Adv. Math. Fin. App., 2024, **9**(2), P.367-381

Advances in Mathematical Finance & Applications
www.amfa.iau-arak.ac.ir
Print ISSN: 2538-5569
Online ISSN: 2645-4610
Doi: 10.22034/amfa.2021.1923297.1565

Case Study

The Improved Semi-Parametric Markov Switching Models for Predicting Stocks Prices

Hossein Naderi[a], Mehrdad Ghanbari[a, *], Babak Jamshidi Navid[a], Arash Nademi[b]

[a] *Department of Accounting, Kermanshah Branch, Islamic Azad University, Kermanshah, Iran.*
[b] *Department of Statistics, Ilam Branch, Islamic Azad University, Ilam, Iran.*

ARTICLE INFO
Article history:
Received 2021/02/17
Accepted 2021/10/19

Keywords:
Strategies for buying and selling
Kernel function
EM algorithm
Markov Switching Models

ABSTRACT

The modeling of strategies for buying and selling in Stock Market Investment have been the object of numerous advances and uses in economic studies, both theoretically and empirically. One of the popular models in economic studies is applying the Semi-parametric Markov Switching models for forecasting the time series observations based on stock prices. The Semi-parametric Markov Switching models for these models are a class of popular methods that have been used extensively by researchers to increase the accuracy of fitting processes. The main part of these models is based on kernel and core functions. Despite of existence of many kernel and core functions that are capable in applications for forecasting the stock prices, there is a widely use of Gaussian kernel and exponential core function in these models. But there is a question if other types of kernel and core functions can be used in these models. This paper tries to introduce the other kernel and core functions can be offered for good fitting of the financial data. We first test three popular kernel and four core functions to find the best one and then offer the new strategy of buying and selling stocks by the best selection on these functions for real data.

1 Introduction

The index of the stock exchange and its subsets in the financial markets, as one of the most important benchmarks of the movement of stock prices of stock companies, is of particular importance. The Stock Exchange Index is derived from the stock price movements of all companies in the market, and thus enables the analysis of the price movement in the stock market. Understanding and examining the behavior of this index and its subsidiaries has always been the focus of researchers, economists and capital market activists since the formation of capital markets. Nowadays, there is a lot of research on stock indexes in financial markets of different countries to scientifically model the movement of stock price

* Corresponding author. Tel.: +98918388754
E-mail address: *mehrdadghanbary@yahoo.com*

information. In addition, accurate forecasting of the stock price trends of companies in the form of modeling that is relevant to the overall index process and its subsidiaries is very important in providing useful trading strategies.

Predicting stock prices is one of the most complex issues because the stock market is essentially non-linear and influenced by probabilistic issues. In addition, the stock market is affected by a variety of factors including political and social crises, economic performance, investor behavior, global prices and more. Following the efforts of economic and statistical scientists, new methods have been developed to predict prices in the stock market. Nowadays, nonlinear models of Markov switching, as well as non-parametric and semi-parametric estimation methods, are among these methods. Recent research into the prediction of stock price trends suggests that these models outperform traditional methods such as the ARIMA, GARCH, and Regression models. (See Nademi and Nademi [16], Chang et al. [4], Von Ganske [22], Billio et al. [2], Doaei et al. [11], Davoodi et al. [8], Aghaeefaret al. [1]).

The many investigations in economic and financial mathematics focused on what makes an investor profitable in the stock market. These studies can aid the researchers to decrease the investment risks and increase opportunities for high return of gaining. One of the important questions in the stock market is when the investors can buy the stocks and when they can sell their stocks. In research economic papers, there are two aspects of analysis: fundamental and technical analysis. In fundamental aspects, the researchers find the reasons of changing stock prices, in response to reasons of changing prices that caused from exogenous geopolitical events, supply disruptions or financial operation of the companies and etc. But technical analysis noted more the statistical and probabilities rules governed by processes of the data. In aspect of technical analysis, there are a lot of models in time series to capture the stock prices.

The Semi-parametric Markov switching models are the popular models in time series that are applied most widely in financial and economic data. These models exhibit abrupt changes in behavior of time series data, called switches of regimes, where the switching between the regimes is controlled by a hidden Markov Chain process. In semi-parametric class of algorithms, a special function, called kernel function and core functions, are used. The selection of proper kernel and core function is important item for estimating the parameters. Such that, if we apply the proper types, we can have a fast and unbiased estimating process. So, offering the best kernel and core functions for estimation algorithms can be essential for modelling process. In this paper, we first focus on selecting the best kernel and core functions in a special class of Markov switching models called semi-parametric Markov switching offered by Nademi and Farnoosh [15] for modeling the time series data and then offer the new strategy of buying and selling stocks by the best selected kernel and core function of this model on real data.

In the next section, the theoretical fundamentals and research background of this field will be introduced. Section 3 discuss on the offered kernel and core functions. Finally, section 4 probe the best selection of these functions and offer the new buying and selling strategy for stock markets.

2 Theoretical Fundamentals and Research Background

2.1 Research Background

The forecasting and offering strategies of stock buying and selling has been the object of plentiful expansions and applications over the past two decades, both theoretically and empirically. There are several attempts in this field. Pourzamani et al. [19] Compared stock buying and selling strategies in long-term investment using filtering, buying and holding methods and the moving average of the market. They showed the moving average of the market and the return of the buying and holding method is higher than the moving average method. Rastegar and Dastpak [20] offered a model entitled" Developing a High-Frequency Trading system with Dynamic Portfolio Management using Reinforcement Learning in Iran Stock Market". Results showed that, the proposed model outperformed the Buy and Hold strategy in Normal and Descending markets. Davallou and Meskinimood [7] examinated of trading strategy based on Stochastic Dominance. They showed the pricing of the random dominance factor in the Tehran Stock Exchange is approved. Sharif-far et al. [21] applied the assessment of the optimal Deep Learning Algorithm on Stock Price Prediction (Long Short-Term Memory Approach). The results showed better performance of LSTM architecture with Drop Out layer than RNN model. Pashaei and Dehkharghani [18] examined stock market modeling using Artificial Neural Network and compared with classical linear models

In the other hand, the most widely applied type of models is AR-ARCH models. The combination of autoregressive (AR) processes and autoregressive conditionally heteroscedastic (ARCH) processes, the so-called AR-ARCH process, are well created and very general models.

These findings clearly show a potential source of unknown structure, to explain that the form of the variance is relatively inflexible and held fixed throughout the entire sample period. Hence the estimates of an AR-ARCH model may suffer from a substantial bias in the persistence parameter. So, models in which the parameters are allowed to change over time may be more feasible for modeling processes. Recently, The Markov Switching models have repeatedly applied for making switching regimes processes which allow for more flexibility in modeling data which only show locally a homogeneous behavior. The Markov Switching models are the popular models in time series that are applied most widely in financial and economic data. These models exhibit abrupt changes in behavior of time series data, called switches of regimes, where the switching between the regimes is controlled by a hidden Markov Chain process. (See Chang et al. [4], Von Ganske [22], Billio et al. [2], Di Persio and Frigo [9-10], Neale et al. [17]).

Recently, Semi-parametric Markov switching models have repeatedly applied for making switching regimes processes and every of them offer an algorithm for estimating the parameters. In this respect, the combination of parametric and nonparametric methods, called semi-parametric algorithms, are popular and most broadly applied. (See Chan and Wang [3], Chang et al. [5], Chen et al. [6], Gupta et al. [14], Gu and Balasubramanian [13], Nademi and Farnoosh [15]). In the next subsection, the theory of these models will be explained.

