University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 27/2023 (434)

Predicting DJIA, NASDAQ and NYSE Index Prices Using ARIMA and VAR Models

Sahil Teymurzade
Robert Ślepaczuk

Warsaw 2023

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Predicting DJIA, NASDAQ and NYSE index prices using ARIMA and VAR models

Sahil Teymurzade [a], *Robert Ślepaczuk* [b*]

[a] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group*
[b] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group, Department of Quantitative Finance*
* *Corresponding author: rslepaczuk@wne.uw.edu.pl*

Abstract: This paper implements automated trading strategies with buy/sell signals based on Autoregressive Integrated Moving Average (ARIMA) and Vector autoregression (VAR) models. ARIMA and VAR models are compared based on several forecast error measures and investment performance statistics. The data used in this thesis are daily closing prices of Dow Jones Industrial Average, NASDAQ Composite and NYSE Composite indices. The trading period covers 20 years of data from 2000-11-30 to 2020-11-30. The sensitivity analysis is made by changing the initial parameters to test how robust the methods are to these changes. Results show that although ARIMA model performed remarkably well during the volatile periods, VAR based strategy had better investment performance and was less robust to the changes compared to the ARIMA based strategy. Additionally, we have found that error metrics might be insufficient to evaluate performance of forecasting models, as VAR with higher forecast errors outperformed ARIMA model in algorithmic trading strategies.

Keywords: ARIMA model, VAR model, time series analysis, algorithmic trading strategies, investment systems, statistical models, forecasting stock prices

JEL codes: C4, C14, C45, C53, C58, G13

Note: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Introduction

There are many research papers which attempted to predict stock prices. Accurate stock price forecasts can help investors in making trade decisions, maximizing returns, and minimizing risks. Different forecasting models are applied by investors when predicting future prices. In recent years, as studies on artificial intelligence, machine learning, and deep learning have increased, algorithms have started to be used extensively in transactions in investment markets. The introduction of computers and more advanced systems in the trading process is called algorithmic or automated trading. The algorithmic trading made the data analysis and decision-making process much faster and more rational than the human approach.

In this paper stock price indices were predicted by using forecasts from Autoregressive Integrated Moving Average (ARIMA) and Vector autoregression (VAR) models in algorithmic trading strategies. The trading signals were generated based on the forecasts from ARIMA and VAR models. Signals and trades were generated for the last 20 years period for the following equity indices: Dow Jones Industrial Average (DJIA), NASDAQ Composite index (NASDAQ) and NYSE Composite index (NYSE). The main aim of the paper was to evaluate the quality of ARIMA and VAR models forecasts and the performance of ARIMA and VAR models in algorithmic trading strategies. Three hypotheses were tested in this paper.

(H1) *ARIMA and VAR models have similar forecasting power*. The intuition behind this hypothesis is based on the results of the discussed research papers that were devoted to the comparative analysis of forecasting models (explained in the literature review). The results from similar papers revealed that in some instances ARIMA outperformed the VAR model based on error metrics while in some instances the VAR model had lower forecast errors compared to ARIMA models. However, ARIMA and VAR forecast errors were not significantly different. The outcomes obtained from the different papers were dependent on the characteristics of particular data samples based on which the models were estimated.

(H2) *VAR models should be more robust to the changes than ARIMA models*. The intuition behind this hypothesis is based on the fact that VAR models incorporate more variables than ARIMA models, hence it makes VAR models be better at describing different dynamics of the time series. Therefore, it is assumed that the VAR model can bring more realistic results in case of the changes to the parameters.

(H3) *The model with more accurate forecasts might not perform better when applied to an algorithmic investment strategy.* The intuition for this assumption is based on the fact that the error metrics describe the overall accuracy of generated forecasts. However, the investment performance of a trading strategy is evaluated by computing different performance and risk metrics after transforming these forecasts into investment signals. Hence the model with lower error metrics might not have higher investment performance statistics.

This paper has six chapters. The first one is the literature review which discusses the papers focusing on the impact of algorithmic trading strategies in the financial markets, papers analyzing different algorithmic trading strategies based on statistical and machine learning methods, and ARIMA and VAR model comparisons. The second chapter describes the data used in this research. The third chapter focuses on methodology and explains forecasting techniques, signal generation methods, and performance measures. The fourth chapter presents the empirical results of each method and their performance metrics. The fifth chapter shows sensitivity analysis. The conclusion part summarizes, verifies hypotheses, and provides recommendations for future research.

1. Literature review

Forecasting the future prices of stocks has been an attractive topic for many researchers. Due to its complexity and potential unprecedented impact on the markets, it is a difficult task to predict the future stock prices correctly, especially the long-term trends. However, even though the stock markets are volatile, and it is often considered chaotic, by using different algorithms applied to the historical data it is possible to forecast the future prices to some extent.

One of the main advantages of algorithmic trading over human traders is that algorithmic trading makes trading decisions automatically and without being influenced by emotions and biases which can often be a challenge for human traders to control. However automated trading processes should often be controlled by humans and processes to avoid possible technological issues and errors that can happen due to various reasons such as internet disconnection, etc.

Algorithmic trading changed the dynamics of the financial markets dramatically. The effect of algorithmic trading on the market has been studied by many researchers. