Advances in Industrial Engineering, Summer 2021, 55(3): 233-247
DOI: 10.22059/aie.2021.328205.1795
RESEARCH PAPER

Stock Return Forecasting Using the Bayesian Model Approach in Tehran Securities Exchange

Ladan Mohebian, Alireza Heidarzadeh Hanzaee *

Department of Financial Management, Tehran North Branch, Islamic Azad University, Tehran, Iran.

Received: 03 August 2018, Revised: 19 October 2021, Accepted: 25 October 2021

Abstract

In the present study, stock returns were predicted using the Bayesian model approach in the Tehran Stock Exchange. Therefore, the research hypothesis based on the Bayesian method has higher accuracy in predicting returns than Autoregressive models was developed and tested. In order to test the research hypothesis, information related to the index of 30 selected industries in the Tehran Stock Exchange during the period from 2017-03-25 to 2020-08-24 was used. The index return was predicted based on two methods for 30 out-of-sample data. First, autoregressive models were fitted on the returns of each index and then the next 30 days of returns were predicted based on these models. Then, after identifying the optimal model lags through the Bayesian model averaging method, autoregressive models were fitted with the optimal lags and the next 30 days predictions were obtained under this method. In order to compare the accuracy of the methods in predicting the return, RMSE and MAE criteria were used and the values of these error criteria were compared using Wilcoxon nonparametric pairwise comparison tests. The results showed that the Bayesian method leads to an increase in the accuracy of model prediction in out of sample data.

Keywords:
Return Prediction;
Autoregressive Model;
Bayesian Model

Introduction

In recent years the stock market has become an integral part of the world's economy. Any fluctuations can impact each country's economy and companies' financial budget [1] . The stock exchange has always been at the center of attention for investment due to its high return. Meanwhile, some influential factors cause unpredictable behavior and prices in the stock market [2] . The trend is defined as a significant price change during a specified period of time. Trends are classified as ascending, descending and neutral. So far, a variety of models with different specifications have been applied to predict the market trend. Since an appropriate prediction will result in noticeable profits so the presentation of a model for accurate prediction has its own importance [3] .

Although the most tendency of prediction of the stock price has focused on time series and classic models [4] , they are only able to do prediction with acceptable approximation through spaces with limited changes, *i.e.*, they fail to predict environmental changes with accurate approximation because of variable conditions in the stock market [5] . Therefore, it is essential to look for modern models and intelligent systems for prediction. In this regard, methods include neural networks, phase of logic, genetic algorithm and postmodern algorithm have been

* Corresponding author: (A. Heidarzadeh Hanzaee)
Email: a_heidarzadeh@iau-tnb.ac.ir

developed for finding solutions in different scientific fields [6] . Among the above-mentioned methods, the nervous system technique has been at center of attention due to its non-linear identity is capable to face rough data, process quickly and access intelligent predicting models for modeling and prediction [7] . Financial reports assume that to make future predictions, investors pay much attention to a set of variables that enable them to predict future return, while there is a great number of variables and uncertainty about those variables that define dominant accurate model on proficiency and price is high [8] .

Recent studies have demonstrated different methods exist in monetary markets to decrease uncertainty levels in investment, however no one is known as pretty accurate and perfect [9] . So, it is necessary for investment managers, petite investors and analysts to provide appropriate and accurate predicting methods with minimum errors in future share cost. Concerning the numbers of predicting models of share price and return selection and identification of optimum model is known frustrating and non-experimental. In this regard, Bayesian model was selected as a competitive method to identify the optimal model among lots of available models [10] . Although there have been methods focusing on identifying optimal variables by means of classic statistics, the Bayesian approach, considering the stochastic nature of the model parameters, leads to the identification of the model with the most fit to the data. Therefore, in this study, index returns are predicted using the Bayesian approach and compared with the usual method of classical time series models to evaluate the advantage of this method in selecting parameters and predictor variables as well as reducing the amount of forecast error. Although different methods of Bayesian estimation have been studied in forecasting, the Bayesian model averaging (BMA) has not been studied so far and there is no evidence of how this method performs in predicting future returns, compared to classical statistical methods.

Literature review and hypothesis development

The prediction of stock returns means a process that determines the future stock value or other purchased monetary tools [11] . Successful prediction of stock prices can provide magnificent profit. The efficient market hypothesis (EMH) indicates that stock prices reflect all available information in the market and any price changes formed according to unrevealed recent information will not be basically predictable [12] . Some researchers have opposite ideas and in their point of view, there are large numbers of methods that enable them to access price information in the future [13]. EMH indicates that the stock price is a function of information and logical expectation and recently revealed information about the company’s outlook roughly are reflected in the current stock price [14] . It means that the whole known general information about a company which includes its price history is hidden in the current stock price [15] . According to this fact, the price changes reflect new information, market changes, or random movement around available information [16] . Malkiel [17] claimed that it is impossible to predict stock prices accurately within price history. As a result, he claimed that stock prices should be predicted by a statistical process, which means the daily price variation is known as a random quantity and is unpredictable [18] .

The EMH received special attention; however, critics point out some cases in which the real market experience is quite different from unpredictability [19] . Several arguments have been raised around this category that says some analysts can predict stocks better [20] . Methods applied by the second group of researchers to predict stock returns can be classified into several main categories which are detailed as follows.

Fundamental analysis

Fundamental analysts evaluate a company's past performance as well as the credibility of its accounts. Many performance ratios are created in this regard that help the fundamental analyst evaluate stock credit, such as the price-to-earnings ratio (P/E). What fundamental stock market analysis seeks to achieve, is to discover the true value of stocks, which can be compared to the value traded in stock markets, and thus to assess whether stocks in the market are undervalued [21] . Real value can be realized in different ways and basically the same principle. The principle is that a company is worth collecting all its future profits [22] . These future profits must also be discounted to the current value. This principle is well in line with the theory that trade is entirely about profit and nothing else. Unlike technical analysis, fundamental analysis is more of a long-term strategy [23] .