2.2 Theoretical Fundamentals

This section consists of two subsections. In the first subsection, we introduce the Markov switching model introduced by Nademi and Farnoosh [15] and in the second subsection, their algorithm for estimating the parameters will be reviewed. Note that, their semi-parametric algorithm is a part of more general algorithm as EM algorithm that apply for class of Markov switching models.

2.2.1. The Model

Suppose $Y_1, \dots, Y_N$ are part of a strictly stationary time series that are generated by the following semi-parametric switching model

$$Y_t = \sum_{k=1}^{M} Z_{tk}\big(\mu(Y_{t-1}, Y_{t-2}; \rho_k) + \sigma(Y_{t-1}, Y_{t-2}; \omega_k, \alpha_k, \beta_k)u_{t,k}\big), \tag{1}$$

such that,

$$\mu(Y_{t-1}, Y_{t-2}; \theta_k, \rho_k) = g(Y_{t-1}, \theta_k)\xi_k(Y_{t-1}) + \rho_k\big(Y_{t-1} - g(Y_{t-2}, \theta_k)\xi_k(Y_{t-2})\big), \quad (2)$$

and

$$\sigma^2(Y_{t-1}, Y_{t-2}; \omega_k, \alpha_k, \beta_k) = \omega_k + \alpha_k Y_{t-1}^2 + \beta_k Y_{t-2}^2,$$

with

$$Z_{tk} = \begin{cases} 1 & Q_t = k \\ 0 & otherwise, \end{cases}$$

where switching between the regimes is controlled by a hidden Markov chain Q_t, with values in $\{1, \dots, M\}$, and the residuals $u_{t,k}, t = 1, \dots, N, k = 1, \dots, M$ are i.i.d. random variables with mean 0 and variance 1. $Z_t = (Z_{t1}, Z_{t2}, \dots, Z_{tM})^T$ are random variables which assume as values of the unit vectors $e_1, e_2, \dots, e_M \in R^M$, i.e. exactly one of the Z_{tk} is 1, and the others are 0. The stationary distribution of the hidden regime process is given by the $M \times M$ transition probability matrix A, i.e, $A_{jk} = pr(Q_t = k|Q_{t-1} = j)$. $\xi_k(x)$ is a nonparametric adjustment factor and $g(x, \theta_k)$ is a known function of x and θ_k , called the core function.

2.2.2. The EM Algorithm Based on Semi-Parametric Method

Supposing the definition of $Y^{(N)} = (Y_1, Y_2, \dots, Y_N)$ and $Z^{(N)} = (Z_1, Z_2, \dots, Z_N)$, Nademi and Farnoosh [15] applied a special class of log likelihood function, called "complete" log likelihood function, by the following form

$$l_c\big(v, A|Y^{(N)}, Z^{(N)}\big) = log\pi_{q_1} \sum_{t=2}^{N} logA_{q_{t-1},q_t} + \sum_{t=2}^{N}\sum_{k=1}^{M} Z_{tk} log \frac{1}{\sigma(Y_{t-1}, Y_{t-2}; \omega_k, \alpha_k, \beta_k)} \varphi\left(\frac{Y_t - \mu(Y_{t-1}, Y_{t-2}; \theta_k, \rho_k)}{\sigma(Y_{t-1}, Y_{t-2}; \omega_k, \alpha_k, \beta_k)}\right)$$

Where $v = (\theta_1, \dots, \theta_M, \rho_1, \dots, \rho_M, \omega_1, \dots, \omega_M, \alpha_1, \dots, \alpha_M, \beta_1, \dots, \beta_M)^T \in V$ and $\varphi(.)$ is the normal density with mean $\mu(.)$ and standard deviation $\sigma(.)$. The word "complete" refer to this definition that if we would have observed the complete data $\big(Y^{(N)}, Z^{(N)}\big)$, instead of just $Y^{(N)}$, we could maximize the complete data log likelihood (see Franke et al. [9]), instead of ordinary log likelihood.

Applying this method led to use of Expectation and Maximization algorithm known as EM algorithm. The EM algorithm repeats between drawing the unseen variables Z_{tk} by conditional expectations Z_{tk} given the seen data $Y^{(N)}$ and using a elementary estimate of the parameters on the one phase (E-step), and by maximizing $l_c\big(v, A|Y^{(N)}, Z^{(N)}\big)$ to get an update of approximations of A or v on the other phase (M-step). These two phases until assuring certain stopping criteria are iterated. The algorithm offered with the EM algorithm can be summarized to the following steps.

E-step: Suppose $\hat{\pi}_1, \dots, \hat{\pi}_M, \hat{\rho}_1, \dots, \hat{\rho}_M, \hat{\theta}_1, \dots, \hat{\theta}_M, \hat{\omega}_1, \dots, \hat{\omega}_M, \hat{\alpha}_1, \dots, \hat{\alpha}_M$ and $, \hat{\beta}_1, \dots, \hat{\beta}_M$ are given. S0, the conditional expectation of the unseen variables Z_{tk} given $Y^{(N)}$are calculated by

$$C_{tk} = E[Z_{tk}|Y^{(N)}] = \frac{\alpha_k^t \beta_k^t}{\sum_{i=1}^{M} \alpha_i^t \beta_i^t} \quad k = 1, \dots, M \quad t = 1, \dots, N,$$

where α_i^t and β_i^t are estimated by following recursive relations

$$\alpha_j^{t+1} = \varphi\left(Y_{t+1}; \mu(Y_t, Y_{t-1}; \theta_j, \rho_j), \sigma(Y_t, Y_{t-1}; \omega_j, \alpha_j, \beta_j)\right) \sum_{k=1}^{M} A_{kj} \alpha_k^t,$$

and

$$\beta_j^t = \sum_{k=1}^{M} \beta_k^{t+1} \varphi\big(Y_{t+1}; \mu(Y_t, Y_{t-1}; \theta_k, \rho_k), \sigma(Y_t, Y_{t-1}; \omega_k, \alpha_k, \beta_k)\big) A_{jk},$$

where $\varphi\big(Y_{t+1}; \mu(Y_t, Y_{t-1}; \theta_k, \rho_k), \sigma(Y_t, Y_{t-1}; \omega_k, \alpha_k, \beta_k)\big)$ is the normal density with mean $\mu(.)$ and standard deviation $\sigma(.)$.

M-step: Suppose the approximations C_{tk} for the unseen variables Z_{tk} are given. Then, the transition probabilities are calculated by

$$\hat{A}_{ij} = \frac{\sum_{t=1}^{N} \delta_{ij}^{t,t+1}}{\sum_{t=1}^{N} C_{ti}},$$

where $\delta_{ij}^{t,t+1}$ are the joint conditional probability of $Q_t = i$ and $Q_{t+1} = j$ given the entire sequence of observations ($Y^{(N)}$) estimated by

$$\delta_{ij}^{t,t+1} = p\big(Q_t = i, Q_{t+1} = j \big| Y^{(N)}\big) = \frac{\beta_j^{t+1} \varphi\left(Y_{t+1}; \mu(Y_t, Y_{t-1}; \theta_j, \rho_j), \sigma(Y_t, Y_{t-1}; \omega_j, \alpha_j, \beta_j)\right) A_{ij} \alpha_i^t}{\sum_{k=1}^{M} \alpha_k^t \beta_k^t}.$$

