out with real-time historical stock market data of Reliance Industries limited. The pivotal objective of this study is to estimate the model parameters that measure the holistic changes in money values of different shares in a stock market. The finance and portfolio managers can make use of this study for designing optimal resource planning by understanding the relations of invisible factors to visible states. The development of user interface dynamic dashboards will make this work more popular and reachable to the community at large.

1. Introduction

A stock market is a market place where people can buy, sell and trade in publicly listed company's stock. It offers a framework for frictionless securities trading. It is important to note that a person can only trade in the stock market through a registered stockbroker through an electronic means (Demat form). Shares are

✉ : Gulbadin Farooq Dar
Email: gulbadinstst.pu@gmail.com

Extended authors information available after reference list of the article.

purchased or sold with the intention of favorable rate of return from the investment the investors made in the stock market. A rising stock market is a solid indication of the country's economic growth. An increase in share price is linked to an increase in investments, whereas a reduction in share price decreases in investments in the stock market. A stock exchange is a place or organization where stock traders (including individuals and businesses) can trade equities either buy or sell. A significant corporation's stock is frequently traded on a number of different exchanges throughout the world and in India we have the two largest stock exchanges namely, the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). In terms of market capitalization, the BSE and NSE are among the top five stock exchanges in the world for developing economies. The normal phenomenon in financial markets is that asset values are constantly changing and regardless of how well-informed an investor is having, the market probability can never be eliminated.

The stock price is theoretically determined by its closing share prices that is by observing the supply and demand of shares; however, there are a variety of hidden factors that influence the stock price other than the supply and demand of the shares in the stock market, including, economic factors, performance of the company, changes in county’s currency value, military factors, change in covid cases, political stability of the country, investor faith on stock market, amount of imports and exports of the country etc. Predicting the value of a stock offers substantial arbitrage profit opportunities, which is a major driver of this study. We have used the HMM to analyse and forecast the stock market trend of Reliance Industries limited in this research. Since the variety of factors influence the stock value of Reliance industries Ltd. in which the fluctuation of the INR (Indian Rupee) against the USD (US Dollar) is assumed to be a hidden state/influencing factor which influence the stock price of Reliance Industries Limited. The two states of INR/USD such as INR decrease (D) and INR increase (I) are the two hidden states and the fall (F) and rise (R) in the closing share prices of Reliance Industries Ltd. are the two visible states in our two-by-two hidden Markov model. The main objective in this paper is to analysis and forecast the share price of Reliance industries which is influenced by the change in INR with respect US dollar.

1.1. Literature Review

Hidden Markov model (HMM) is a statistical prediction model that estimates, analysis and predicts hidden states in terms observational/visible data. Every observation\visible state of an HMM has a specific probability distribution corresponding to a possible hidden state. Baum *et al.* (1966) first proposed HMM theory in the late 1960s, and Baum and Petrie created its mathematical foundations. Again, Baum *et al.* (1970) developed a maximizing approach where a single observation is used to calibrate the HMM parameters. Levinson *et al.* (1983) proposed an MLE approach for HMM that incorporated repeated observation training under the premise that all of the observations are independent and Li, *et al.* (2000) proposed an HMM training for different mutually dependent observations.

In order to analysis and forecast the stock market price for any profitable company, investors, and the shareholders in making a safe and confident decision to make a successful stock trading in the stock market Adam (2009), has made it an attractive and popular research area. Given the difficulty of making a meaningful prediction based on publicly available historical data, we must be asked: how can historical data be used to make a meaningful prediction about future stock market behaviour? Initially, technical and fundamental analysis presented a solution by Butler and Malaikah (1992) on the basis of assumption that previous data is sufficient to anticipate future behavior. Choi (1999) presented the degrees of association between share prices and all available data are measured in real-time according to the Efficient Market Hypothesis (EMH). The applications of a stochastic process to investigate stock market trend behaviour is not new. Using the stochastic model, several research articles have been published to anticipate and analyze share market moment at various times. Recently, GF Dar *et al.* (2022) applied Markov chain model (MCM) in order to analyze and predict the stock market movement of Tata Consultancy Services Ltd (TCS Ltd). They observed that the probabilities for the state of high loss in share prices of TCS are less than that the state of high gain in the long run recommended that investing in TCS share is a good choice of investment for investors to make capital gain. Huang (2015) introduced an absorbing Markov chain to create a Markov chain model to study the stock price volatility of Taiwanese corporation HTC. T R Padi, *et al.* (2022) used MCM for the stock market trend analysis and prediction in the context of Indian stock market. Real-time data of the daily closing share price was obtained from the historical price of NSE (National Stock Exchange) for Nifty banks to conduct a detailed and exploratory data analysis. Again, T R Padi, *et al.* (2022) provides a discrete-time MCM to analyze the behaviour of stock market prices with reference to State Bank of India (SBI) which is one of the leading commercial banks in India. The analysis is carried out on the basis of past three years daily historical real time data of closing share prices. Doubleday and Esunge (2011) applied discrete time Markov chain, to identify the diversified portfolio of shares and stock markets as a whole on the Dow Jones Industrial Average (DJIA).

Markov chain model is used only when the states under study are directly visible to us. However, it sometimes happens that the stock prices are influenced by many hidden influencing factors such as country's currency value, political factors, socio economic factors, military factors, imports and exports of the nation, political factors etc. In an HMM, these influencing factors are called as hidden states and the predictions about the visible states are made on the basis of these hidden factors. Hassan and Baikunth Nath (2005) employed HMM to anticipate stock price for an interconnected market. Nguyen (2018) used the AIC, BIC, and HQ information criteria to find the ideal number of states for the HMM, as well as discuss the applicability of HMM in stock trading. Kritzman *et al.* (2012) employed a two-state HMM to forecast market volatility, inflation, and the company's production index. Guidolin and Timmermann (2007) applied HMM with four hidden states and multiple visible states to investigate asset allocation decisions on the basis of regime change in asset returns. Ang and Bekaert (2002) used the regime shift model for the allocation of international asset. Again, Nguyen and Nguyen (2015) used HMM for single visible data to forecast the regimes of some economic indicators and to select equities based on their performance during the projected schemes. Gupta *et al.* (2012) applied HMM in order to predicting stock closing prices utilizing multiple observation data such as open, close, low and high. HMM was utilized by Poonam Somani *et al.* (2014) to find out the patterns from previous data sets that matched the current day's stock price behavior. Tuyen (2013) used a normally distributed HMM to get the optimal MCM for the real time data by applying it to historical VN-Index data. Holzmann *et al.* (2016) studied the number of states in the HMM and determined the state with the most fluctuation corresponding the recent financial crisis. Again, Liu *et al.* (2017) employed an HMM in order to determine the time changing distribution of the returns in Chinese stock market since 2005. Later, Fu *et al.* (2018) used an HMM to conduct a quantitative timing research on the CSI 300 Index, and they were successful in recognizing market status and generating positive results. In the same year, Chen and Xia (2018) proposed a piecewise continuous HMM temporal probability strategy to address the market's commodity classification problem. HMM was utilized by Eun-Chong Kim *et al.* (2019) to detect the stages of specific assets and to suggest an investing strategy based on price movements. Recently, EO Amiens, IO Osamwonyi (2022) used HMM with single observation to estimate the closing share price of the selected manufacturing companies from the Nigerian Stock Exchange (NSE). They used the real time data from 22 Nov, 2013 to 6 July 2018 and the data were partitioned into two datasets for training and testing purpose. From the results obtained, they suggested that HMM should be adopted in practice for the stock price analysis and forecasting. Again, GF Dar *et al.* (2022) developed an HMM for a proper understanding of finance variables in the stock market. Stochastic modelling with HMMs is carried out for exploring various parameters of the model. It is observed from the empirical real time historical closing share price data analysis that there is the maximum likelihood of rising the share prices of HDFC bank in consecutive two days and the system will be on the rising state from the 19^{th} day onwards. Qingqing Chang and Jincheng Hu (2022) made the use of an assumption that the hidden state of the HMM is a first-order Markov process and proposes a one-step method to predict the next hidden state of a financial time series when the sequence has two types of hidden states. From result of the hidden state, they also proposed to choose different models to predict its value according to the different amount of financial time series sample data.

From the existing literature, it is observed that the probability of the next day share prices and then steady state probabilities using MCM and an HMM are obtained. However, in this paper, the mathematical expressions for the probability distributions for both the visible states V_1 and V_2 (which represents the falling and rise in the share price of Reliance Industries Ltd.) with one day, two-day and three days sequence are derived. The formulae for the statistical measures like mean, variance, coefficient of variation, skewness and kurtosis along with the generating functions such as moment generating function, probability generating function and the characteristics function of corresponding proposed probability distribution are also derived which is not observed in the existing literature. In order to forecast the future behaviour of both the hidden and the visible states, the steady state or the stationary probability distributions are calculated from the real time historical closing share price data of Reliance industries Limited.