Technical analysis

Technical analysts never pay attention to the company’s principles. They are looking for setting prices according to past price trends (a type of time series analysis). Many Patterns such as head and shoulder or cup and saucer as well as techniques like exponential moving average (EMA), oscillators, support and resistant level, or motion and volume indices are being used [24] . Technical analysis is more used than long term strategies for short term strategies. So, for goods and in bilateral markets that business people concentrate on short term price movement, technical analysis is more common [25] . Some of the basic assumptions used in this analysis are that everything significant about a company is already priced in its stock, another is that the price is moving in the process, and finally that the date (prices) tend to be repeated [26] .

Machine learning

Thanks to computers and the expansion of the digital era, stock market prediction was transferred to the field of technology. The most prominent methods include artificial neural networks (ANN) and genetics algorithms (GA). The artificial neural network is considered as an approximate mathematic function [27] . The commonest form of ANN used to predict the stock market is the feed-forward network which can update the network by error algorithm. To predict price with ANN, there are usually two methods to predict time horizons: independent and joint [28] .

The independent approach uses a single ANN for each horizon for instance 1 day, 2 days, or 5 days. Its upside is that the errors of network prediction for each horizon have no effects on the other because every time horizon is unique as usual. In this method, the error of a time horizon may divide its error with another horizon which can weaken the application of the method for prediction [29] .

Time series

One of the general methods for the prediction of stock returns is the use of classic predicting statistical methods and time series models. In these models, the values of return or future price are modeled through its past values, and in particular, this prediction can be made through the fundamental variables of the company [30] . One of the advantages of this method over machine-based methods is the presentation of an analytical model and accurate specification of the governing model, but experimental results show that the accuracy of this method is lower compared to algorithmic and machine methods [31] . Among the consecutive time series models in modeling and forecasting returns, we can mention ARMA, GARCH, fractional models, EGARCH, TGARCH, etc. [32] .

Hypothesis development

Numerous studies have been performed on the predictive power of regression and non-regression models of future values of variables. For example, Chen [32] started to check 24 specifications of a company entitled the prediction of share return based on The Bayesian model. He claimed that the group of effective specifications estimated by the Bayesian model can be added to the period before and after 2003. The results come from his research show high returns in internal samples as well as external ones.

Gao [33] carried out a research entitled selecting share management for investment via the Bayesian model to improve investment management for investors. He first extracted monetary indexes of shares located in China then selected those indexes which highly affected share growth based on The Bayesian model. Afterward, he classified them to high proficiency shares and other ones. The results have proved those shares classified as high proficiency by The Bayesian model have had more growth that means his method was efficient.

Atkins et al. [34] studied the fluctuation of the stock market via a comparison of GARCH and implicit and achievable fluctuation models. The results indicated the implicit fluctuation follow a predictable pattern and verify a connection between implicit fluctuation and share index return. Implicit fluctuation has a worse function than periodical prediction. Meanwhile, a model which combines the asymmetric GARCH model with implicit fluctuation through ARMA models is known as an ideal model.

Wang et al. [11] investigated share bargaining by using the PE ratio. They used the Bayesian network modeling to cope with the monetary behavior and basic investment. They concluded estimated PE ratio can be used as formative protection for experts' decisions to set up an automatic share bargaining system.

Hamidian et al. [35] carried out a research contributing to the prediction of the share negative return of accepted companies in Iran's capital market. So as to achieve the goal, the data deal with 180 companies was obtained via the systematic omitting system, as well as criteria such as leverage, function, turning, oscillation, quality and torpedo were used to predict negative proficiency. The result has indicated that all the above criteria are known as effective ones in order to predict negative proficiency [28] by using inquire data method established a research to predict share index. For stock market prediction, 16 financial statements and share index as labels divided by 4 gaps were studied. Later, in order to predict various types of classification including decision tree neural system and compound method (random jungle, simple Biz and tree decision. Dealing with the outcomes and the effectiveness of each method, it became obvious that the compound method and tree decision which have an 87.06 rate of function were known as the best way for prediction.

Alamatian and vafaie Jahan [3] via the prediction of share price in Iran's stock market based on the Bayesian Network and Markov hidden market claimed that share behavior and its changing trends are the most complicated mechanism that has always been noticed. The stock market is influenced by both internal and external factors, the effective external factors such as political and social issues are immeasurable, therefore, to predict stock market trends we need to focus on internal effective factors.

This research suggests a compound system in accordance with The Bayesian Network and hidden Markov model to predict the daily stock market trends. Fakhari et al. [7] through a research considered the function of The Bayesian neural network and Lunberg Markovat in comparison with classic models to predict the share price of investing companies. To design a predicting model with a neural system, daily market price and technical monetary index as input variables were used. Moreover, to design the ARIMA model final daily data price as an input variable and also final new day data price as output variable was considered.

The results from the Bayesian neural network indicated fewer errors and more powerful predictions than the ARIMA model. The outcomes of research showed the Bayesian neural network has more efficiency to use market investment opportunities which can help investors to choose suitable portfolios and gain more proficiency.

Barzegari Khaneghah et al. [36] predicted share return through financial ratio. The outcome clarified that beneficial ratios play a more important role in share proficiency than other financial ratios. Meanwhile, assets return ratio and owner's rights have the most abilities to explain share return.

Salehirad and Habibifard [37] in order to study and examine time-series data contribute to share index in the monetary market compare maximum verification method with The Bayesian, finally by selecting the best model, they gained prediction made by the Bayesian model for the future stock price.

A review of previous studies shows that the methods based on classical statistical patterns, although having an explicit form, show less accuracy in prediction. Neural network-based methods and Bayesian-based methods will generally provide higher accuracy in forecasting. Hence, the research hypothesis is defined in this way:

H_0: Bayesian model averaging is more accurate than autoregressive models in predicting index return.

H_1: Bayesian model averaging is not more accurate than autoregressive models in predicting index return.