The probabilities $\pi_1, \dots, \pi_M$ are approximated by

$$\widehat{\pi_k} = \frac{1}{N} \sum_{t=1}^{N} C_{tk}, \; k = 1, \dots, M,$$

and the functions $\mu(x, y; \theta_k, \rho_k)$ are estimated by

$$\mu(x, y; \hat{\theta}_k, \hat{\rho}_k) = g(x, \hat{\theta}_k) \hat{\xi}_k(x) + \hat{\rho}_k \left(x - g(y, \hat{\theta}_k) \hat{\xi}_k(y)\right),$$

such that, $(\hat{\theta}_k, \hat{\rho}_k)$ gets from $(\hat{\theta}_k, \hat{\rho}_k) = argmin Q_n(\theta_k, \rho_k), \theta_k, \rho_k \in \mathrm{V}, |\rho_k| < 1$ for $k = 1, \dots, M$, where $Q_n(\theta_k, \rho_k)$ is

$$Q_n(\theta_k, \rho_k) = \sum_{t=2}^{N} C_{tk} (Y_t - g(Y_{t-1}, \theta_k) - \rho_k (Y_{t-1} - g(Y_{t-2}, \theta_k)))^2,$$

and $\hat{\xi}_k(x)$ is

$$\hat{\xi}_k(x) = \frac{\sum_{t=2}^{N} C_{tk}\left[k\left(\frac{Y_{t-1}-x}{h_k}\right)g(Y_{t-1},\hat{\theta}_k)Y_t + k\left(\frac{Y_{t-2}-x}{h_k}\right)g(Y_{t-2},\hat{\theta}_k)Y_{t-1}\right]}{\sum_{t=2}^{N} C_{tk}\left[k\left(\frac{Y_{t-1}-x}{h_k}\right)g^2(Y_{t-1},\hat{\theta}_k) + k\left(\frac{Y_{t-2}-x}{h_k}\right)g^2(Y_{t-2},\hat{\theta}_k)\right]}, \qquad (3)$$

where $k(.)$ is a Gaussian Kernel function and $(\omega_k, \alpha_k, \beta_k)$ are estimated by

$$(\hat{\omega}_k, \hat{\alpha}_k, \hat{\beta}_k) = Arg\, max \sum_{t=2}^{N} C_{tk}\, log \frac{1}{\sigma(Y_{t-1}, Y_{t-2}; \omega_k, \alpha_k, \beta_k)} \varphi\left(\frac{\hat{e}_{tk}}{\sigma(Y_{t-1}, Y_{t-2}; \omega_k, \alpha_k, \beta_k)}\right),$$

for $k = 1, \ldots, M$, where $\hat{e}_{tk} = Y_t - \mu(x, y; \hat{\theta}_k, \hat{\rho}_k)$ denotes the sample residuals. The optimal selection of the bandwidth h_k are also estimated by

$$\hat{h}_k = \arg\max_{h_k} \sum_{t=2}^{N} C_{tk}\left[Y_t - g(Y_{t-1},\hat{\theta}_k)\hat{\xi}_k(Y_{t-1}) - \hat{\rho}_k\left(Y_{t-1} - g(Y_{t-2},\hat{\theta}_k)\hat{\xi}_k(Y_{t-2})\right)\right]^2.$$

The estimate of the parameters are obtained by iterating these two steps (E-step and M-step) until convergence.

In estimating the functions $\hat{\xi}_k(x)$ and core function, they applied the Gaussian Kernel function and exponential core function. But there is a question that if other types of kernel function and core function can improve performance of the semi-parametric algorithm. We want to trial some other types of kernel and core functions that are popular in mathematics field. These functions consist of: Uniform and Triangle. We also apply the Gaussian kernel to compare this function with the candidate kernel functions. We also test some core functions to compare the ability of them in improvement of the EM algorithm.

3 Materials and Methods

The type of study used in this study is correlational studies. The statistical population of this study includes all indices as well as stock prices of companies listed on the Tehran Stock market. The sample data consist of financial observations, including of Bank's index (industry group) of Tehran Stock market the period March 25, 2018 to March 19, 2019, downloaded from "http://tse.ir/archive.html", where the sample size is 243. So, the variable of our study is Bank's index which is presented by Y_t (Bank's index in time t).

In the first step, we must determine the number of regimes in observations. This can demonstrate by drawing the sample path of data and observing changes trends as increasing and decreasing function. Figure 1 (blue line) shows the sample path the data. Applying this plot, we applied the step function (red line), the down step (regime=1) and upper step (regime=2), to indicate the regimes such that we named increasing trends and decreasing trends by regimes =1 and regimes=2, respectively.

In the second step, we apply the semiparametric model (1) to fit the observations and then, we use of EM algorithm to estimate the parameters of the model. Note that, in one part of M-step, we must select a kernel function ($k\left(\frac{Y_{t-1}-x}{h_k}\right)$) and core function ($g(x, \hat{\theta}_k)$) .So, In the third step, the selected core functions are as follows

$$g_1(x,\theta) = \exp(x\theta), g_2(x,\theta) = \exp(\sqrt{x}\theta), g_3(x,\theta) = \frac{\exp(x\theta)}{1+\exp(x\theta)}, g_4(x,\theta) = \theta\sin(x).$$

The kernel functions are selected by

$$k_1(u) = \frac{1}{2} I_{[-1,1]}(u), k_2(u) = (1-|u|)I_{[-1,1]}(u), k_3(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}u^2\right),$$

where $k_1(u), k_2(u)$ and $k_3(u)$ are called Uniform, Triangle and Gaussian kernel functions, respectively. For the sake of simplicity in definition of the models, we will call the semi-parametric Markov switching models by MS-SEMI-k(i)-G(i) based on $k_i(u)$ and $g_i(x,\theta)$.
For comparing the models, we apply two indices the square Root of Mean Squared Error (RMSE) and classifying index "Max$C_{tk}, k = 1, \dots, M$". The square Root of Mean Squared Error (RMSE) is defined by the following form:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(Y_t - \hat{Y}_t\right)^2},$$

and the classifying index "Max$C_{tk}, k = 1, \dots, M$" is defined by: "Y_t is belonging to regime k if and only if $C_{tk} = \underset{i=1,\dots,M}{\text{Max}} C_{ti}$". This index is suitable in evaluation of the models, such that high amounts of this index show the correction classifying of the best model in categorizing the observations in the right regimes. We also apply a step function to show the regimes based on behavior the observations (see Nademi and Farnoosh [15] and Nademi and Nademi [16]).

4 Results

In the first step, we focus on the plot of the data. Figure 1 (blue line) shows the sample path the data. Based on to this plot, we applied the step function (red line), the down step (regime=1) and upper step (regime=2), to indicate the regimes such that we named increasing trends and decreasing trends by regimes =1 and regimes=2, So, we considered Twelve Semi-parametric Markov Switching models (MS-SEMI-K(i)-G(j), i=1,2,3, j=1,2,3,4) based on three kernel functions and four core functions with two regimes (M=2).

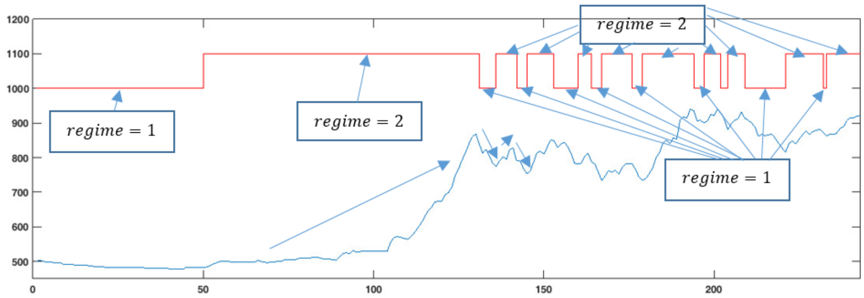

Fig.1. Bank's index data (blue line) and Step function (red line).