1.2. Discrete-time Markov chains and HMMs

A Markov chain model (MCM) is a random or stochastic model that displays a succession of possible events, with the probability of each event exclusively determined by the state obtained in the instantaneous previous event. The Markov chain means that given X_t (the current state) the state X_{t+1} (future state) depends only upon X_t but not on the previous $X_{t-1}, X_{t-2}, X_{t-3}, \ldots X_1, X_0$. Because of its Markovian features, powerless interest in accurate information, and predicting behavior with many preferences, the Markov model is important in statistics. Mathematically, the Markov property or the memoryless property represents that if $X_n, n = 0,1,2,\ldots t+1$ is a stochastic process with discrete state space S, then the Markov property can be formulated in equation (1.2.1).

$$P(X_{t+1} = x_{t+1} \mid X_t = x_t, X_{t-1} = x_{t-1}, \ldots, X_1 = x_1, X_0 = x_0) = P(X_{t+1} = x_{t+1} \mid X_t = x_t) \qquad (1.2.1)$$

For all, $\text{t} = 0, 1, 2, 3, \ldots$ and for all states $x_0, x_1, \ldots, x_t, x_{t+1}$. The state-space of the process is a countable set S containing all of X_i's possible values. The condition of the any state of the system may change over time. The transition probability, represented as a_{ij}, is the likelihood that the process will move from state i in the n^{th} step to state j in the $(n+1)^{th}$ step. Hence $a_{ij} = P(X_{n+1} = x_{n+1} \mid X_n = x_n)$ for all $i, j \in S$ and $n \geq 0$, $1 \leq i, j \leq n$.

An HMM is a statistical Markov model where the system under consideration is assumed to be a Markov process, which we call it say X, with hidden, or invisible states. It is assumed that there is another process, we call it Y, whose movement is influenced by the hidden process X. Suppose X_n and Y_n be two discrete-time stochastic processes such that $\text{n} \geq 1$, then the pair $\{X_n, Y_n\}$ is an HMM if X_n follows Markov property (1.2.1) whose trend behaviour is not directly visible and the process Y_n follows the following equation (1.2.2).

$$P(Y_m = y_m / X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = P(Y_m = y_m / X_n = x_n) \qquad (1.2.2)$$

or every $n \geq 1$. In our probabilistic model, HMM allows us to discuss and talk about both the visible and hidden events that we think of as causal factors. Therefore, HMM can be defined as a stochastic model in which the hidden states are supposed to follow a Markov property, and it outperforms the other models in terms of accuracy. The parameters of an HMM (λ) are denoted by A, B and π and are determined using the supplied or given input values. HMM is written as λ = (S, V, A, B, π) where S ={H_1,H_2, … ,H_N}is a set or collection of N possible hidden/invisible states, the set V={V_1,V_2,…,V_M} are M possible visible states, A is a square matrix of dimension N that is called the transition probability matrix (TPM), B is also an $N \times M$ rectangular matrix of observation probabilities which is termed as observed probability matrix (OPM) and finally the N dimensional vector π contains the probability of hidden states which is known as the initial probability vector (IPV). These parameters A, B and π of HMM satisfy the following conditions presented in equation (1.2.3).

$$\sum_{j=1}^{N} a_{ij} = \sum_{k=1}^{M} b_{jk} = \sum_{i=1}^{N} \pi_i = 1, \forall\ i, j = 1, 2, 3, ..., n \text{ and } k = 1, 2, 3, ..., m \quad (1.2.3)$$

For all, $a_{ij}, b_{jk}, \pi_i \geq 0$.

2. Description of the Hidden and Visible States

Since stock price variations assumes Markov dependency and time-homogeneity as per the existing literature, therefore, we design a two by two HMM with two hidden and two visible states. The two hidden states (decreasing and increasing INR/USD) that influence the happening of two visible states (fall and rise in the share value of Reliance Industries Ltd). The difference between the next day INR value and the previous day INR value with respect to USD yields two hidden states. These hidden states are symbolically written as follows.

H_1: When $(x_t - x_{t-1}) < 0$, then the process is in a decreasing state (D) and

H_2: When $(x_t - x_{t-1}) > 0$, then the process is in an increasing state (I).

Where, x_t *is* the current and x_{t-1} is the previous day INR value with respect USD. Since no two consecutive closing prices are from the real time historical in the given period are observed equal in, therefore there is no any observation in the difference $x_t - x_{t-1}$ is equal to zero.

Similarly, the two visible states are also obtained on the basis of the difference between the current and previous day's closing share price of Reliance industries limited and hence we write these two visible states as fallows.

V_1: When $(y_t - y_{t-1}) < 0$, then there is a fall in the share price of Reliance (F).

V_2: When $(y_t - y_{t-1}) > 0$, then there is a rise in the share price of Reliance (R).

Where, y_t *is* the current and y_{t-1} is the previous day's closing share price of Reliance industries limited. From the historical real time closing share price data, we have observed that there is no any difference $y_t - y_{t-1}$ equal to zero. In the empirical study, we will use the hidden states D and I and the visible states F and R presented in the table 1.

Table 1: Description of hidden and visible states.

Invisible (Hidden) states	**Visible(observation) states**
H_1=D: Decreasing Rupee value	V_1=F: Fall of Reliance share price
H_2=I: Increasing Rupee value	V_2=R: Rise of Reliance share price

2.1. Model parameters

In this section, we describe briefly the parameters of our two by two HMM such as the transition probability matrix (TPM), observed probability matrix (OPM) and an initial probability vector (IPV). the model parameters (TPM, OPM, IPV) are presented in the following subsections 2.1.1, 2.1.2 and 2.1.3 respectively.

2.1.1. Transition Probability Matrix (TPM)

Since the closing price difference of INR/USD has been separated into two states (D and I), the TPM will also include these two states, resulting in a 2 by 2 TPM. The TPM is a matrix that represents the probabilities of reaching to some hidden state from another hidden state in one step. The elements of TPM are all positive, and the rows must add up to one. The TPM for our two by two HMM is denoted by matrix A whereas the elements of matrix A are $a_{ij}, i, j = 1, 2$, the transition probabilities among hidden states from ith to jth and are defined as $a_{ij} = P(X_n = j / X_{n-1} = i) \geq 0$ such that $\sum_{j=1}^{2} a_{ij} = 1, i = 1, 2$. Therefore, the our TPM is of the form;

$$A = \begin{array}{c c} & \begin{array}{cc} D & I \end{array} \\ \begin{array}{c} D \\ I \end{array} & \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}_{2\times 2} \end{array}$$

2.1.2. Observed/Emission Probability Matrix (OPM)

Observed probability matrix (OPM) consists of transition probabilities from the hidden to visible states. Observed probability matrix our two by two HMM is denoted by matrix B and the elements of OPM are denoted as $b_{ij} \geq 0, \forall\ i, j = 1, 2$, (the probabilities from the ith hidden state to the jth visible state, i.e., the transition between the states) and are defined as $b_{ij} = P(y_m = j / x_n = i) \geq 0, i, j = 1, 2$ Such that $\sum_{j=1}^{2} b_{ij} = 1, i = 1, 2$. Therefore, the two-by-two OPM for our study will be of the form;

$$B = \begin{array}{c c} & \begin{array}{cc} F & R \end{array} \\ \begin{array}{c} D \\ I \end{array} & \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}_{2\times 2} \end{array}$$

2.1.3. Initial Probability Vector (IPV)

Initial probability vector (IPV) the vector of initial probabilities of the hidden/invisible states of an HMM. IPV is denoted by π (the vector contains the probabilities of decrease and increase in the INR value with respect to the USD) and the elements of IPV are written as $\pi_i = P(X_0 = i)\ \forall, i = 1, 2$ such that $\sum_{i=1}^{2} \pi_i = 1$.

Therefore, the IPV will be $\pi = [\pi_1, \pi_2]$.

2.2. Schematic diagram

HMM has been applied to predict economic systems and share market values in the subject of financial mathematics. In this paper we propose a statistical prediction model for the time series data of share values of Reliance industries limited using HMM. The Decrease (H_1) and Increase (H_2) of the INR value with respect to US dollar are two hidden (Invisible) states which influence the visible states like Fall (V_1) and Rise (V_2) of the stock value of Reliance as defined in the previous section. This technique will enable financial specialists in determining the optimal moment to purchase and sell stocks by automating the process of adjusting stock price indexes based on technical analysis. The arrow marks depict the probability of from and to the state of the system. The schematic diagram for the present study with two hidden and two visible states is given in Fig. 1.

Schematic diagram of two by two hmm

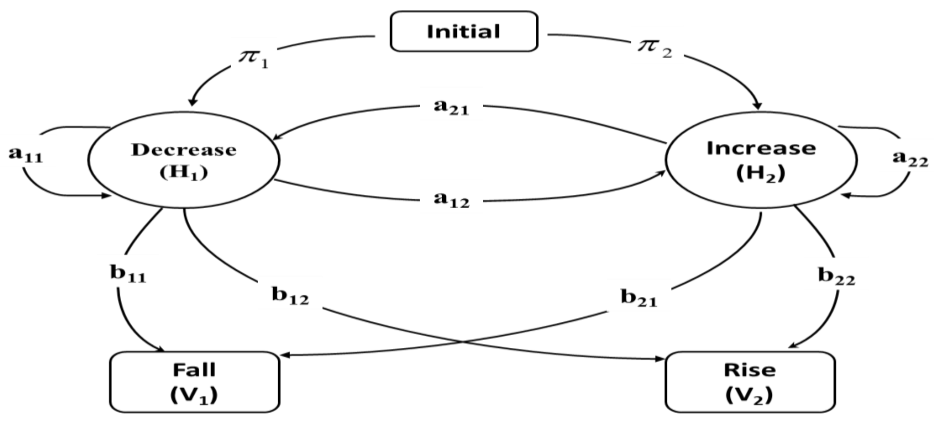

Fig. 1:

From the above state transition diagram, $a_{ij}, b_{ij}, \; i, j = 1, 2$ and $\pi_i, \; i = 1, 2$ are the elements of TPM, OPM, and IPV.