Methodology

The statistical population of the research includes 30 selected industries indexes in Tehran Stock Exchange and daily index values are used from 2017.03.25 to 2020.08.24. Studied industries are as follows:

Insurance and retirement, cement, investments, chemicals, drugs, tile and ceramic, sugar, non-metal minerals, automobiles, media electric gadgets, machineries, metal products, main metals, rubber, oil products, publications, paper, computer, leather products, banks, metals, coal, agriculture, engineering, non-sugary foods, oil extraction without excavation, retail, communication and textiles.

Index return is known as a variable in this research which is calculated by the logarithm of current index value divided by the previous value in the last period:

$$R_{\mathrm{t}} = \mathrm{Ln}\left(\frac{\mathrm{S_t}}{\mathrm{S_{t_0}}}\right) \tag{1}$$

In this relationship, R_{t} represents the $t - t_0$days return, $\mathrm{S_t}$ the value of index at the end of day t and $\mathrm{S_{t_0}}$ the value of index at the end of day t_0.

In order to examine the research hypothesis, index data are collected over T days and a single day return is estimated for them. Then, the following procedure are done to test research hypothesis. We used to methods of autoregressive and Bayesian models to predict returns as follows:

Autoregressive model

Autoregressive model for index return is calculated using Eq. 2:

$$R_t = \alpha + \sum_{i=1}^{k} \beta_i R_{t-i} + \varepsilon_t \qquad (2)$$

Where, K is the length of lags which is calculated by autocorrelation function of returns and Akaike and Schwarz criterion. So, we have: $\varepsilon_t \sim N(0, \sigma^2)$

Following the estimation of regression coefficients β_i, the predicted returns are obtained by autoregressive model during [t, T]. To examine the accuracy, RMSE and MAE are used. Furthermore, to check the acceptable accuracy of the prediction, mean comparing test of residuals are used during [t, T].

Bayesian model averaging

The Bayesian method in this research is Bayesian model averaging. In order to compare results with the autoregressive method, return lags are considered as independent variables in this method, unlike Sousa and Sousa [8] in which major economic indexes are considered as independent variables. In Bayesian model averaging, different combination of independent variables is considered, and models are considered as selected models for prediction.

For each model, one probability is defined based on prior information that can be the probable foundation for model accuracy. In other words, if K lags of return values are considered as independent variables, it will be 2^k number of suggested models that predict returns. The correctness probability of each individual 2^k model is equal to P_i (i=1, 2, ..., k). If there is no previous information about the priority of a model in comparison with others, each probability P_i equals to 1 divided by 2^k.

If the priority of models relatively seems clear, the value of P_i is different from each other and a model that has more fitness with data has higher P_i with this condition that the sum of all P_is equal to 1. In this research, all models have equal fitness and the value of P_is for each i is the same. After coefficient estimation, the value of the Schwarz criterion is calculated and a model with lower Schwarz fitness is introduced as an optimal model.

After selection of the final model, RMSE and MAE criteria are calculated in order to measure the accuracy of predicting. Finally, the accuracy of the model is tested by comparing the mean test.

In order to find out the prediction precision of two methods, RMSE and MAE criteria were compared by two models as well as a mean comparison for error prediction was done by two models. Data analysis has been checked by descriptive and inferential statistics. The stationarity of data was checked by the Dickey-Fuller test and the autoregressive model fitted to the data. For checking primary assumptions of model, the ARCH test and Breusch-Pagan test were applied on errors. SPSS software version 22nd was used to compare mean comparison, prior to this comparison, the normal assumption of error values was checked by Kolmogorov–Smirnov test. Furthermore, for Bayesian analysis, statistical R software packages were used.

Results

Table 1 shows a descriptive statistic of index daily returns by mean, median, std., max and min values. The average of daily returns for selected industries in this table indicates that the index of the Media industry with the daily average of 0.0132 obtains the highest average among others and the minimum average return belongs to the Computer industry which equals to 0.0040. Meanwhile, the minimum and maximum values of return indexes were calculated by logarithm of two successive days for index values which do not comply with the +5 and -5 rule. Based on standard deviations, the Publication industry has the highest return fluctuation which equals to 0.4781 and the lowest one relating to the Investment industry which equals to 0.0187.

Table 1. Descriptive statistics of returns

Industry	Code	Mean	Median	Max	Min	Std.
Insurance and Retirement	Ind1	0.0076	0.0088	0.0476	-0.0452	0.0230
Cement	Ind2	0.0074	0.0093	0.0535	-0.0400	0.0210
Investments	Ind3	0.0077	0.0086	0.0933	-0.0430	0.0187
Chemicals	Ind4	0.0060	0.0051	0.0793	-0.0581	0.0204
Drugs	Ind5	0.0065	0.0069	0.0927	-0.0435	0.0232
Tiles and Ceramics	Ind6	0.0080	0.0107	0.1948	-0.0464	0.0247
sugar	Ind7	0.0066	0.0072	0.1251	-0.0426	0.0213
Nonmetal Minerals	Ind8	0.0071	0.0089	0.0730	-0.0501	0.0207
Automobiles	Ind9	0.0089	0.0097	0.1134	-0.0485	0.0266
Media	Ind10	0.0132	0.0184	0.1328	-0.0717	0.0334
Electrical Gadgets	Ind11	0.0076	0.0091	2.3539	-2.3364	0.3639
Machineries	Ind12	0.0073	0.0087	0.0510	-0.0461	0.0197
Metal Products	Ind13	0.0072	0.0101	0.1067	-0.0502	0.0259
Main Metals	Ind14	0.0064	0.0045	2.2662	-2.2647	0.1796
Rubber	Ind15	0.0068	0.0094	0.0765	-0.0498	0.0259
Oil Products	Ind16	0.0065	0.0062	2.3676	-2.3440	0.4428
Publications	Ind17	0.0067	0.0084	2.3602	-2.3534	0.4781
Paper Products	Ind18	0.0062	0.0074	0.0827	-0.0516	0.0287
Computers	Ind19	0.0040	0.0040	0.0913	-0.0642	0.0220
leather Products	Ind20	0.0114	0.0025	0.3486	-0.0513	0.0347
Banks	Ind21	0.0071	0.0064	0.0983	-0.0486	0.0222
Metal Minerals	Ind22	0.0057	0.0031	0.0913	-0.0490	0.0230
Coal	Ind23	0.0067	0.0079	0.0883	-0.0513	0.0321
Agriculture	Ind24	0.0080	0.0104	0.0705	-0.0513	0.0276
Engineering	Ind25	0.0067	0.0056	0.0723	-0.0869	0.0318
Non-Sugary Foods	Ind26	0.0073	0.0097	0.0650	-0.0440	0.0196
Oil Extraction Except Excavation	Ind27	0.0061	0.0040	0.1241	-0.1355	0.0337
Retail Except Transportation	Ind28	0.0060	0.0019	0.1578	-0.0512	0.0269
Information and Communication	Ind29	0.0055	0.0000	0.1549	-0.0513	0.0318
Textiles	Ind30	0.0060	0.0059	0.0768	-0.1156	0.0234

Table 2. The stationarity ADF test

Industry	ADF stat.	sig.	Industry	ADF stat.	sig.
Ind1	-11.61342	0.000	Ind16	-9.014740	0.000
Ind2	-10.89691	0.000	Ind17	-18.90906	0.000
Ind3	-12.36596	0.000	Ind18	-13.42475	0.000
Ind4	-12.42275	0.000	Ind19	-13.04046	0.000
Ind5	-11.01548	0.000	Ind20	-11.87558	0.000
Ind6	-11.81756	0.000	Ind21	-6.927888	0.000
Ind7	-11.36234	0.000	Ind22	-13.82615	0.000
Ind8	-11.32862	0.000	Ind23	-13.09668	0.000
Ind9	-6.541667	0.000	Ind24	-12.20042	0.000
Ind10	-11.62215	0.000	Ind25	-13.23409	0.000
Ind11	-17.11835	0.000	Ind26	-13.11152	0.000
Ind12	-11.65160	0.000	Ind27	-15.13329	0.000
Ind13	-12.29263	0.000	Ind28	-13.73915	0.000
Ind14	-12.19422	0.000	Ind29	-13.32717	0.000
Ind15	-12.07828	0.000	Ind30	-14.00574	0.000

Also, the Dickey-Fuller test was used to examine stationarity of daily index returns. The results are demonstrated in Table 2. According to Table 2, we can conclude that the returns are stationary and it is possible for modeling of index returns with no needs of using Box-Cox transformations.

In order to fit autoregressive models, autocorrelation and partial autocorrelation functions are used as the main methods to diagnose lags. They would not necessarily be able to diagnose optimal lags themselves. So, in this research optimal lags were diagnosed in accordance with some lags which lead to the least AIC. Table 3 indicates the results of fitness for each selected industry. In this table, β_i represents the estimated coefficient for the ith lag of the autoregressive model. According to Table 3 autoregressive model for each selected industry was fit with different lags. The number of lags relies on minimizing AIC. It is clear that the analysis of results for each model individually is not important and only the prediction accuracy of models is highly considered to check the precision of predictions.

Table 3: Autoregressive models fitted to index returns

Industry	α	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Ind1	0.0088**	0.387**							
Ind2	0.0085**	0.471**	-0.087	0.136*	-0.066	-0.068	0.150**		
Ind3	0.0086**	0.340**	-0.040	0.182**					
Ind4	0.0071**	0.345**							
Ind5	0.0082**	0.433**							
Ind6	0.0087**	0.448**	-0.160**	0.248**	-0.139*				
Ind7	0.0072**	0.463**	-0.112						
Ind8	0.0087**	0.456**	-0.171**	0.157*	-0.131*				
Ind9	0.0107**	0.307**	0.023	0.201**					
Ind10	0.0125**	0.454**	-0.112*						
Ind11	0.0088	-0.374**	-0.116*						
Ind12	0.0081**	0.386**							
Ind13	0.0086**	0.325**							
Ind14	0.0075**	0.350**	-0.038	0.222**					
Ind15	0.0082**	0.356**							
Ind16	0.0071	-0.417**	-0.342**	-0.338**	-0.182**	-0.060	0.189**	0.017	-0.170**
Ind17	0.0082	-0.352**	-0.262**						
Ind18	0.0082**	0.274**							
Ind19	0.0056**	0.263**							
Ind20	0.0111**	0.380**							
Ind21	0.0078**	0.428**	-0.136*	0.294**	-0.129*				
Ind22	0.0066**	0.287**	-0.040	0.179**					
Ind23	0.0067**	0.331**	-0.103						
Ind24	0.0096**	0.334**							
Ind25	0.0084**	0.295**							
Ind26	0.0087**	0.280**							
Ind27	0.0063**	0.183**							
Ind28	0.0073**	0.313**	-0.125*						
Ind29	0.0054*	0.278**							
Ind30	0.0055**	0.279**	-0.094	0.128*					

Note: (*) significant at level 0.05, (**) significant at level 0.01

According to the results of Table 4, the significance level of the Ljung-Box test indicates that there exist no serial-correlation among error terms. In other words, the structure of index return determines explicitly by its previous lags and has no longer connection with other lags. The significance levels obtained by ARCH Test indicated heterogeneity of variance of error terms. Moreover, the significance level of the K-S test also showed the empirical distribution of residuals follows a normal distribution.

Based on the r-squared obtained for models it looks obvious fitting model for Oil Products (Ind16) industry with r-squared equal to 0.3199 has a significant role in expressing index return changes, meanwhile, the Oil Extraction without Excavation industry (Ind27) has the least R^2 which equals to 0.0333. AIC values indicated optimal Akaike for each model in accordance with identified lags.

Table 4. Goodness of fit indexes of autoregressive models

Industry	R^2	AIC	Ljung-Box	ARCH	K-S
Ind1	0.3341	0.3002	0.9208	-1472.99	0.1497
Ind2	0.4967	0.1545	0.9360	-1535.01	0.2303
Ind3	0.4616	0.9708	0.9983	-1596.67	0.1552
Ind4	0.2820	0.0761	0.9524	-1570	0.1203
Ind5	0.1722	0.4379	0.4603	-1476.74	0.1883
Ind6	0.5907	0.9999	0.8795	-1414.21	0.1979
Ind7	0.7377	1.0000	0.9688	-1506.94	0.1845
Ind8	0.9744	0.1428	0.9322	-1537.77	0.1848
Ind9	0.4108	0.0874	0.7722	-1396.99	0.1682
Ind10	0.1071	0.4001	0.8826	-1241.38	0.1786
Ind11	0.0523	0.0580	0.9252	147.83	0.1253
Ind12	0.4858	0.2347	0.8913	-1552.28	0.1492
Ind13	0.3536	0.1905	0.9751	-1382.64	0.1047
Ind14	0.4367	0.0737	0.9889	-1530.8	0.1845
Ind15	0.1760	0.0796	0.8531	-1393.76	0.1266
Ind16	0.0578	0.0532	0.8612	236.98	0.3199
Ind17	0.0575	0.0587	0.7897	343.93	0.1425
Ind18	0.3441	0.3710	0.6288	-1314.14	0.0751
Ind19	0.1888	0.0666	0.9118	-1494.56	0.0699
Ind20	0.0599	1.0000	0.5891	-1210.02	0.1441
Ind21	0.6968	0.9866	0.8542	-1492.94	0.2107
Ind22	0.2122	0.0611	0.9813	-1483.24	0.1159
Ind23	0.0673	0.4103	0.9910	-1241.15	0.0999
Ind24	0.1480	0.7379	0.9259	-1346.4	0.1114
Ind25	0.1036	0.0927	0.7783	-1255.12	0.0869
Ind26	0.5423	0.0574	0.7545	-1522.67	0.0782
Ind27	0.0865	0.9551	0.9633	-1187.32	0.0333
Ind28	0.0739	0.0944	0.8680	-1362.03	0.0812
Ind29	0.1055	0.9931	0.8001	-1247.4	0.0776
Ind30	0.0778	0.4182	0.9478	-1388.75	0.0848

The results of fitting autoregressive models on index returns are presented using the lags identified in the Bayesian averaging method. For this purpose, 30 proposed lags have been considered for each model and among the models with different combinations of autoregressive lags, 5 top models with optimal lags have been identified based on the Schwartz criterion and the optimal model among these 5 models is the basis of the analysis. The reason of using the Schwartz criterion in this method is its proximity to the Akaike criterion and matching the criteria for selecting optimal lags in both methods and, as a result, increasing the comparability of results between the two methods. Table 5 shows the results of fitting autoregressive models based on the proposed lags of the Bayesian averaging model. It should be noted that in this method, consecutive lags are not necessarily estimated and the detected lags are based only on the suitability of the model.

Table 5. Estimated Bayesian averaging models

Industry	Optimal lags	Intercept	1^{st} lag	2^{nd} lag	3^{rd} lag	4^{th} lag	5^{th} lag	6^{th} lag
Ind1	1	0.0054^{**}	0.4105^{**}					
Ind2	1	0.0046^{**}	0.4181^{**}					
Ind3	1,3	0.0045^{**}	0.3416^{**}	0.1521^{**}				
Ind4	1,7	0.0047^{**}	0.3434^{**}	0.0422				
Ind5	1	0.00452^{**}	0.4365^{**}					
Ind6	1	0.0055^{**}	0.3777^{**}					
Ind7	1	0.0042^{**}	0.3865^{**}					
Ind8	1	0.0054^{**}	0.4000^{**}					
Ind9	1,3	0.0054^{**}	0.2984^{**}	0.2156^{**}				
Ind10	1	0.0073^{**}	0.4269^{**}					
Ind11	1,10,12	0.0119	-0.3368^{**}	0.1621^{**}	-0.157^{**}			
Ind12	1	0.0046^{**}	0.4001^{**}					
Ind13	1,24	0.0073^{**}	0.3121^{**}	-0.1617^{**}				
Ind14	1,3	0.0037^{**}	0.3517^{**}	0.1908^{**}				
Ind15	1	0.0052^{**}	0.3525^{**}					
Ind16	1,2,3,4,6,8	0.0161	-0.3980^{**}	-0.3163^{**}	-0.172^{**}	0.2126^{**}	-0.156^{**}	-0.311^{**}
Ind17	1,2	0.0127	-0.3534^{**}	-0.2646^{**}				
Ind18	1	0.0059^{**}	0.2770^{**}					
Ind19	1	0.0044^{**}	0.2491^{**}					
Ind20	1	0.0074^{**}	0.3772^{**}					
Ind21	1,3	0.0037^{**}	0.3808^{**}	0.1795^{**}				
Ind22	1,3	0.0040^{**}	0.2603^{**}	0.1824^{**}				
Ind23	1	0.0040^{*}	0.2978^{**}					
Ind24	1,12	0.0042^{*}	0.3576^{**}	0.1632^{**}				
Ind25	1,23	0.0066^{**}	0.2897^{**}	-0.1478^{*}				
Ind26	1	0.0063^{**}	0.2601^{**}					
Ind27	1	0.0049^{*}	0.1750^{**}					
Ind28	1	0.0052^{**}	0.2738^{**}					
Ind29	1,10	0.0037	0.2454^{**}	0.1403^{*}				
Ind30	1,19	0.0041^{**}	0.1928^{**}	-0.1831^{**}				

Note: (*) significant at level 0.05, (**) significant at level 0.01

According to the results of Table 6, it can be seen that the significance level of the Lejang Box test, ARCH effects test and Kolmogorov-Smirnov test showed that the initial assumptions of the models were established. Based on the R^2 obtained for the models, it is observed that the model fitted to the petroleum products industry (Ind16) with R^2 equal to 0.3016 had the greatest power in explaining the changes in index returns, while the lowest value of the R^2 with the same interpretation is related to the oil extraction industry except exploration (Ind27) with R^2 equal to 0.0271.