In the second and third steps, we apply the semiparametric model (1) and focus on selecting the core function for every kernel function and then compare the best models based on the best kernel function. Table 1 lists the estimated parameters for kernel function Uniform where we have the semi-parametric Markov switching models MS-SEMI-K(1)-G(1), MS-SEMI-K(1)-G(2), MS-SEMI-K(1)-G(3) and MS-SEMI-K(1)-G(4). The results of RMSE criteria for these models show the model MS-SEMI-K(1)-G(4) has the minimum RMSE (0.0725) among the other models. So, we can say that the core function $g_4(x,\theta) = \theta \sin(x)$ is proper for capturing the data. After this core function, the model MS-SEMI-K(1)-G(1) with RMSE, 0.0731 can be selected for observations with the structure with Uniform kernel function. Figure 2 shows the classification of the data based on index $Max(C_{tk})$ where the values of

$Max(C_{t1}, C_{t2})$ for four models, which are all greater than 0.5, show the ability of semi-parametric models in classifying the data. On the other hand, in the MS-SEMI-K(1)-G(4) model the belonging probabilities greater than 0.85 indicate that this model is more powerful than the other models in classifying the observations.

Table 1. The Estimated Parameters Based on Uniform Kernel Function

The Parameters	MS-SEMI-K(1)-G(1)	MS-SEMI-K(1)-G(2)	MS-SEMI-K(1)-G(3)	MS-SEMI-K(1)-G(4)
θ_1	-1.1501	-3.3104	-1.3116	-1.1032
θ_2	-6.4251	-4.3148	-3.2190	-6.2163
ρ_1	0.4216	0.3148	0.5184	0.4084
ρ_2	0.6218	0.7315	0.4023	0.5032
ω_1	0.0015	0.0051	0.0010	0.0021
ω_2	0.0001	0.0016	0.0003	0.0013
α_1	0.0037	0.0204	0.0216	0.0001
α_2	0.0028	0.0381	0.0203	0.0061
β_1	0.0104	0.0204	0.0016	0.0053
β_2	0.0265	0.0110	0.0367	0.0011
π_1	0.3401	0.3721	0.5169	0.3606
π_2	0.6599	0.6279	0.4831	0.6394
A_{12}	0.6112	0.5143	0.4035	0.6203
A_{21}	0.3150	0.3048	0.4318	0.3498
h_1	0.0012	0.0102	0.0122	0.0016
h_2	0.0034	0.0131	0.0351	0.0010
$RMSE$	0.0731	0.0945	0.0871	0.0725

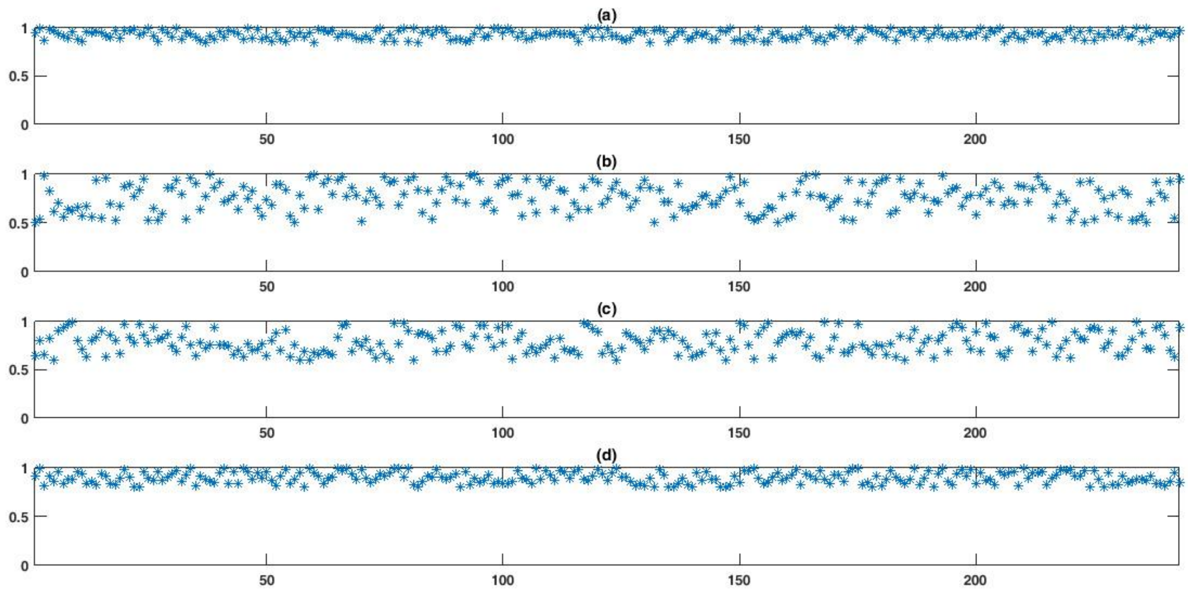

Fig.2. Max(C_{t1}, C_{t2}) for models: (a). MS-SEMI-K(1)-G(1), (b). MS-SEMI-K(1)-G(2), (c). MS-SEMI-K(1)-G(3), (d). MS-SEMI-K(1)-G(4).

Table 2 consists of the estimated parameters based on Triangle kernel function. This table indicate the RMSE for models MS-SEMI-K(2)-G(1), MS-SEMI-K(2)-G(2), MS-SEMI-K(2)-G(3) and MS-SEMI-K(2)-G(4). According to the results, we can find that the model MS-SEMI-K(2)-G(1), with RMSE 0.0945, has the minimum amount of RMSE, comparing the other models. So, the best core function, for Triangle kernel function, is $g_1(x,\theta) = \exp(x\theta)$. After this core function, the model MS-SEMI-K(2)-G(2), with RMSE 0.1134, has the minimum RMSE among the other models. Figure 3 shows the classification of the data based on index $Max(C_{tk})$ for Triangle kernel function where the values of $Max(C_{t1}, C_{t2})$ for four models, which are all greater than 0.5, show the ability of semi-parametric models in classifying the data. On the other hand, in the MS-SEMI-K(2)-G(1) model the belonging probabilities greater than 0.70 indicate that this model is more powerful than the other models in classifying the observations based on Triangle kernel function.

Table 2. The Estimated Parameters Based on Triangle Kernel Function

The Parameters	MS-SEMI-K(2)-G(1)	MS-SEMI-K(2)-G(2)	MS-SEMI-K(2)-G(3)	MS-SEMI-K(2)-G(4)
θ_1	-1.2510	-1.1024	-2.0012	-3.1004
θ_2	-3.5462	-3.6412	-2.1640	-2.3145
ρ_1	0.4325	0.5148	0.2489	0.1624
ρ_2	0.6245	0.6489	0.4327	0.2031
ω_1	0.0026	0.0514	0.0031	0.0321
ω_2	0.0010	0.0379	0.0049	0.1202
α_1	0.0311	0.0302	0.0319	0.1304
α_2	0.0051	0.0942	0.0402	0.0181
β_1	0.0645	0.0234	0.0521	0.0094
β_2	0.0713	0.0824	0.0601	0.0081
π_1	0.5732	0.2859	0.3289	0.3186
π_2	0.4268	0.7141	0.6711	0.6814
A_{12}	0.3489	0.6150	0.6502	0.4316
A_{21}	0.4685	0.2462	0.3186	0.2018
h_1	0.0013	0.0046	0.0027	0.0003
h_2	0.0048	0.0487	0.0062	0.0042
$RMSE$	0.0945	0.1134	0.1246	0.1215