3. Probability Distributions for Visible States

In order to derive the probability distributions for all visible states $V_r, r = 1, 2$ of our two by two HMM, we need to define the random variable $X_r(\omega_t), r = 1, 2$ and $t = 1, 2, 3, \ldots$ which represents the occurrence of $V_r, r = 1, 2$ in a single day, two days sequence, three days sequence and so on. Here r denotes the rth visible state and t denotes the length of sequence. The probability distributions of $V_r, r = 1, 2$ in a single day, two days sequence and three days sequences along with their respective descriptive statistical measures are derived in the following sections 3.1, 3.2 and 3.3 respectively.

3.1. Probability distributions of $V_r, r = 1, 2$ in a single day

In this section, we have obtained the formula for computing the probability of visible states ($V_r, r = 1, 2$). These probabilities are obtained by the effect and influence of invisible/hidden states ($H_q, q = 1, 2$). Hence $V_r, r = 1, 2$ can happen either by the happening of H_1 or H_2. Therefore, the formulae for obtaining the probability of visible states $V_r, r = 1, 2$ is obtained in equation (3.1.1).

$$P(V_r) = \sum_{i=1}^{2} P(V_r H_q) = \sum_{i=1}^{2} \pi_i b_{ir} \geq 0, \quad for \;\; r = 1, 2 \tag{3.1.1}$$

Generally, for an M by N HMM, the probability of each and every M visible states are obtained using equation (3.1.2).

$$P(V_r) = \sum_{i=1}^{N} \pi_i b_{ir} \geq 0, \; \forall \; r = 1, 2, 3, \ldots, M \text{ , such that} \sum_{r=1}^{M} P(V_r) = 1 \tag{3.1.2}$$

I. Probability distributions

Let $X_r(\omega_1), r = 1, 2$ be a random variable that represents the frequency of $V_r, r = 1, 2$ in one day sequence. Then the random variable $X_r(\omega_1) = 0, 1$ and the probability mass function (PMF) of $X_r(\omega_1)$ is derived in the equation (3.1.3).

$$P(X_r(\omega_1) = x_r) = \begin{cases} \sum_{i=1}^{2} \pi_i b_{ir}, & x_r = 1 \\ (1 - \sum_{i=1}^{2} \pi_i b_{ir}), & x_r = 0 \\ 0, & otherwise \end{cases} \tag{3.1.3}$$

Where for r=1, we will get the probability distribution of state F, and when r=2, we will get the probability distribution of state R.

II. Statistical measures

The formulae for the k^{th} movement about origin is derived in equation (3.1.4).

$$\mu_k' = \sum_{i=1}^{2} \pi_i b_{ir} \quad r = 1,2 \tag{3.1.4}$$

Since the expression is independent of k, then all the moments about origin are equal and therefore, the formulae for the mean and variance of the above derived probability distributions are presented in equations (3.1.5) and (3.1.6) respectively.

$$\mu_1' = \sum_{i=1}^{2} \pi_i b_{ir} \quad r = 1,2 \tag{3.1.5}$$

$$\mu_2 = \sum_{i=1}^{2} \pi_i b_{ir} (1 - \sum_{i=1}^{2} \pi_i b_{ir}),\ r = 1,2 \tag{3.1.6}$$

Now in order to derive the expression for the coefficient of skewness and the coefficient of kurtosis, the third and the fourth central moment are required to obtain. The formulae for obtaining the third and fourth central moment are derived as in the equation (3.1.7) and (3.1.8) respectively.

$$\mu_3 = \sum_{i=1}^{2} \pi_i b_{ir} (1 - 3\sum_{i=1}^{2} \pi_i b_{ir} + 2(\sum_{i=1}^{2} \pi_i b_{ir})^2) \tag{3.1.7}$$

$$\mu_4 = \sum_{i=1}^{2} \pi_i b_{ir} (1 - 4\sum_{i=1}^{2} \pi_i b_{ir} + 6(\sum_{i=1}^{2} \pi_i b_{ir})^2 - 3(\sum_{i=1}^{2} \pi_i b_{ir})^3) \tag{3.1.8}$$

Hence the coefficient of skewness and coefficient of kurtosis are presented in equation (3.1.9) and (3.1.10) respectively.

$$\beta_1 = (1 - 2\sum_{i=1}^{2} \pi_i b_{ir})(\sum_{i=1}^{2} \pi_i b_{ir} (1 - \sum_{i=1}^{2} \pi_i b_{ir}))^{-1} \tag{3.1.9}$$

$$\beta_2 = (1 - 4\sum_{i=1}^{2} \pi_i b_{ir} + 6(\sum_{i=1}^{2} \pi_i b_{ir})^2 - 3(\sum_{i=1}^{2} \pi_i b_{ir})^3)(1 - \sum_{i=1}^{2} \pi_i b_{ir})^{-1} \tag{3.1.10}$$

Furthermore, values of γ_1 and γ_2 for the proposed probability distributions are given in equations (3.1.11) and (3.1.12) respectively.

$$\gamma_1 = (1-2\sum_{i=1}^{2}\pi_i b_{ir})^{1/2}(\sum_{i=1}^{2}\pi_i b_{ir}(1-\sum_{i=1}^{2}\pi_i b_{ir}))^{-1/2} \quad (3.1.11)$$

$$\gamma_2 = (1-4\sum_{i=1}^{2}\pi_i b_{ir} + 6(\sum_{i=1}^{2}\pi_i b_{ir})^2 - 3(\sum_{i=1}^{2}\pi_i b_{ir})^3)(1-\sum_{i=1}^{2}\pi_i b_{ir})^{-1} - 3 \quad (3.1.12)$$

III. Generating function

The expressions of various generating functions such as moment generating function (MGF), probability generating function (PGF), and characteristic functions (C.F.) of the proposed probability distributions are derived in this section. The formulae for MGF, PGF and CF are given in the expressions (3.1.13), (3.1.14), and (3.1.15) respectively:

a). Moment generating function

$$M_x(t) = 1-(1-e^t)\sum_{i=1}^{2}\pi_i b_{ir}, r = 1,2 \quad (3.1.13)$$

b). Probability generating function

$$P_x(t) = \sum_{i=1}^{2}\pi_i b_{ir}, r = 1,2 \quad (3.1.14)$$

c). Characteristics function

$$\phi_x(t) = 1-(1-e^{it})\sum_{i=1}^{2}\pi_i b_{ir}, r = 1,2 \text{ and } i = \sqrt{-1} \quad (3.1.15)$$

3.2. Probability distributions of $V_r, r = 1,2$ in the run of two visible states

The probability of sequence of two visible states is obtained on the effect of all possible invisible/hidden states' sequences. The probability of the sequence of two visible states can occur by the happening of four different possible combinations of hidden states. Mathematically, we write the probability of (V_l, V_m) for all (l, m = 1,2) can happen jointly with the happening of (H_1, H_1), (H_1, H_2), (H_2, H_1) and (H_2, H_2). Therefore, the formulae for obtaining the probability of the sequence of two visible states is presented in equation (3.2.1).

$$P(V_l, V_m) = \left\{\sum_{i=1}^{2}\pi_i b_{il}\left[\sum_{j=1}^{2}a_{ij}b_{jm}\right]\right\}, \forall\ l, m = 1,2 \quad (3.2.1)$$

For a specific combination say (V1, V2) using the above expression, we can obtain the probability of (V1 V2) as in equation (3.2.2).

$$P(V_1,V_2)=\left\{\sum_{i=1}^{2}\pi_i b_{i1}\left[\sum_{j=1}^{2}a_{ij}b_{j2}\right]\right\} \tag{3.2.2}$$

In the same way, the expressions for obtaining the probability of (V_1, V_1), (V_2, V_1) and (V_2, V_2) can be obtained by substituting the values of l and m.

I. Probability distributions

Suppose $X_r(\omega_2), r=1,2$ be a random variable which denotes the frequency of visible state $V_r, r=1,2$ happens in the run of length two days, then $X_r(\omega_2)$ takes values 0, 1and 2. Mathematically, we say that a random variable $X_r(\omega_2)$ is said to fallow the derived probability distribution if it takes the integral 0,1,2 and its PMF is given in equation (3.2.3).

$$P(X_r(\omega_2)=x_r)=\begin{cases}\sum_{j=1}^{M}\left\{C_x^2 b_{jr}{}^x(1-b_{jr})^{2-x}(\delta)\right\}, & x=0,1,2\\ 0 & , otherwise\end{cases} \tag{3.2.3}$$

Where, $\delta=\sum_{i=1}^{N}\pi_i a_{ij}$ and $a_{ij}, b_{ij}, \pi_i\ \forall i, j, r=1,2$ have the usual meaning. This model deals with two hidden two visible states that N=M=2. For every state, there is a corresponding probability distribution. Putting the value of r=1, we will get the probability distribution of the visible state V_1=F and when r=2 in equation 1, we will get the probability distribution of V_2=R.