In order to compare the performance of both methods in predicting the returns, 30 out-of-sample observations in both methods based on fitted models were predicted and RMSE and MAE error criteria were estimated for each model separately for each industry. Table 7 shows the results of estimating forecast error criteria for both methods and separately for each industry.

Table 6. Goodness of fit indexes of Bayesian averaging models

Industry	R^2	BIC	Ljung-Box	ARCH Effects	K-S
Ind1	0.4584	0.8347	0.3603	-44.7203	0.1669
Ind2	0.3391	0.8074	0.2410	-46.2569	0.1717
Ind3	0.5276	0.9414	0.9882	-36.1584	0.1546
Ind4	0.3290	0.8923	0.0569	-30.4499	0.1168
Ind5	0.1358	0.6130	0.6561	-51.4096	0.1873
Ind6	0.6992	0.5752	1.0000	-35.9948	0.1396
Ind7	0.9544	0.4456	1.0000	-38.4469	0.1474
Ind8	0.9069	0.3214	0.2596	-41.8372	0.1580
Ind9	0.3924	0.9117	0.0943	-37.5714	0.1590
Ind10	0.1266	0.2971	0.6180	-49.4459	0.1814
Ind11	0.0628	0.3147	0.0533	-31.5300	0.1544
Ind12	0.8672	0.6527	0.3309	-41.7085	0.1576
Ind13	0.2892	0.8462	0.5378	-24.9122	0.1186
Ind14	0.4554	0.9866	0.0769	-44.7827	0.1812
Ind15	0.1293	0.9971	0.0750	-30.0850	0.1205
Ind16	0.0641	0.6364	0.0535	-69.4370	0.3016
Ind17	0.0597	0.8133	0.0700	-30.3710	0.1363
Ind18	0.4575	0.7206	0.2925	-15.8222	0.0728
Ind19	0.3331	0.8990	0.0703	-11.8241	0.0589
Ind20	0.0687	0.6471	1.0000	-35.5637	0.1382
Ind21	0.7372	0.3541	0.9319	-49.3930	0.1950
Ind22	0.0865	0.9624	0.0641	-20.7136	0.1048
Ind23	0.0719	0.6856	0.8407	-19.6430	0.0858
Ind24	0.1094	0.8144	0.4123	-34.2474	0.1486
Ind25	0.2630	0.8117	0.1539	-20.4220	0.1039
Ind26	0.6887	0.7992	0.1387	-13.2564	0.0639
Ind27	0.1077	0.9265	0.9692	-2.8288	0.0271
Ind28	0.1122	0.6012	0.1824	-12.7935	0.0623
Ind29	0.0646	0.7808	1.0000	-12.8938	0.0785
Ind30	0.0791	0.8044	0.6338	-12.2658	0.0763

Table 7. Predictive error criteria under out of sample observations

Industry	Autoregressive Model		Bayesian Model	
	RMSE	MAE	RMSE	MAE
Ind1	0.02912	0.02545	0.01758	0.01543
Ind2	0.02152	0.01977	0.01238	0.01141
Ind3	0.02338	0.02019	0.01345	0.01155
Ind4	0.03197	0.02783	0.02477	0.02133
Ind5	0.02684	0.02417	0.01533	0.01379
Ind6	0.01939	0.01585	0.01236	0.00994
Ind7	0.01915	0.01625	0.01173	0.00995
Ind8	0.02510	0.02107	0.01550	0.01310
Ind9	0.03645	0.03111	0.03588	0.03053
Ind10	0.03261	0.02707	0.01867	0.01563
Ind11	0.57163	0.16290	0.57111	0.16615
Ind12	0.20931	0.01741	0.01241	0.01036
Ind13	0.02991	0.02576	0.02632	0.02292
Ind14	0.55871	0.16176	0.55794	0.15803
Ind15	0.03070	0.02834	0.02041	0.01875
Ind16	0.56248	0.16845	0.55702	0.16478
Ind17	0.58056	0.17653	0.57710	0.16838
Ind18	0.03366	0.02973	0.02483	0.02190
Ind19	0.03343	0.03054	0.02526	0.02304
Ind20	0.03267	0.02548	0.02098	0.01650
Ind21	0.02262	0.01915	0.02413	0.02146
Ind22	0.03523	0.03197	0.03685	0.03216
Ind23	0.03535	0.03129	0.02480	0.02193

Industry	Autoregressive Model		Bayesian Model	
	RMSE	MAE	RMSE	MAE
Ind24	0.03376	0.03019	0.03350	0.02958
Ind25	0.03862	0.03246	0.03744	0.03169
Ind26	0.02458	0.02107	0.01838	0.01579
Ind27	0.03546	0.03143	0.02925	0.02605
Ind28	0.03738	0.03449	0.02728	0.02551
Ind29	0.03687	0.03186	0.03764	0.03201
Ind30	0.00704	0.00443	0.01104	0.01022

In order to compare the prediction accuracy of the models, descriptive indices of error values are presented in Table 8.

Table 8. Description of error criteria

	Autoregressive Model		Bayesian Model	
	RMSE	MAE	RMSE	MAE
Mean	0.10090	0.04413	0.09503	0.03899
Mode	0.03264	0.02808	0.02478	0.02167
Std.	0.16636	0.04965	0.18799	0.05051
Min	0.00704	0.00443	0.01103	0.00993
Max	0.58056	0.17653	0.57710	0.16838
K-S	2.7240	2.430	2.666	2.303
(sig.)	(0.000)	(0.000)	(0.000)	(0.000)

According to the results in Table 8, the average error values for RMSE in the autoregressive model equals to 0.1009 and in the Bayesian model equals to 0.0950. Also, the average of error values for MAE in the autoregressive model equals to 0.0441 and in the Bayesian model equals to 0.0389. It is estimated that the Bayesian model has performed better in reducing the prediction error. To examine this difference for each industry paired comparison test was used. Prior to choosing the type of comparison test, the hypothesis of normal distribution among error criterion was tested by Kolmogorov–Smirnov test and the outcome showed that experimental distribution of error criterion was not normal. Therefore, a nonparametric Wilcoxon test was used to compare statistically error values of prediction. The results of this test are presented in Table 9.