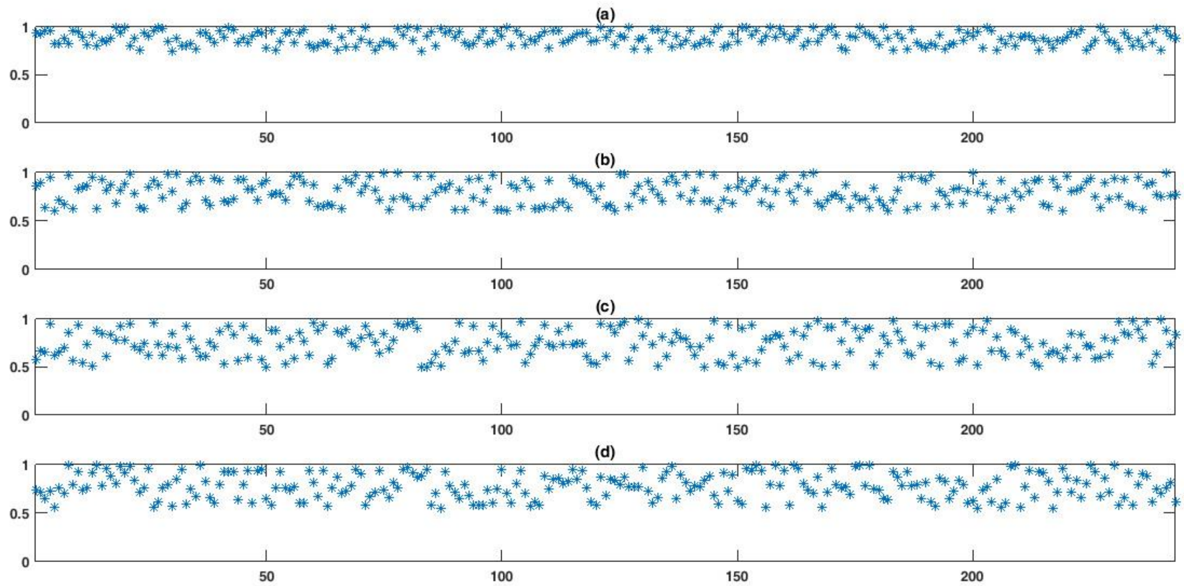

Fig. 3. $\text{Max}(C_{t1}, C_{t2})$ for models: (a). MS-SEMI-K(2)-G(1), (b). MS-SEMI-K(2)-G(2), (c). MS-SEMI-K(2)-G(3), (d). MS-SEMI-K(2)-G(4).

Table 3 shows the estimated parameters based on Gaussian kernel function. Comparing the amounts of RMSE for the models MS-SEMI-K(3)-G(1), MS-SEMI-K(3)-G(2), MS-SEMI-K(3)-G(3) and MS-SEMI-K(3)-G(4), we can see the best model is the model MS-SEMI-K(3)-G(1) in which RMSE is 0.0703. Therefore, we can find that the best core function for Gaussian kernel function is $g_1(x,\theta) = \exp(x\theta)$. After this model, the core function $g_4(x,\theta) = \theta \sin(x)$ of the model MS-SEMI-K(3)-G(4) can be offer as the proper core function based on Gaussian kernel function. Figure 4 shows the classification of the data based on index $Max(C_{tk})$ for Gaussian kernel function where the values of $Max(C_{t1}, C_{t2})$ for four models, which are all greater than 0.6, show the ability of semi-parametric models in classifying the data. On the other hand, in the MS-SEMI-K(3)-G(1) model the belonging probabilities greater than 0.87 indicate that this model is more powerful than the other models in classifying the observations based on Gaussian kernel function.

Table 3. The Estimated Parameters Based on Gaussian Kernel Function

The Parameters	MS-SEMI-K(3)-G(1)	MS-SEMI-K(3)-G(2)	MS-SEMI-K(3)-G(3)	MS-SEMI-K(3)-G(4)
θ_1	-1.0162	-3.0510	-1.4692	-0.8472
θ_2	-2.3201	-2.4210	-4.3150	-1.3496
ρ_1	0.1240	0.3160	0.4682	0.4081
ρ_2	0.4201	0.5015	0.4182	0.2486
ω_1	0.0213	0.0105	0.0013	0.0003
ω_2	0.0010	0.0315	0.0041	0.0008
α_1	0.0114	0.0021	0.0203	0.0214
α_2	0.0315	0.0184	0.0344	0.0804
β_1	0.0243	0.0648	0.0510	0.0034
β_2	0.0152	0.0921	0.0025	0.0107
π_1	0.3812	0.5501	0.3091	0.3329
π_2	0.6188	0.4499	0.6809	0.6671
A_{12}	0.5142	0.4210	0.3162	0.4316
A_{21}	0.3168	0.5147	0.1482	0.2154
h_1	0.0032	0.0032	0.0018	0.0002
h_2	0.0102	0.0012	0.0054	0.0011
$RMSE$	0.0703	0.0791	0.0849	0.0748

Comparing the RMSE of the models, we can find that the MS-SEMI-K(3)-G(1) has the least amount of RMSE among the others. So, the selected kernel function and core function for the data are Gaussian kernel and $g_1(x,\theta) = \exp(x\theta)$, respectively. But, with a closer look, one see that the Uniform kernel can be a strongly rival for selecting the kernel function. Because some RMSE's for this kernel, regardless in selecting the core function $g_1(x,\theta)$, is better than the Gaussian kernel. This indicate that in modeling data, the researchers should not focus on just one kernel or core function.

After estimating the models for the data, we forecasted future of the data for four lags of time. Figures 5, 6 and 7 consist of the estimated joint conditional probabilities for twelve the models. The estimated joint conditional probabilities were defined by $\delta_{ij}^{t,t+1} = p(Q_t = i, Q_{t+1} = j | Y^{(N)})$, such that we can write the joint conditional probability matrix

$$\delta_{ij}^{T} = \begin{pmatrix} p(Q_t = 1, Q_{t+1} = 1 | Y^{(N)}) & p(Q_t = 1, Q_{t+1} = 2 | Y^{(N)}) \\ p(Q_t = 2, Q_{t+1} = 1 | Y^{(N)}) & p(Q_t = 2, Q_{t+1} = 2 | Y^{(N)}) \end{pmatrix},$$

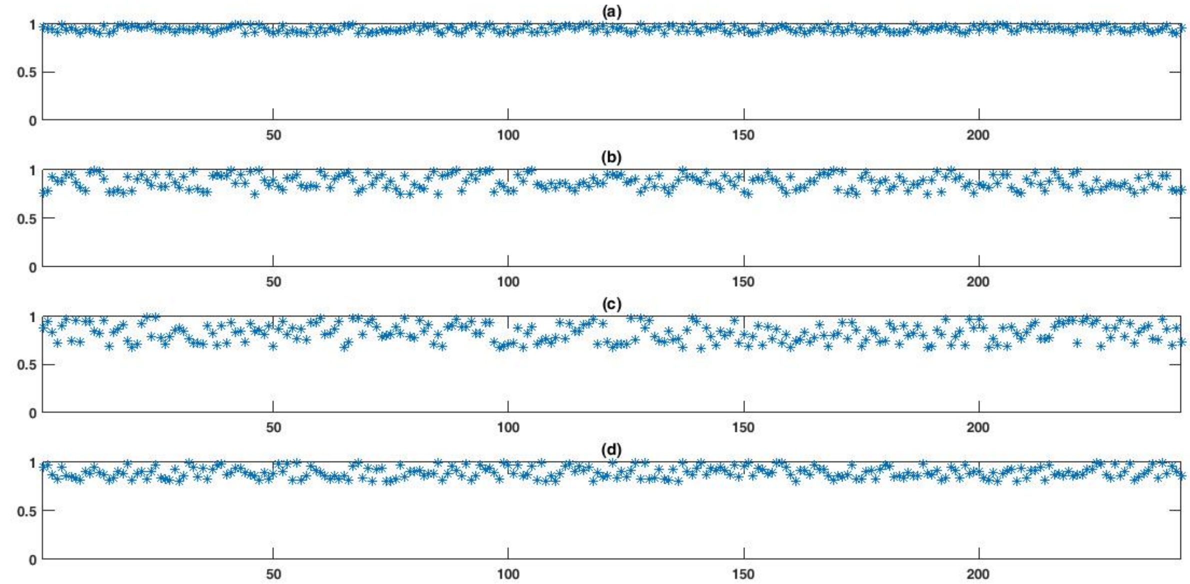

Fig. 4. Max(C_{t1}, C_{t2}) for models: (a). MS-SEMI-K(3)-G(1), (b). MS-SEMI-K(3)-G(2), (c). MS-SEMI-K(3)-G(3), (d). MS-SEMI-K(3)-G(4).