II. Statistical measures

The measures based on moments for the present study are obtained. The expressions for k[th] raw moment is given in equation (3.2.4), formula for mean is derived in equation (3.2.5) and the expression for variance for the derived probability distributions is given in equation (3.2.6).

$$\mu_k{}' = 2\sum_{j=1}^{2}\left[b_{jk}\left\{1+\left(2^{r-1}-1\right)b_{jk}\right\}(\delta)\right], r=1,2 \textit{ and } k=1,2,3,... \tag{3.2.4}$$

Therefore, the formulae for mean and variance of the proposed probability distributions are

$$\mu_1{}'=2\sum_{j=1}^{2}b_{jk}[\delta]=2\theta \tag{3.2.5}$$

$$\mu_2=2\theta(1-2\theta)+2\phi \tag{3.2.6}$$

where, $\theta=\sum_{j=1}^{2}b_{jk}[\delta]$, $\phi=\sum_{j=1}^{2}b_{jr}{}^2[\delta]$, $\delta=\sum_{i=1}^{2}\pi_i a_{ij}$ and r=1, 2.

In order to find the shape and peekedness of the probability distributions, we need to obtained the third and fourth central moments. The formulae for obtaining the third and the fourth central moments are derived in equations (3.2.7) and (3.2.8) respectively.

$$\mu_3 = 2\theta(4\theta)(2\theta-1)+6(2\theta-1)\phi \quad (3.2.7)$$

$$\mu_4 = 2\theta\left(4\theta-1\right)\left(-3\left(2\theta\right)^2+6\theta-1\right)+\left(12\left(2\theta\right)^2-24\left(2\theta\right)+7\right)\phi \quad (3.2.8)$$

Similarly, the Pearson's coefficients of skewness and kurtosis are respectively derived in equations (3.2.9) and (3.2.10).

$$\beta_1 = \left(2\theta\left(4\theta\right)\left(2\theta-1\right)+6\left(2\theta-1\right)\phi\right)^2\left(2\theta\left(1-2\theta\right)+2\phi\right)^{-3} \quad (3.2.9)$$

$$\beta_2 = \left[2\theta\left(4\theta-1\right)\left(-3\left(2\theta\right)^2+6\theta-1\right)+\left(12\left(2\theta\right)^2-24\left(2\theta\right)+7\right)\phi\right]\left[2\theta\left(1-2\theta\right)+2\phi\right]^{-2} \quad (3.2.10)$$

Also, the values of γ_1 and γ_2 for the proposed probability distributions are given in equations (3.2.11) and (3.2.12) respectively.

$$\gamma_1 = \left(2\theta\left(4\theta\right)\left(2\theta-1\right)+6\left(2\theta-1\right)\phi\right)*\left(2\theta\left(1-2\theta\right)+2\phi\right)^{\frac{-3}{2}} \quad (3.2.11)$$

$$\gamma_2 = \left[2\theta\left(4\theta-1\right)\left(-3\left(2\theta\right)^2+6\theta-1\right)+\left(12\left(2\theta\right)^2-24\left(2\theta\right)+7\right)\phi\right]\left[2\theta\left(1-2\theta\right)+2\phi\right]^{-2}-3 \quad (3.2.12)$$

where, $\theta = \sum_{j=1}^{2} b_{jr}\left[\delta\right]$, $\phi = \sum_{j=1}^{2} b_{jr}{}^2\left[\delta\right]$, $\delta = \sum_{i=1}^{2}\pi_i a_{ij}$ and r=1, 2.

III. Generating functions

The generating functions of the proposed probability distributions such as the MGF, PGF and characteristic function are obtained in this section. These functions are obtained in order to find the moments and other descriptive statistical parameters. The formulae for the MGF, PGF and CF are given in the expressions (3.2.13), (3.2.14), and (3.2.15) respectively:

a). Moment generating function

$$M_x(t) = \sum_{j=1}^{2}\left[\left\{\left(1-b_{jr}\right)\left(1-b_{jr}+2e^t b_{jr}\right)+e^{2t}b_{jr}{}^2\right\}\delta\right], \forall r = 1,2 \quad (3.2.13)$$

b). Probability generating function

$$P_x(s) = \sum_{j=1}^{2}\left[\left\{(1-b_{jr})^2+2sb_{jr}(1-b_{jr})+s^2 b_{jr}{}^2\right\}\delta\right] \quad (3.2.14)$$

c). Characteristic function

$$\phi_x(t) = \sum_{j=1}^{2}\left[\left\{\left(1-b_{jr}\right)\left(1-b_{jr}+2e^{it}b_{jr}\right)+e^{2it}b_{jr}{}^2\right\}\delta\right] \quad (3.2.15)$$

Such that $i=\sqrt{-1}$ and $r=1,2$

3.3. Probability distributions $V_r, r=1,2,3$ in three days sequence

In the present sections, we have developed the formulae for the probability of getting three days visible states which are influenced by the effect of eight different combinations of hidden states. For instance, by applying these formulae, we can find out three days fall, three days rise, probability of first two days fall and the third day rise and so on in the stock market. The formulae for three days visible states are presented in equation (3.3.1).

$$P(V_l,V_m,V_n)=\sum_{k=1}^{2}\left[\left(\sum_{j=1}^{2}a_{kj}b_{jn}\right)b_{km}\left(\sum_{i=1}^{2}\pi_i a_{ik}b_{il}\right)\right],\ \forall\ l,m,n=1,2,3 \qquad (3.3.1)$$

For a particular sequence (V_1, V_2, V_3), we put l=1, m=2 and n=3 in equation (3.3.1) and we will get the $P(V_1,V_2,V_3)$ that is the probability of first day V_1, second day V_2 and third day V_3 in equation (3.3.2).

$$P(V_1,V_2,V_3)=\sum_{k=1}^{2}\left[\left(\sum_{j=1}^{2}a_{kj}b_{j3}\right)b_{k2}\left(\sum_{i=1}^{2}\pi_i a_{ik}b_{i1}\right)\right] \qquad (3.3.2)$$

I. Probability distributions

Now in order to derive the probability distribution for visible state $V_r, r=1,2,3$ in the run of three consecutive trading days, suppose $X_r(\omega_3), r=1,2,3$ be the random variable which denotes the occurrence of the r^th visible state $V_r, r=1,2$ in a sequence of length 3. Then, $X_r(\omega_3)=0,1,2,3$. Therefore, the random variable $X_r(\omega_3), r=1,2$ is said to follow the proposed probability distribution if it takes the integral values of 0,1, 2, 3 and its probability mass function is presented in equation (3.3.3).

$$P(X_r(\omega_3)=x_r)=\begin{cases}\sum\limits_{K=1}^{2}\left[\left(1-b_{kr}\right)\sum\limits_{j=1}^{2}a_{kj}\left(1-b_{jr}\right)\sum\limits_{i=1}^{2}\pi_i a_{ik}\left(1-b_{ir}\right)\right], & x_r=0\\ \sum\limits_{k=1}^{2}\left[\sum\limits_{j=1}^{2}a_{kj}\left(1-b_{jr}\right)\left\{\sum\limits_{i=1}^{2}\pi_i a_{ik}\left(b_{ir}+b_{kr}\left(1-2b_{ir}\right)\right)\right\}+\left(1-b_{kr}\right)\sum\limits_{j=1}^{2}a_{kj}b_{jr}\cdot\sum\limits_{i=1}^{2}\pi_i a_{ik}\left(1-b_{ir}\right)\right], & x_r=1\\ \sum\limits_{k=1}^{2}\left[b_{kr}\sum\limits_{j=1}^{2}a_{kj}\left(1-2b_{jr}\right)\cdot\sum\limits_{i=1}^{2}\pi_i a_{ik}b_{ir}+\sum\limits_{j=1}^{2}a_{kj}b_{jr}\cdot\sum\limits_{i=1}^{2}\pi_i a_{ik}\left(b_{ir}+b_{kr}(1-b_{ir})\right)\right], & x_r=2\\ \sum\limits_{k=1}^{2}\left[\left(\sum\limits_{j=1}^{2}a_{kj}b_{jr}\right)b_{kr}\left(\sum\limits_{i=1}^{2}\pi_i a_{ik}b_{ir}\right)\right], & x_r=3\\ 0, & \text{otherwise}\end{cases} \qquad (3.3.3)$$

II. Mean and Variance

The expression for the k^th moment about origin is derived as in the following equation (3.3.4).

$$\mu_k^{'} = \beta + 2^k \gamma + 3^k \delta, k = 1, 2, 3, \ldots \tag{3.3.4}$$

Where the value of α, β, γ *and* δ are respectively given in equations (3.3.5), (3.3.6), (3.3.7) and (3.3.8).