Table 9. Comparison of method prediction error

Test (BMA - AR)	RMSE	MAE
Wilcoxon Stat.	-4.1650	-4.1650
sig.	0.000	0.000

According to the results of the Wilcoxon test in Table 9, it can be seen that the significance level of the test for both RMSE and MAE criteria is less than 0.05 and shows a significant difference between the values of these error criteria between the two methods. The negative test statistic, which is equal to 6 decimal places for both criteria, shows that each of the RMSE and MAE criteria under the Bayesian model was smaller than the values of these criteria under the classical autoregressive model. Therefore, it can be concluded that Bayesian model averaging has higher accuracy in predicting returns than autoregressive models and therefore the research hypothesis has been confirmed at the error level of 0.05.

Conclusion

In the present research, stock returns were predicted using the Bayesian model approach in Tehran Stock Exchange. According to it, the hypothesis of the research based on the Bayesian model has higher accuracy in predicting returns than autoregressive models that were developed and tested. In order to examine the hypothesis, information related to the index of 30 selected industries in the Tehran Stock Exchange during the period from 2017.03.25 to 2020.08.24 was used. The index return was predicted based on two methods for 30 out of sample data. First, autoregressive models were fitted on returns of each index and then the next 30 days of returns were predicted based on these models. Then after identifying the optimal model lags through the Bayesian Model Averaging method, autoregressive models were fitted with the optimal lags and the next 30 days predictions were obtained under this method.

In order to compare the accuracy of the methods in predicting the return, RMSE and MAE criteria were used, and the values of these error criteria were compared using Wilcoxon Nonparametric pairwise comparison tests. The results showed that the Bayesian Method leads to an increase in the accuracy of the prediction model in out of sample data. Classic methods for return prediction rely on the fitness of simple or multiple linear regressive models which include the history of information about price return, or other variables related to return.

In algorithm methods, researchers paid special attention to the nervous system and ultra-technical algorithms in different types.

In compound methods also both classic and predicting algorithms have been considered. Meanwhile, all research has more tendencies to use previous information to build up predicting models. Previous information includes a set of information about the primary distribution of a parameter or a variable which result into the formation of previous distribution and the Bayesian model, e.g., [11] ,[32] ,[33] ,[38] basically carried out on the Bayesian model.

The advantage of the Bayesian analysis contributes to modeling of predicting model is that it adds primary information to a model which increases the chance of access to optimal models.

All mentioned research emphasized on the high potential of the Bayesian Models in return prediction and the result of the research hypothesis is parallel with the result of those researches.

In the researcher's Point of view Bayesian models through a combination of different sets of predicting variables (including return lags in this research) recognize the best model based on prefabricated criterion (Schwarz Criteria in this research). The advantage of this analysis relatively to classic analysis is that in the latest one recognition of all components of previous predicting variable is extremely time consuming, certainly human error looks very effective in recognition of optimal model and effective lags. However, by Bayesian models, it is possible to weigh each component that indicates its importance we can optimize the optimal model based on previous information about predicting variables. Needless to say, there is a limitation for the number of model parameters. Particularly as long as the number of parameters increases such as successive and long-term lags in classic modeling, the number of error predictions goes up, so model accuracy in return prediction will decrease.

But Bayesian approach filters effective lags in the model, and it leads to fewer errors in prediction. Therefore, the results from this research have been predictable and it is expected adding further information to predicting issues (under Bayesian analysis) can improve model accuracy for the prediction of out of sample data.

Future research

Due to the greater ability of Bayesian models to predict returns than classical models, it is suggested that in cases where there is information based on technical analysis of the stock, in order to predict future returns, the Bayesian methods and weighted available models using specific probabilities allocation be used. Adding additional information to the problem can increase the accuracy of the prediction. Also, the use of macroeconomic variables and indicators affecting the market, such as producer price index, exchange rate, etc. can be effective in improving the results.

References

[1] Hadi Doulabi, N., M. Rastegar, P. Mohammadi, *Measuring the Stock Liquidity Using a Market Microstructure Approach.* Advances in Industrial Engineering, 2020. 54(3), 311-331. doi: 10.22059/jieng.2021.325016.1770

[2] Kumar, G., and A. K. Misra, *Closer view at the stock market liquidity: A literature review*. Asian Journal of Finance and Accounting, 2015. 7(2), 35-57.

[3] Alamatian, Z., M. Vafaei Jahan, *Iran Stock Market Prediction Based on Bayesian Networks and Hidden Markov Models.* 2017. 8(33), 283-298.

[4] Quah, H., J. Haman, and D. Naidu, *The effect of stock liquidity on investment efficiency under financing constraints and asymmetric information: Evidence from the United States*. Accounting and Finance, 2021. 61, 2109-2150

[5] Będowska-Sójka, B., *Commonality in liquidity measures*. The evidence from the Polish stock market. 2019.

[6] Johann, T., S. Scharnowski, E. Theissen, C. Westheide, and L. Zimmermann, *Liquidity in the German stock market.* Schmalenbach Business Review, 2019. 71(4), 443-473

[7] Fakhari, H., M. Valipour Khatir, M. Mousavi, *Investigating Performance of Bayesian and Levenberg-Marquardt Neural Network in Comparison Classical Models in Stock Price Forecasting*. Financial Research Journal, 2017. 19(2), 299-318. doi: 10.22059/jfr.2017.214203.1006264

[8] Sousa J. M.,R. M. Sousa, *Asset Returns Under Model Uncertainty: Evidence from the Euro Area*, the US and the UK, Comput Econ, 2017. DOI 10.1007/s10614-017-9696-2

[9] Rather, A. M., A. Agarwal, V. N. Sastry, *Recurrent neural network and a hybrid model for prediction of stock returns*. Expert Systems with Applications, 2015. 42(6), 3234-3241

[10] Susannah H. M., C. Curme, A. Avakian, Y. Dror, H. Kenett, E. Stanley, T. Preis, *Quantifying Wikipedia Usage Patterns Before Stock Market Moves*. Scientific Reports, 2013. 3: 1801.