for the models. This matrix can offer the strategy of buying and selling stock in financial markets. Such that, the probability elements of the matrix indicate the behavior of the data in passing time "t" to "t+1". According to structure of the Semi-parametric model (1) and dependency degree of observation Y_t that is degree of 2, based on dependency of observation Y_t on Y_{t-1} and Y_{t-2}, we can see that the plots can predict strongly the behavior of observations just for two lags of future time and after that there is not a certain decision for behavior the observations for future lags. Such that, these figures show that after the lag 4 the joint conditional probabilities almost converge to the same amount.

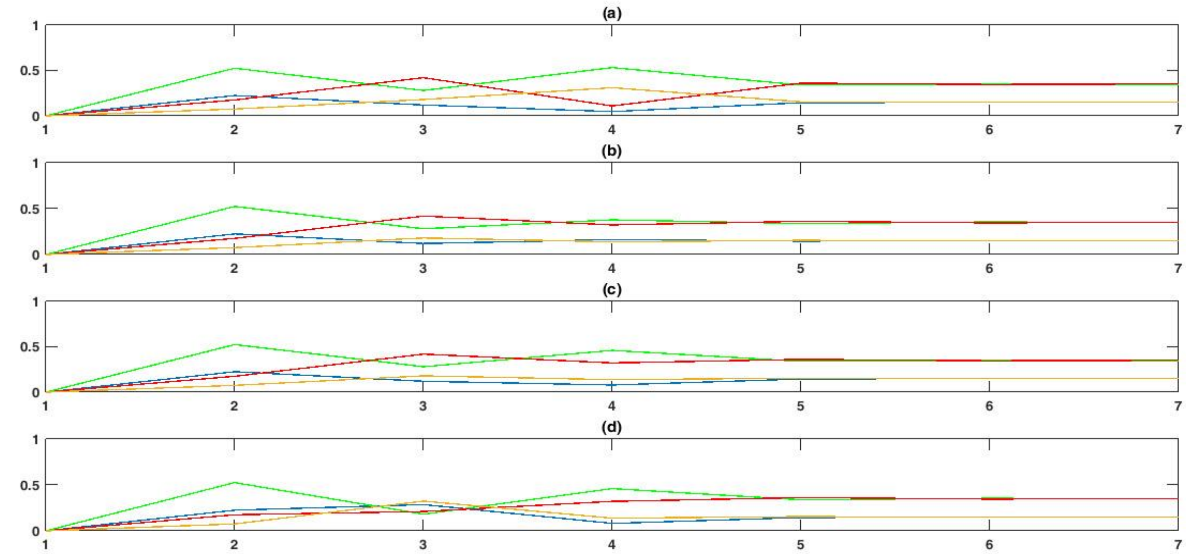

Fig. 5. (a). MS-SEMI-K(1)-G(1), (b). MS-SEMI-K(1)-G(2), (c). MS-SEMI-K(1)-G(3), (d). MS-SEMI-K(1)-G(4), the colors of green, brown, red and blue are the joint conditional probability of ($Q_t = 1, Q_{t+1} = 2$), ($Q_t = 1, Q_{t+1} = 1$), ($Q_t = 2, Q_{t+1} = 1$) and ($Q_t = 2, Q_{t+1} = 2$), respectively.

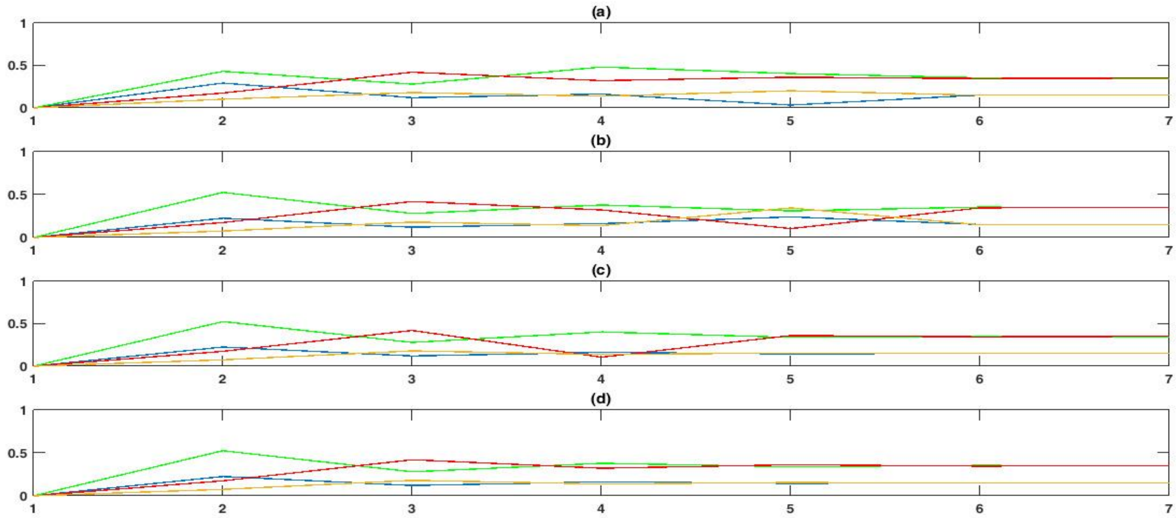

Fig. 6. (a). MS-SEMI-K(2)-G(1), (b). MS-SEMI-K(2)-G(2), (c). MS-SEMI-K(2)-G(3), (d). MS-SEMI-K(2)-G(4), the colors of green, brown, red and blue are the joint conditional probability of ($Q_t = 1, Q_{t+1} = 2$), ($Q_t = 1, Q_{t+1} = 1$), ($Q_t = 2, Q_{t+1} = 1$) and ($Q_t = 2, Q_{t+1} = 2$), respectively.