$$\alpha = \sum_{K=1}^{2}\left[\left(1-b_{kr}\right)\sum_{j=1}^{2} a_{kj}\left(1-b_{jr}\right)\sum_{i=1}^{2}\pi_i a_{ik}\left(1-b_{ir}\right)\right] \tag{3.3.5}$$

$$\beta = \sum_{k=1}^{2}\left[\sum_{j=1}^{2} a_{kj}\left(1-b_{jr}\right)\left\{\sum_{i=1}^{2}\pi_i a_{ik}\left(b_{ir}+b_{kr}\left(1-2b_{ir}\right)\right)\right\}+\left(1-b_{kr}\right)\sum_{j=1}^{2} a_{kj}b_{jr}.\sum_{i=1}^{2}\pi_i a_{ik}\left(1-b_{ir}\right)\right] \tag{3.3.6}$$

$$\gamma = \sum_{k=1}^{2}\left[b_{kr}\sum_{j=1}^{2} a_{kj}\left(1-2b_{jr}\right).\sum_{i=1}^{2}\pi_i a_{ik}b_{ir}+\sum_{j=1}^{2} a_{kj}b_{jr}.\sum_{i=1}^{2}\pi_i a_{ik}\left(b_{ir}+b_{kr}(1-b_{ir})\right)\right] \tag{3.3.7}$$

$$\delta = \sum_{k=1}^{2}\left[\left(\sum_{j=1}^{2} a_{kj}b_{jr}\right)b_{kr}\left(\sum_{i=1}^{2}\pi_i a_{ik}b_{ir}\right)\right] \tag{3.3.8}$$

Therefore, the formulae for mean and variance of the proposed probability distribution are presented in equation (3.3.9) and (3.3.10) respectively.

$$\mu_1^{'} = \beta + 2\gamma + 3\delta \tag{3.3.9}$$

$$\mu_2 = \beta\left(1-\beta-4\gamma\right)+4\gamma\left(1-\gamma-3\delta\right)+3\delta\left(3-3\delta-2\beta\right) \tag{3.310}$$

III. Other Statistical parameters

The value of μ_3 and μ_4 of the probability distributions are also derived in equation (3.3.11) and (3.3.12) respectively.

$$\mu_3 = \beta\left(1-\beta-10\gamma\right)+24\gamma\left(4-4\gamma-33\delta\right)+3\delta\left(9-8\delta-21\beta\right) \tag{3.3.11}$$

$$\mu_4 = \beta+16\gamma+81\delta-\left(\beta+2\gamma+3\delta\right)\left[\beta(4-5\beta-32\gamma)+\gamma(32-44\gamma-168\delta)+\delta(108-153\delta-66\beta)\right] \tag{3.3.12}$$

Therefore, the shaping measures that is the value of β_1 and β_2 of the probability distributions are derived in equation (3.3.13) and (3.3.14) respectively.

$$\beta_1 = \frac{\left[\beta\left(1-\beta-10\gamma\right)+24\gamma\left(4-4\gamma-33\delta\right)+3\delta\left(9-8\delta-21\beta\right)\right]^2}{\left[\beta\left(1-\beta-4\gamma\right)+4\gamma\left(1-\gamma-3\delta\right)+3\delta\left(3-3\delta-2\beta\right)\right]^3} \tag{3.3.13}$$

$$\beta_2 = \frac{\beta+16\gamma+81\delta-(\beta+2\gamma+3\delta)\left[\beta(4-5\beta-32\gamma)+\gamma(32-44\gamma-168\delta)+\delta(108-153\delta-66\beta)\right]}{\left[\beta(1-\beta-4\gamma)+4\gamma(1-\gamma-3\delta)+3\delta(3-3\delta-2\beta)\right]^2} \quad (3.3.14)$$

Furthermore, we have derived the formulae for obtaining the value of γ_1 and γ_2 for the proposed probability distributions. The value of γ_1 and γ_2 are presented in equations (3.3.15) and (3.3.16) respectively.

$$\gamma_1 = \left[\beta(1-\beta-10\gamma)+24\gamma(4-4\gamma-33\delta)+3\delta(9-8\delta-21\beta)\right]\left[\beta(1-\beta-4\gamma)+4\gamma(1-\gamma-3\delta)+3\delta(3-3\delta-2\beta)\right]^{-3/2} \quad (3.3.15)$$

$$\gamma_2 = \frac{\beta+16\gamma+81\delta-(\beta+2\gamma+3\delta)\left[\beta(4-5\beta-32\gamma)+\gamma(32-44\gamma-168\delta)+\delta(108-153\delta-66\beta)\right]}{\left[\beta(1-\beta-4\gamma)+4\gamma(1-\gamma-3\delta)+3\delta(3-3\delta-2\beta)\right]^2}-3 \quad (3.3.16)$$

IV. Generating functions

Finally, for the probability distributions of the random variable $X_r(\omega_3), r=1,2,3$ the formulae for various generating functions such as moment generating function (MGF), probability generating function (PGF), and characteristic functions (C.F.) are derived in equations (3.3.17), (3.3.18) and (3.3.19) respectively:

a). Moment generating function (MGF)

$$M_x(t) = \alpha + e^t\beta + e^{2t}\gamma + e^{3t}\delta \quad (3.3.17)$$

b). Probability generating function (PGF)

$$P_x(t) = \alpha + t\beta + 2t\gamma + 3t\delta \quad (3.3.18)$$

c). Characteristic function

$$\phi_x(t) = \alpha + e^{it}\beta + e^{2it}\gamma + e^{3it}\delta \quad (3.3.19)$$

The key idea behind our new approach to HMM is to estimate the parameters of HMM (A, B and π) using the training dataset. we know the previous day closing share price for a given stock at the market, and our goal is to predict the next day's closing share price using this information. All the derived and generated expressions are then used in the real-time historical closing share price data set of Reliance Industries limited to analyze and predict its next day's closing share price and to forecast its long-term behavior.

4. Statistical Data Analysis

In this paper, we have retrieved two data sets from the source. The primary data set deals with invisible states and the second data set deals with visible states.

The daily data regarding the changing values in INR with respect to USD is considered hidden factor and the daily changes in the closing share prices of Reliance industries limited is visible data set. The basic aim in this paper is to forecast the share prices of Reliance using HMM. The daily data of INR with respect USD from Jan. 02, 2018 to Dec. 24, 2019 consisting of 486 observations is retrieved from the website of Reserve Bank of India (https://in.investing.com/currencies/usd-inr-historical-data). Similarly, the daily data of the closing share prices of Reliance are obtained from www.yahoofinance.com for the same time period. The necessary variable needed for computation like date and closing prices of INR and closing share prices of Reliance industries limited are extracted from the corresponding sources.

4.1. Stock market trend of Reliance and INR value

Fig. 2 depicts a side-by-side comparison of the movement of the Indian rupee value with respect to USD and the Reliance stock market trend from January 02, 2018, to December 24, 2019.

Stock market movement of reliance vs change in inr/usd

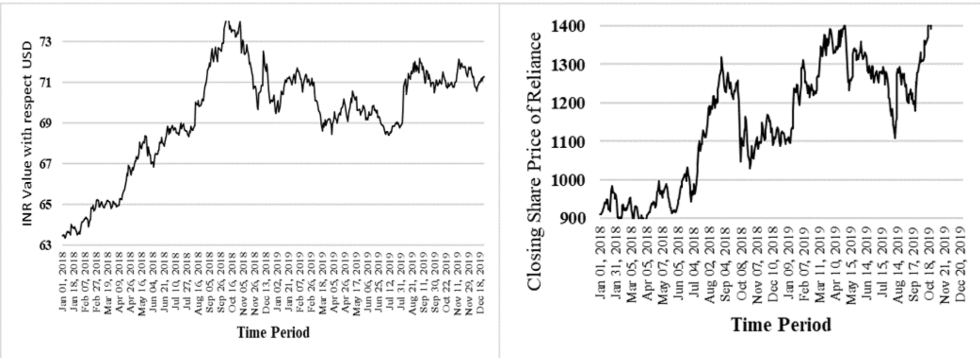

Fig. 1

Since the correlation between the visible state (Reliance share prices) and the invisible state (change in INR) for the time period from February 2, 2018 to December 24, 2019 is 0.63, we can conclude that the hidden and visible states are positively correlated and that the two stocks move in the same direction. This diagram also shows that as the INR fluctuates, the share price of Reliance fluctuates as well.

4.2. Change in INR and share prices of Reliance

Fig. 3 depicts the daily movement in INR/USD against the volatility in Reliance's share prices. It is clear that the fluctuation in Indian rupee value has a substantial impact on the stock price of Reliance Industries. The increase and decrease in the INR correspond to the rise and fall in Reliance's stock market movement.