[11] Wang H., R. Chatpatanasiri, P. Sattayatham, *Stock Trading Using PE ratio: A Dynamic Bayesian Network Modeling on Behavioral Finance and Fundamental*, Investment School of Mathematics, Suranaree University of Technology, 2017. THAILAND.

[12] Zhou, X., J. Wang, X. Yang, B. Lev, Y. Tu, S. Wang, *Portfolio selection under different attitudes in fuzzy environment*. Information Sciences, 2018. 462, 278-289

[13] Bollen, J., M. Huina, X. Zeng, *Twitter mood predicts the stock market*. Cornell University, 2010. Retrieved November 7.

[14] Zhang, Y., D. W. Gong, X. Y. Sun, Y. N. Guo, *A PSO-based multi-objective multi-label feature selection method in classification*. Scientific reports, 2017. 7(1), 1-12.

[15] Schroeder, P., I. Kacem, G. Schmidt, *Optimal online algorithms for the portfolio selection problem, bi-directional trading and-search with interrelated prices*. RAIROOperations Research, 2019. 53(2), 559-576

[16] Mokhtarzadeh, M., R. Tavakkoli-Moghaddam, C. Triki, Y. Rahimi, *A hybrid of clustering and meta-heuristic algorithms to solve a p-mobile hub location–allocation problem with the depreciation cost of hub facilities.* Engineering Applications of Artificial Intelligence, 2021. 98, 104121.

[17] Malkiel, Burton G., *A Random Walk Down Wall Street (6th ed.).* W.W. Norton & Company, Inc. 1973.

[18] Ball P., *Counting Google searches predicts market movements*. Nature. 2013. doi:10.1038/nature.2013.12879

[19] Li, X., Y. H. Huang, S. C. Fang, Y. Zhang, *An alternative efficient representation for the project portfolio selection problem*. European Journal of Operational Research, 2020. 281(1), 100-113.

[20] Beckmann, M., *Doctoral Thesis: Stock Price Change Prediction Using News Text Mining*. COPPE/Federal University of Rio de Janeiro, 2017.

[21] Landsman, Z., U. Makov, T. Shushi, *A generalized measure for the optimal portfolio selection problem and its explicit solution.* Risks, 2018. 6(1), 19.

[22] Garcia, F., J. González-Bueno, J. Oliver, R. Tamošiūnienė, *A credibilistic meansemivariance-PER portfolio selection model for Latin America.* Journal of Business Economics and Management, 2019. 20(2), 225-243.

[23] Jacaruso, L. C., *A method of trend forecasting for financial and geopolitical data: inferring the effects of unknown exogenous variables*. Journal of Big Data, 2018. 5 (1): 47.

[24] Sobhanifard, F., M. Shahraki, *An Integrated Neural Networks and MCMC Model to Predicting Bank's Efficiency.* Advances in Industrial Engineering, 2020. 54(1), 1-14. doi: 10.22059/jieng.2021.312818.1743

[25] Wanke, P., M. D. A. Azad, C. P. Barros, *Predicting effciency in Malaysian Islamic bank: A two-stage TOPSIS and neural networks approach*. Business and Finance, 2016. 36, 485–498.

[26] Zhang, Y., L. Wu, *Stock Market Prediction of S&P 500 via combination of improved BCO Approach and BP Neural Network*. Expert Systems with Applications, 2009.36 (5): 8849–8854.

[27] Atashgar, K., N. Rafiee, *Identification of the change point in panel data using simultaneously EWMAA and CUSUM*. Advances in Industrial Engineering, 2019. 53(1), 471-481.

[28] Harvey, A., P. Kattuman, *Time series models based on growth curves with applications to forecasting coronavirus*. Harvard Data Science Review, 2020.

[29] Kambouroudis, D. S., D. G. McMillan, K. Tsakou, *Forecasting stock return volatility: a comparison of GARCH, Implied volatility, and realized volatility models*. Journal of futures markets, 2016. 36(12), 1127-1163.

[30] Cho, H., *Change-Point Detection in Panel Data Via Double CUSUM Statistic*, Electronic Journal of Statistics, 2016. 10(2), 2000-2038.

[31] Enomoto, T., Y. Nagata, *Detection of Change Points in Panel Data Based on Bayesian MT Method*, Total Quality Science, 2016. 2(1), 36-47.

[32] Chen S., *Predicting Stock Returns Using Firm Characteristics: A Bayesian Model Averaging Approach*, Department of Economics and Finance, City University of Hong Kong, 2018.

[33] Gao Z., *Stock Investment Selection Management Based on Bayesian Method, Advances in Economics,* Business and Management Research, 2018. Vol. 75, 407-413.

[34] Atkins A., M. Niranjan, E. Gerding, *Financial news predicts stock market volatility better than close price*, The Journal of Finance and Data Science, 2018. 4(2), 120-137.

[35] Hamidian M., S. Boostani , H. Mashhadi, *Predicting negative returns on stocks of companies listed on the Iranian capital market,* Journal of Decisions and Operations Research, 2019. 4(2), 30-40. (In Persian)

[36] Barzegari Khanagha, J., Z. Jamali, *Predicting Stock Returns with Financial Ratios; An Exploration in Recent Researches*. Journal of accounting and social interests, 2016. 6(2), 71-92. (In Persian)

[37] Salehirad M. R., N. Habibifard, *Comparing of Bayesian Model Selection Based on MCMC Method and Finance Time Series*. Financial knowledge of security analysis, 2012. 5(15), 59-67. (In Persian)

[38] Dehghan Dehnavi, M., M. Bahrololoum,, M. Peymany Foroushany, S. Raeiszadeh, *Portfolio Selection Optimization Problem Under Systemic Risks.* Advances in Industrial Engineering, 2020. 54(2), 121-140.