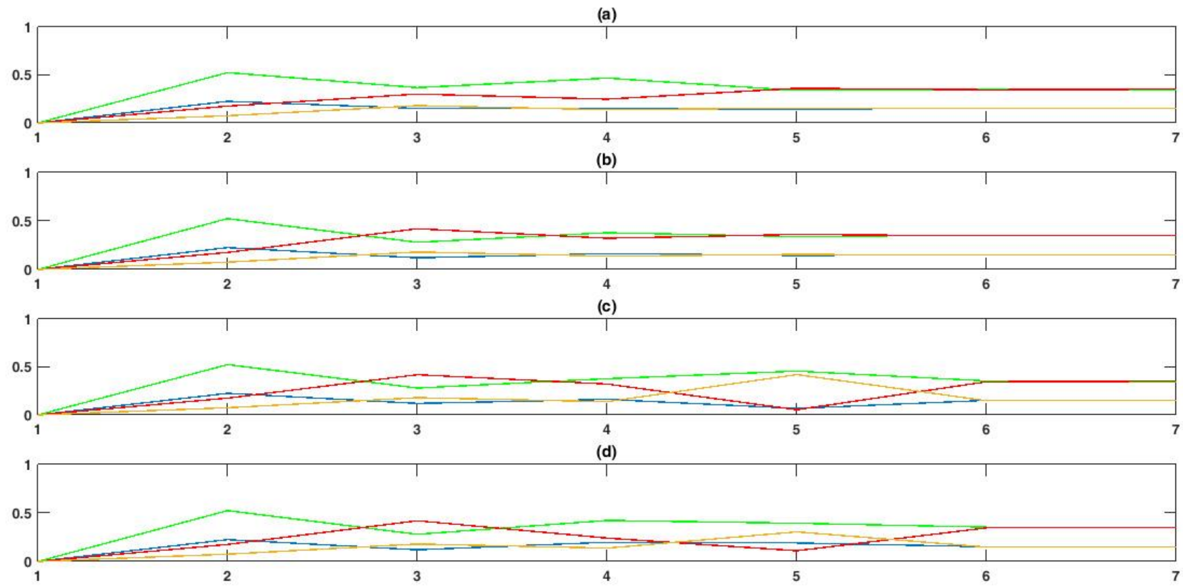

Fig. 7. (a). MS-SEMI-K(3)-G(1), (b). MS-SEMI-K(3)-G(2), (c). MS-SEMI-K(3)-G(3), (d). MS-SEMI-K(3)-G(4), the colors of green, brown, red and blue are the joint conditional probability of ($Q_t = 1, Q_{t+1} = 2$), ($Q_t = 1, Q_{t+1} = 1$), ($Q_t = 2, Q_{t+1} = 1$) and ($Q_t = 2, Q_{t+1} = 2$), respectively.

Table 5 shows the estimated joint conditional probability matrix for the observations of the Banks index (industry group) for the selected time period based on the twelve models.
According to the result of this Table for the best selected model (MS-SEMI-K(3)-G(1)), we can see the most probabilities among the elements of the matrices $\delta_{ij}^{t,t+1}$for period (t=243, t+1=244), (t=244, t+1=245) and (t=245, t+1=246) are 0.5247, 0.3680 and 0.4670, respectively, that are belong to the switching between the regimes from 1 to 2 ($Q_t = 1, Q_{t+1} = 2$). This offers that the strategy of buying

the stocks in period of time t=243 to t= 245. In the other hand, the results indicate that the most probabilities among the element of the matrices $\delta_{ij}^{t,t+1}$for period (t=246, t+1=247) is 0.3610. This shows the switching between the regimes is from 2 to 1 ($Q_t = 2, Q_{t+1} = 1$) that offers the strategy of selling the stocks in period of time t=246. On the other hand, if we select the second proper model (MS-SEMI-K(1)-G(4)), the most probabilities among the elements of the matrices $\delta_{ij}^{t,t+1}$for period (t=243, t+1=244) and (t=245, t+1=246) are 0.5256 and 0.4602, respectively, that are belong to the switching between the regimes from 1 to 2 ($Q_t = 1, Q_{t+1} = 2$). This offers that the strategy of buying the stocks in period of time t=243 and t= 245. But, the most probabilities among the elements of the matrices $\delta_{ij}^{t,t+1}$for period (t=244, t+1=245) is 0.3234 that is belong to the switching between the regimes from 2 to 2 ($Q_t = 2, Q_{t+1} = 2$) that show the trend of the data will stay in increasing trend. This state indicate that the strategy of buying for traders in time of t=244. Although, the probability of the switching between the regimes from 1 to 2 ($Q_t = 1, Q_{t+1} = 2$) (0.1810) is weak, comparing the other probabilities.

Table 5. The Estimated Joint Conditional Probability for the Models

The model	Time period			
	t=243, t+1=244	t=244, t+1=245	t=245, t+1=246	t=246, t+1=247
MS-SEMI-K(1)-G(1)	$\begin{pmatrix}0.2243 & 0.5252\\0.1751 & 0.0753\end{pmatrix}$	$\begin{pmatrix}0.1198 & 0.2797\\0.4203 & 0.1802\end{pmatrix}$	$\begin{pmatrix}0.0475 & 0.5321\\0.1102 & 0.3102\end{pmatrix}$	$\begin{pmatrix}0.1451 & 0.3395\\0.3605 & 0.1550\end{pmatrix}$
MS-SEMI-K(1)-G(2)	$\begin{pmatrix}0.2249 & 0.5243\\0.1757 & 0.0752\end{pmatrix}$	$\begin{pmatrix}0.1200 & 0.2806\\0.4193 & 0.1801\end{pmatrix}$	$\begin{pmatrix}0.1619 & 0.3775\\0.3225 & 0.1381\end{pmatrix}$	$\begin{pmatrix}0.1457 & 0.3387\\0.3613 & 0.1543\end{pmatrix}$
MS-SEMI-K(1)-G(3)	$\begin{pmatrix}0.2253 & 0.5242\\0.1755 & 0.0750\end{pmatrix}$	$\begin{pmatrix}0.1201 & 0.2807\\0.4193 & 0.1800\end{pmatrix}$	$\begin{pmatrix}0.0792 & 0.4602\\0.3225 & 0.1382\end{pmatrix}$	$\begin{pmatrix}0.1452 & 0.3391\\0.3609 & 0.1548\end{pmatrix}$
MS-SEMI-K(1)-G(4)	$\begin{pmatrix}0.2247 & 0.5256\\0.1746 & 0.0750\end{pmatrix}$	$\begin{pmatrix}0.2854 & 0.1810\\0.2102 & 0.3234\end{pmatrix}$	$\begin{pmatrix}0.0810 & 0.4602\\0.3209 & 0.1379\end{pmatrix}$	$\begin{pmatrix}0.1449 & 0.3381\\0.3619 & 0.1551\end{pmatrix}$
MS-SEMI-K(2)-G(1)	$\begin{pmatrix}0.2923 & 0.4310\\0.1747 & 0.1020\end{pmatrix}$	$\begin{pmatrix}0.1201 & 0.2796\\0.4205 & 0.1798\end{pmatrix}$	$\begin{pmatrix}0.1598 & 0.4811\\0.3214 & 0.0377\end{pmatrix}$	$\begin{pmatrix}0.0331 & 0.4041\\0.3613 & 0.2015\end{pmatrix}$
MS-SEMI-K(2)-G(2)	$\begin{pmatrix}0.2010 & 0.4903\\0.1416 & 0.1671\end{pmatrix}$	$\begin{pmatrix}0.1246 & 0.2591\\0.4128 & 0.2035\end{pmatrix}$	$\begin{pmatrix}0.1633 & 0.3785\\0.3215 & 0.1367\end{pmatrix}$	$\begin{pmatrix}0.2402 & 0.3105\\0.1042 & 0.3451\end{pmatrix}$
MS-SEMI-K(2)-G(3)	$\begin{pmatrix}0.2250 & 0.5254\\0.1747 & 0.0749\end{pmatrix}$	$\begin{pmatrix}0.1224 & 0.2910\\0.4175 & 0.1691\end{pmatrix}$	$\begin{pmatrix}0.1548 & 0.4030\\0.1054 & 0.3368\end{pmatrix}$	$\begin{pmatrix}0.1449 & 0.3387\\0.3613 & 0.1551\end{pmatrix}$
MS-SEMI-K(2)-G(4)	$\begin{pmatrix}0.2437 & 0.5349\\0.2134 & 0.0080\end{pmatrix}$	$\begin{pmatrix}0.1181 & 0.2471\\0.5118 & 0.1230\end{pmatrix}$	$\begin{pmatrix}0.1502 & 0.3584\\0.3413 & 0.1501\end{pmatrix}$	$\begin{pmatrix}0.1449 & 0.3387\\0.3613 & 0.1551\end{pmatrix}$
MS-SEMI-K(3)-G(1)	$\begin{pmatrix}0.2253 & 0.5247\\0.1751 & 0.0749\end{pmatrix}$	$\begin{pmatrix}0.1512 & 0.3680\\0.3010 & 0.1798\end{pmatrix}$	$\begin{pmatrix}0.1471 & 0.4670\\0.2480 & 0.1379\end{pmatrix}$	$\begin{pmatrix}0.1451 & 0.3390\\0.3610 & 0.1549\end{pmatrix}$
MS-SEMI-K(3)-G(2)	$\begin{pmatrix}0.2317 & 0.4902\\0.1564 & 0.1217\end{pmatrix}$	$\begin{pmatrix}0.1202 & 0.2802\\0.4198 & 0.1798\end{pmatrix}$	$\begin{pmatrix}0.1531 & 0.33694\\0.3108 & 0.1667\end{pmatrix}$	$\begin{pmatrix}0.1451 & 0.3390\\0.3610 & 0.1549\end{pmatrix}$
MS-SEMI-K(3)-G(3)	$\begin{pmatrix}0.2038 & 0.5011\\0.1698 & 0.1253\end{pmatrix}$	$\begin{pmatrix}0.1227 & 0.2712\\0.4037 & 0.2024\end{pmatrix}$	$\begin{pmatrix}0.1584 & 0.3698\\0.3157 & 0.1561\end{pmatrix}$	$\begin{pmatrix}0.0679 & 0.4570\\0.0541 & 0.4210\end{pmatrix}$
MS-SEMI-K(3)-G(4)	$\begin{pmatrix}0.2102 & 0.5063\\0.1502 & 0.1333\end{pmatrix}$	$\begin{pmatrix}0.1215 & 0.2641\\0.4005 & 0.2139\end{pmatrix}$	$\begin{pmatrix}0.1981 & 0.4230\\0.2410 & 0.1379\end{pmatrix}$	$\begin{pmatrix}0.1906 & 0.3940\\0.1104 & 0.3050\end{pmatrix}$