Change in inr and closing price of reliance

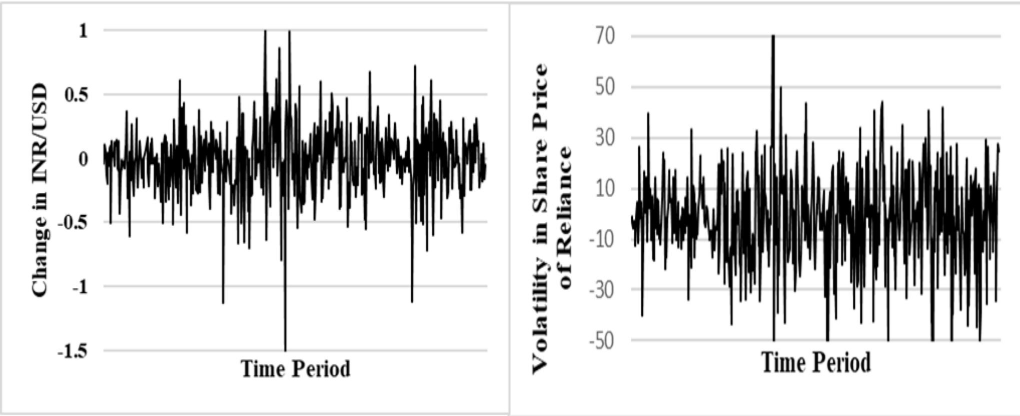

Fig. 3

4.3. Frequency diagrams of hidden and visible states

As described in section 2, data on the change in INR and Reliance share prices are transformed into labels such as decrease (D) and increase (I) for hidden states and fall (F) and rise (R) for visible states. The frequency of states D and I, as well as visible states F and R, are shown in table 2. In the period from January 02, 2018 to December 24, 2019, INR decreased 236 times and increased 249 times, as shown in table 2. In the same time span, the closing share prices of Reliance have fallen 238 times and risen 247 times.

Table 2: Frequency table for invisible and visible states.

	Invisible states		**Visible states**	
States	D	I	F	R
Frequency	236	249	238	247

The above table shows that as the value of the INR decreases, the share price of Reliance falls in almost the same proportion, and as the INR increases with respect to USD, the share price rises in the same ratio.

4.4. Frequency diagrams for the transition states

The frequency diagram represents the relation of transition states in terms of transition frequencies of both visible and invisible/ influencing factors. Table 3 and Fig. 4 shows a comparison between how INR value and the share prices of

Reliance are changing in terms of transition frequencies of hidden and visible states (D, I and F, R).

Table 3: Transition frequencies of hidden and visible states.

		Hidden States		Visible States	
Hidden States		D	I	F	R
	D	114	122	113	129
	I	122	126	126	117

These frequencies are expressed in diagrammatic representation in Fig. 4 that represents a relation between how INR/USD and the share prices of Reliance are changing in terms of transition of hidden and visible states (D, I and F, R).

TRANSITION FREQUENCIES OF INVISIBLE AND VISIBLE STATES

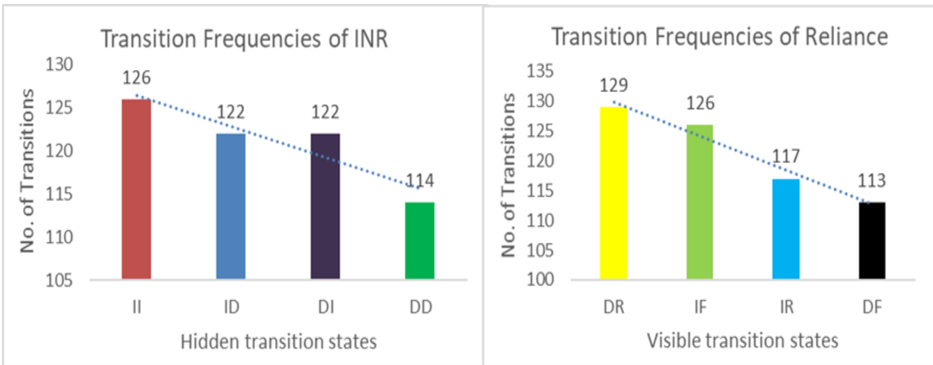

Fig. 4

The above diagram clearly shows that the decrease in the INR is directly proportional to the fall in the share values of Reliance Industries limited whereas when there is an increase in INR value, the share prices of Reliance rise.

5. Results and Discussion

In this section, we have estimated the parameters of HMM, the TPM, OPM and the IPV for the real time historical data of INR and Reliance Industries limited. The Transition probability matrix, observed probability matrix and an initial probability vector presented in section 4.1, 4.2 and 4.3 respectively.

5.1. Transition probability matrix

Microsoft Excel is used to calculate the TPM for the INR value (from the previous day's value to the current day's rupee value). The elements of the TPM represent the chance that the system's state will change to state j from state i in one step. The TPM our two by two HMM is estimated from the real time historical closing price of INR/USD is presented in matrix A given below.

$$A = \begin{array}{c} \\ D \\ I \end{array}\begin{array}{c} \begin{array}{cc} D & I \end{array} \\ \begin{bmatrix} 0.483051 & 0.516949 \\ 0.491935 & 0.508065 \end{bmatrix} \end{array}$$

This matrix A reveals that the probability of the system will go to decreasing state on the next day given that a decreasing state was observed on the previous day is 48.3%, and there are 51.7% that the INR will increase with respect to USD in the next day given that the INR is in decreasing today. Similarly, the second row of the matrix demonstrates that if the INR is in an increasing state, there is a 49% likelihood that it will move to decreasing state on the next day, whereas the INR will remain in the same state with a 51% chance. It can be observed from the above matrix A that when there is the increase in the INR in the previous day, there will be increase in the INR in the next day with maximum probability.

5.2. Observed probability matrix

The OPM of the HMM is also a matrix of probabilities of transitioning from hidden state to visible states. In other words, the likelihood of the system's state changing to the j^{th} visible state from the i^{th} hidden state is b_{ij}. The observed probability matrix for our two by two HMM is denoted by B and estimated from the real time historical hidden and visible data sets given below.

$$B = \begin{array}{c} \\ D \\ I \end{array}\begin{array}{c} \begin{array}{cc} F & R \end{array} \\ \begin{bmatrix} 0.466942 & 0.533058 \\ 0.518519 & 0.481481 \end{bmatrix} \end{array}$$

The first row in the above matrix B represents the probability that as the INR decreases today, the share prices of Reliance will fall with 46.69% chances and rise with 53.30 % chances on the next day. Similarly, the second row in matrix B depicts that as the INR increases today, the share price of Reliance will fall with 51.9% likelihood and rise with probability 0.48 on the next day. We have observed from matrix B that when the INR decreases with respect to USD, the closing share price of Reliance industries will rise with maximum probability.

5.3. Initial probability vector

The IPV is the vector of initial probabilities of the hidden states. The two invisible states in this study are decrease and increase of INR with respect to USD. From the table 2, it is observed that 236 times there is a decrease in INR during the given time period and 249 times that the increase in INR with respect USD out of 485. Hence the corresponding probabilities of the states D and I are

48.7% and 51.3% respectively. Hence the IPV of our two by two HMM is presented in vector π given below.

$$\begin{array}{cc} D & I \end{array}$$
$$\pi = \begin{bmatrix} 0.486598 & 0.513402 \end{bmatrix}$$

These parameters of the model such as the TPM, OPM, and IPV are obtained from the real time data sets can best be explained through the diagram called the schematic diagram or the state transition diagram. The schematic diagram of our tow by two HMM with two hidden (H_1=D, H_2=I) and two visible (V_1=F, V_2=R) states is presented in Fig. 5.

SCHEMATIC DIAGRAM OF TWO BY TWO HMM

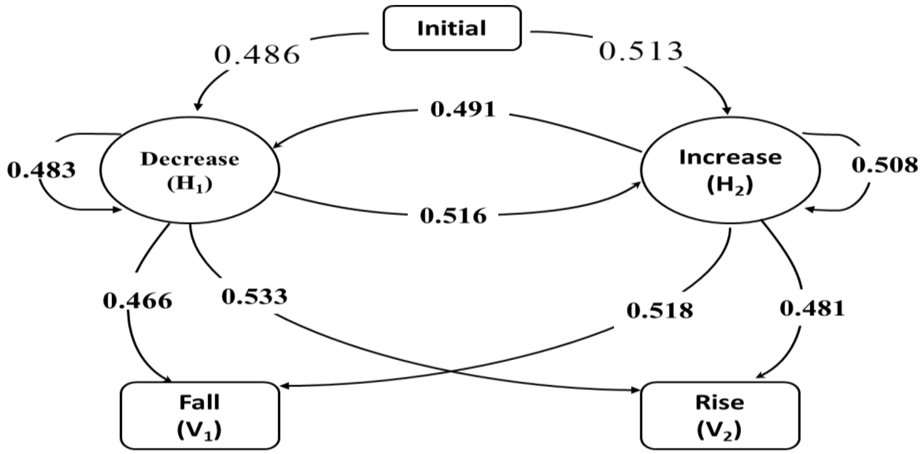

Fig. 5

Microsoft Excel and R Software were used for the mathematical and statistical computations and graphical displays for this study.

6. Probabilities distributions for visible states

In this section, we have calculated the probability of visible states V_1 and V_2 for different time lengths (that is for one day, for two days and for three days sequence) for our two by two HMM by applying the formulae obtained in section 3.1, 3.2 and 3.3. The probability distributions for the visible state particularly the falling state in this study along with their descriptive statistical measures are presented in next subsections 5.1, 5.2 and 5.3 respectively.

6.1. Probabilities distribution for falling state in a single day sequence

On applying equations (3.1.1), the probabilities of the visible states such as V_1 = Falling of share price of Reliance and V_2 = Rising in the share price of Reliance industries respectively are obtained as under.

Table 4: Probabilities of visible states.

States	Probability
$P(V_1) = P(\text{Falling})$	0.493422
$P(V_2) = P(\text{Rising})$	0.506578

The highest probability is associated with rising in the share prices of Reliance. Now, let $X_F(\omega_1)$ be a random variable representing the falling states in one day sequence, then $X_F(\omega_1)$ takes values 0 or 1. Hence the probability distribution of $X_F(\omega_1)$ (falling the share prices of Reliance industries limited) is as fallows.

Table 5: Probability distribution of falling state in single day sequence.

$X(\omega)$	0	1	Total
$P\{X(\omega)\}$	0.506578	0.493422	1

I. Descriptive statistical measures

Various statistical parameters such as the mean, variance, CV, skewness and kurtosis of the above probability distribution are obtained from the real time historical data sets using the above derived formulae. The statistical parameters of the probability distribution are presented in following table 6.

Table 6: Descriptive statistical measures.

Raw moments		Central moments		Coefficient of Skewness		Coefficient of Kurtosis	
μ_1'	0.493422						
μ_2'	0.493422	μ_2	0.249957				
μ_3'	0.493422	μ_3	0.249957	β_1	4.000692	β_2	-1.9214
μ_4'	0.493422	μ_4	-0.12005	γ_1	2.000173	γ_2	-4.9214
CV	101.3244						

6.2. Probabilities distribution for falling state in two days sequence

The formulae for obtaining the probability of sequence of visible states $(V_\alpha, V_\beta) \;\; \alpha, \beta = 1, 2$ of length two are derived in section 3.2 in equation (3.2.1). Therefore, by applying equation (3.2.1), the probability of visible states of length two with the real time historical data are presented in table 7.

Table 7: Probability of visible states in two days sequence.

Sequences	**Probability**	**Sequences**	**Probability**
F, F	0.244078	R, F	0.249291
F, R	0.249291	R, R	0.257339

Here P (F, F) = 0.244078 is the probability that there will be consecutive two days fall and P (R, R) = 0.257339 is the probability of consecutive two days rise in the share prices of Reliance industries limited. In this section, we have obtained the probability distribution of the falling state in two days sequence. The probability distribution of falling state with the real time historical data set using equation (3.2.3) in section 3 is given in the following table.

Table 8: Probability distribution of falling state in two days sequence.

$X(\omega)$	0	1	2	Total
$P\{X(\omega)\}$	0.257339	0.498583	0.244078	1

The mean, variance and other statistical measures of the above probability distribution are presented in the next section.

I. Statistical measures

The mean and variance and other statistical measures of the above derived probability distribution of the random variable $X_F(\omega_2)$ denoting the falling state in the sequence of run two using HMM are estimated from the real time historical data are presented in the following table 9. Therefore, by using equations derived in section 3.3, the mean, variance, coefficient of variation, skewness and kurtosis are obtained in table 9.

Table 9: Descriptive Statistical measures.

Raw moments		Central moments		Coefficient of Skewness		Coefficient of Kurtosis	
μ_1'	0.986739						
μ_2'	1.474895	μ_2	0.501241				
μ_3'	2.451207	μ_3	0.032505	β_1	0.00839	β_2	1.842925
μ_4'	4.403831	μ_4	0.463021	γ_1	0.091598	γ_2	-1.15708
CV	71.74986						

From the results it is observed that in successive occurrence of 2-day sequence, the fall of share value is observed around one day with variance 0.501241.

6.3. Probabilities distribution for falling state in three days sequence

The formulae for obtaining the probability of sequence of visible states $(V_l, V_m, V_n)\ \ l, m, n = 1, 2$ of length two are derived in section 3.3 in equation (3.3.1). Therefore, by applying equation (3.3.1), the probability of visible states of length two with the real time historical data are presented in table 10.

Table 10: Probability of the sequence of three visible states

Sequences	**Probability**	**Sequences**	**Probability**
F, F, F	0.25078	R, F, F	0.052436
F, F, R	0.110231	R, R, F	0.052039
F, R, F	0.102514	R, R, R	0.432

Here P (F, F, F) = 0.25078 is the probability that there will be consecutive three days fall and P (R, R, R) = 0.43200 is the probability of consecutive three days rise in the share prices of Reliance industries limited. In this section, we have obtained the probability distribution of the falling state in three days sequence. The probability distribution of falling state with the real time historical data set using equation (3.3.3) in section 3.3 is given in the following table 11.

Table 11: Probability distribution of falling state in three days sequence.

$X(\omega)$	0	1	2	3	Total
$P\{X(\omega)\}$	0.432	0.154553	0.162667	0.25078	1

The mean, variance and other statistical measures of the above probability distribution are presented in the next section.

I. Statistical measures

Various descriptive statistical measures of the above derived probability distribution of the random variable $X_F(\omega_2)$ denoting the falling state in the sequence of run three using HMM are estimated from the real time historical data are presented in the following table 12. Therefore, by using equations derived in section 3.3, the mean, variance, coefficient of variation, skewness and kurtosis are obtained in table 12.

Table 12: Descriptive Statistical measures.

Raw moments		Central moments		Coefficient of Skewness		Coefficient of Kurtosis	
μ_1'	1.232227						
μ_2'	3.062241	μ_2	1.543858				
μ_3'	8.226949	μ_3	0.648807	β_1	0.114396	β_2	1.469271
μ_4'	23.07041	μ_4	3.502001	γ_1	0.338224	γ_2	-1.53073
CV	100.8354						

From the results it is observed that in successive occurrence of 3-day sequence, the fall of share value is observed more than one day with variance 1.543.

7. Stationary Probability Distributions

In order to know the future behaviour of INR/USD and the closing share prices of Reliance Industries limited, the stationary probability distributions for both the transition probability matrix and for the observed probability matrix are obtained in this section. A stationary probability distribution of a Markov chain is a probability distribution that remains unchanged in the process with the change of time. The stationarity matrix is achieved by multiplying the same matrix number of times until we get the matrix with identical elements in each column. These stationarity conditions for the TPM and OPM are obtained from the real time historical data in the next subsections 6.1 and 6.2 respectively.

7.1. Stationarity for transition probability matrix

Forecasting of long run behavior of INR is very meaningful for investors investing in Reliance industries as its closing share prices is influenced by the change in INR with respect to USD. This long run behavior of INR is observed by determining the higher order TPM. The stationary matrix is obtained by multiplying the TPM (A) in using R software number of times until all the column entries are identical in the matrix. The TPM obtained in section 4.1 is

$$A = \begin{array}{c c} & \begin{array}{cc} D & I \end{array} \\ \begin{array}{c} D \\ I \end{array} & \begin{bmatrix} 0.483051 & 0.516949 \\ 0.491935 & 0.508065 \end{bmatrix} \end{array}$$

The INR/USD data set is an Ergodic Markov chain, which means it is irreducible, positive recurrent, aperiodic, and time homogeneous. The assumption helps in predicting the closing share prices' long-term behaviour. The stationary matrix is created by multiplying the TPM four times in R Software until all we observe the identical column elements. Here the stationary matrix for the INR is obtained in A^4.

$$A^4 = \begin{array}{c} \begin{array}{cc} D & I \end{array} \\ \begin{bmatrix} 0.4876033 & 0.5123967 \\ 0.4876033 & 0.5123967 \end{bmatrix} \end{array}$$

At n=4, the likelihood of the INR value remaining in the same state, regardless of its previous day's state. The higher-order TPM (A^4) computed above demonstrates that the TPM tends to the state of equilibrium or steady-state after the 4^{th} trading day since 486 trading days. Following that, the TPM of INR remains stable for the next few trading days. This steady state of INR interprets as follows. There are 51.2% chances of increase in INR in near future irrespective of its initial states weather it is in D or I and the probability that there will be a decrease in INR is 48% on the fourth trading day and onwards irrespective of its initial states is weather D or I.

7.2. Stationary distribution of OPM

The matrix B obtained in section 4.2, like the matrix A, is an irreducible, positive recurrent, aperiodic matrix that is independent of where we begin. If we continue the chain for a long time, the distribution of Y_n will converge to a constant, i.e., its stationary distribution. The OPM calculated in section 4.2 as follows.

$$B = \begin{array}{c c} & \begin{array}{cc} F & R \end{array} \\ \begin{array}{c} D \\ I \end{array} & \begin{bmatrix} 0.466942 & 0.533058 \\ 0.518519 & 0.481481 \end{bmatrix} \end{array}$$

Following the same procedure as in 6.1, the stationary observed probability matrix is arrived at the 6th step which is presented in matrix B^6.

$$B^6 = \begin{array}{c} \\ D \\ I \end{array}\!\!\begin{array}{c} \begin{array}{cc} F & \qquad R \end{array} \\ \begin{bmatrix} 0.4930869 & 0.5069131 \\ 0.4930869 & 0.5069131 \end{bmatrix} \end{array}$$

The probability of stock prices for Reliance remained stationary at t=6, independent of which state it was on the previous day. The matrix (B^6) computed above shows that after the 6th trading days since 486 trading days, OPM tends to the steady state or the state of equilibrium. After then, the OPM remains same/unaltered for the following trading days. The above stationary distribution B^6 of OPM for Reliance's share prices reveals the following information. There is 49.3% chance that closing share prices of Reliance will witness the state of fall (F) by 6th day irrespective of the nature of previous day’s hidden state (D or I in the INR) and approximately 50.7% chance that the share prices of Reliance will end up be in rise (R) on 6th day and onwards, irrespective of its state on the previous day, no matter whether there is the decrease or increase in INR/USD.