References

[1] Aghaeefar, N., Pourzarandi, M., Kazemi, M., and Minoie, M., Applying Optimized Mathematical Algorithms to Forecast Stock Price Average Accredited Banks in Tehran Stock Exchange and Iran Fara Bourse, *Advances in Mathematical Finance and Applications*, 2019:77-94. doi: 10.22034/amfa.2019.580802.1150

[2] Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H. K., Interconnections between Eurozone and US booms and busts using a Bayesian panel Markov switching VAR model, *Journal of Applied Econometrics*, 2016. doi:10.1002/jae.2501

[3] Chan, K. C. G., and Wang, M. C., Semiparametric modeling and estimation of the terminal behavior of recurrent marker processes before failure events, *Journal of the American Statistical Association*, 2017;112(517): 351-362. DOI: 10.1080/01621459.2016.1140051

[4] Chang, Y., Yongok, C., and Park, J.Y., A New Approach to Model Regime Switching, *Journal of Econometrics*, 2017;196(1): 127-143. doi:10.1016/j.jeconom.2016.09.005

[5] Chang, J., Tang, C. Y. and Wu, Y., Local Independence Feature Screening for Nonparametric and Semiparametric Models by Marginal Empirical Likelihood, *Annals of statistics*, 2016;**44**(2):515-539. doi: 10.1214/15-AOS1374

[6] Chen, C. M., Shen, P. S., Wei, J. C., and Lin, L., A semiparametric mixture cure survival model for left-truncated and right-censored data, *Biometrical journal*, 2017; **59**(2): 270-290. doi:10.1002/bimj.201500267

[7] Davallou, M., and Meskinimood, S., *Performance examination of trading strategy based on Stochastic Dominance*, Financial Knowledge of security analysis (Financial studies), 2019; 12:171-193.

[8] Davoodi, A., and Dadashi, I., Stock price prediction using the Chaid rule-based algorithm and particle swarm optimization (pso), *Advances in Mathematical Finance and Applications*, 2020;197-213. doi:10.22034/amfa.2019.585043.1184

[9] Di Persio, L., Frigo, M., Gibbs sampling approach to regime switching analysis of financial time series, *Journal of Computational and Applied Mathematics*, 2016. doi:10.1016/j.cam.2015.12.010

[10] Di Persio, L., Frigo, M., Maximum Likelihood Approach to Markov Switching Models, *WSEAS Transactions on Business and Economics*, 2015; 12: 239-242.

[11] Doaei, M., Mirzaei, S,A., and Rafigh, M., Hybrid Multilayer Perceptron Neural Network with Grey Wolf Optimization for Predicting Stock Market, *Advances in Mathematical Finance and applications,* 2021; 883-894. doi: 10.22034/amfa.2021.1903474.1452

[12] Franke, J., Stockis, J.P., Kamgaing, J.T., Li, W.K., Mixtures of nonparametric autoregressions, *J. Nonparametric statistics*, 2011; 23: 287-303. doi: 10.1080/10485252.2010.539686

[13] Gu, X., Ma, Y., and Balasubramanian, R., Semiparametric Time to Event Models in the Presence of Error-Prone, Self-Reported Outcomes-with Application to the Women's Health Initiative, *the annals of applied statistics*, 2015; **9**(2): 714-730. doi: 10.1214/15-AOAS810

[14] Gupta, C., Cobre, J., Polpo, A., and Sinha, D., Semiparametric Bayesian estimation of quantile function for breast cancer survival data with cured fraction, *Biometrical journal. Biometrische Zeitschrift,* 2016; **58**(5): 1164-77. doi:10.1002/bimj.201500111

[15] Nademi, A., and Farnoosh, R., Mixtures of autoregressive-autoregressive conditionally heteroscedastic models: semi-parametric approach, *Journal of Applied Statistics,* 2014, **41**(2), 275-293. doi:10.1080/02664763.2013.839129

[16] Nademi, A. and Nademi, Y., Forecasting crude oil prices by a semiparametric Markov switching model: OPEC, WTI, and Brent cases, *Energy Economics,* 2019; 74: 757-766. doi:10.1016/j.eneco.2018.06.020

[17] Neale, M. C., Clark, S. L., Dolan, C. V., Hunter, M. D., Regime Switching Modeling of Substance Use: Time-Varying and Second-Order Markov Models and Individual Probability Plots, *Structural equation modeling: a multidisciplinary journal*, 2016; **23**(2):221-233. doi: 10.1080/10705511.2014.979932

[18] Pashaei, Z., and Dehkharghani, R., Stock Market Modeling Using Artificial Neural Network and Comparison with Classical Linear Models, *Signal and Data Processing*, 2021, 17. doi:10.29252/jsdp.17.4.89

[19] Pourzamani, Z., Heidarpour, F., and Mohammadi, M., Comparative study of stock trading strategy in the long term investment, *Financial engineering and securities management*, 2011:7: 215-230.

[20] Rastegar, M. A., Dastpak, M., Developing a High-Frequency Trading system with Dynamic Portfolio Management using Reinforcement Learning in Iran Stock Market, *Financial research journal*, 2018; 20. doi: 10.22059/jfr.2017.230613.1006415

[21] Sharif-far, A., Araghi, M.K., Vanani, I.R., Fallahshams, M., The Assessment of the optimal Deep Learning Algorithm on Stock Price Prediction (Long Short-Term Memory Approach), *Quarterly Journal of Financial Engineering and Securities Management*,2021.

[22] Von Ganske, J., A Regime Switching Partial Least Squares Approach to Forecasting Industry Stock Returns, *Mimeo Edhec-Risk Institute*, 2016.