8. Conclusion

The Hidden Markov model is a well-established and well-researched method for analyzing and predicting time series behavior from previous historical time series data sets. We suggest the use of an HMM, a novel technique, to forecast the future behaviour in a time series in this paper. This study is based on two parts; in the primary part, the mathematical formulae of the probability distributions of each visible states with single day, two days sequence and three days sequence are derived in section 3.1, 3.2 and 3.3 respectively in equations 3.1.3, 3.2.3 and 3.3.3. In addition to this, for every derived probability distribution, the formulae for various descriptive statistical measures are derived in section 3.1, 3.2, 3.3. Furthermore, the expression for different generating functions such as the moment generating function, probability generating and the function are obtained in section 3.1part III, 3.2 part III and 3.3 part IV. Whereas the second part of the paper is based on verification and the empirical study of all the formulae derived through the real time historical data sets for the effectiveness and the verification of the model. The statistical analysis of an HMM are done using Microsoft excel and R Software. The study of chance processes for which information of each event is significant solely in forecasting the next outcome is obvious from the results produced. The results in an IPV in section 4.3 represents that the likelihood of decreasing and increasing the INR with respect USD are 49% and 51% respectively. The TPM obtained in section 4.1 explains the probability that INR will decrease and increase in the next day given that today's INR is on decreasing state are 48% and 52% respectively and the probability that INR will decrease and increase in the next day given that it is on increasing state are 49% and 51% respectively. In the same way OPM obtained in section 4.2 explains that the likelihood of falling and rising the share price of Reliance industries limited in the next day given that there is decrease in the INR are 47% and 53% respectively and the probability of falling and rising the share price of Reliance given that there is an increase in INR are 52% and 48% respectively.

Probabilities of the visible states in a single day are presented in table 4 where we observe that probability of rising the share price of Reliance in more than its falling. The probability distribution for the falling state in one day length is presented in table 5 and the descriptive statistical measures such as mean, variance, skewness kurtosis and coefficient of variation are presented in table 6. Probabilities of the visible states in two days sequence are presented in table 7 where we observe the likelihood of rising the share price of Reliance on second day given that the first day it was on rising state is maximum as compared to other sequences. The probability distribution along with its statistical parameters are shown in table 8 and 9 respectively. Similarly, the probabilities of three visible states are presented in table 10 in which we have observed that the probability of three consecutive days rise in the share price of Reliance industries limited is maximum with respect to other different sequences.

Finally, the steady-state probability distribution of the OPM is obtained at B^6 that is at t=6, the probability of stock prices for Reliance remained stationary, independent of which state it was on the previous day. The stationarity probability distribution reveals that the probability of fall in the share price of Reliance on 6^{th} day and onwards is 49.3% and probability of rise in the share price of Reliance on 6^{th} day and onwards is 50.7% irrespective of the nature of previous of day's state weather the INR is on decreasing (D) or increasing (I) state. These results suggest that HMM can be effectively used to analysis and forecast time series historical time series data. The proposed model is useful for the investors to predict the variations of share values and will decide when to sell, when to hold and when to purchase a share of certain company say for example Reliance industries in this study. Since the probability distributions are derived for the visible states of length one, two and three, the future work can be done up to n days sequence. Researchers can make use of an HMM with N number of hidden states and M number of visible states.

References

Adam, M. H. M. (2009): Financial markets: the recent experience of a developing economy, *Savings and Development*, 27-40.

Amiens, E. O. and Osamwonyi, I. O. (2022): Stock price forecasting using hidden Markov model, *International Journal of Information and Decision Sciences*, **14(1)**, 39-59.

Ang, A. and Bekaert, G. (2002): International asset allocation with regime shifts, *The review of financial studies*, **15(4)**, 1137-1187.

Baum, L. E. and Petrie, T. (1966): Statistical inference for probabilistic functions of finite state Markov chains, *The annals of mathematical statistics*, **37(6)**, 1554-1563.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. (1970): A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains, *The annals of mathematical statistics*, **41(1)**, 164-171.

Butler, K. C. and Malaikah, S. J. (1992): Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia, *Journal of Banking & Finance*, **16(1**), 197-210.

Chang, Q. and Hu, J. (2022): Application of Hidden Markov Model in Financial Time Series Data, *Security and Communication Networks*, 2022.

Chen, Z., Xia, T., Li, Y. and Pan, E. (2018): Degradation modeling and classification of mixed populations using segmental continuous hidden Markov models, *Quality and Reliability Engineering International*, **34(5)**, 807-823.

Choi, I. (1999): Testing the random walk hypothesis for real exchange rates, *Journal of Applied Econometrics*, **14(3)**, 293-308.

Doubleday, K. J. and Esunge, J. N. (2011): Application of Markov chains to stock trends, *Journal of Mathematics and Statistics*, **7(2)**, 103-106.

Fu, Z. J. and Wu, Q. Q. (2018): Research of market index quantitative timing based on hidden Markov model, *J. Xiamen Univ. Nat. Sci*, **57**, 404-412.

G. F. Dar, T. R. Padi and S. Rekha (2022): Stock Price Prediction Using a Markov Chain Model: A Study for TCS Share Values, *Advances and Applications in Statistics*, **80**, 83-101.

G. F. Dar, T. R. Padi, S. Rekha and Q. F. Dar (2022): Stochastic Modelling for the Analysis and Forecasting of Stock Market Trend Using Hidden Markov Model, *Asian Journal of Probability and Statistics*, **18(1)**, 43-56.

Guidolin, M. and Timmermann, A. (2007): Asset allocation under multivariate regime switching, *Journal of Economic Dynamics and Control*, **31(11)**, 3503-3544.

Gupta, A. and Dhingra, B. (2012): Stock market prediction using hidden Markov models, In 2012 Students Conference on Engineering and Systems, *IEEE*, 1-4.

Hassan, M. R. and Nath, B. (2005, September): Stock market forecasting using hidden Markov model: a new approach, In 5th International Conference on Intelligent Systems Design and Applications (ISDA'05), *IEEE*, 192-196.

Holzmann, H. and Schwaiger, F. (2016): Testing for the number of states in hidden Markov models, *Computational Statistics & Data Analysis*, **100**, 318-330.

Huang, J. C., Huang, W. T., Chu, P. T., Lee, W. Y., Pai, H. P., Chuang, C. C. and Wu, Y. W. (2017): Applying a Markov chain for the stock pricing of a novel forecasting model, *Communications in Statistics-theory and Methods*, **46(9**), 4388-4402.

Kim, E. C, Jeong, H. W. and Lee, N. Y. (2019): Global asset allocation strategy using a hidden markov model, *Journal of Risk and Financial Management*, **12(4)**, 168.

Kritzman, M., Page, S. and Turkington, D. (2012): Regime shifts: Implications for dynamic strategies, *Financial Analysts Journal*, **68(3)**, 22-39.

Levinson, S. E, Rabiner, L. R. and Sondhi, M. M. (1983): An introduction to the application of the theory of probabilistic functions of a Markov process to automatic speech recognition, *Bell System Technical Journal*, **62(4)**, 1035-1074.

Li, X., Parizeau, M. and Plamondon, R. (2000): Training hidden Markov models with multiple observations-a combinatorial method, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **22(4)**, 371-377.

Liu, Z. and Wang, S. (2017): Decoding Chinese stock market returns: Three-state hidden semi-Markov model, *Pacific-Basin Finance Journal*, **44**, 127-149.

Nguyen, N. (2018): Hidden Markov model for stock trading, *International Journal of Financial Studies*, **6(2)**, 36.

Nguyen, N. and Nguyen, D. (2015): Hidden Markov model for stock selection, *Risks*, **3(4)**, 455-473.

Somani, P, Talele, S. and Sawant, S. (2014): Stock market prediction using hidden Markov model, In 2014 IEEE 7th joint international information technology and artificial intelligence conference, *IEEE*, 89-92.

T. R. Padi, G. F. Dar and S. Rekha (2022): Markov Modelling of Indian Stock Market Prices with Reference to State Bank of India, *International Journal of Statistics and Reliability Engineering*, **9(3)**, 481-492.

T. R. Padi, G. F. Dar and S. Rekha (2022): Stock Market Trend Analysis and Prediction Using Markov Chain Approach in the Context of Indian Stock Market, *IOSR Journal of Mathematics*, **18(4)**, 40-48.

Tuyen, L. T. (2013): Markov financial model using hidden Markov model, *International Journal of Applied Mathematics and Statistics*, **40(10)**, 72-83.

Authors and Affiliations

Tirupathi Rao Padi[1], Gulbadin Farooq Dar[2] and Sarode Rekha[3]

Tirupathi Rao Padi
Email: drtrpadi@gmail.com

Sarode Rekha
Email: statisticsrekha@gmail.com

[1]RSMS, Department of Statistics, Pondicherry University, Puducherry, 605014.
[2,3]Department of Statistics, Pondicherry University, Puducherry, 605014.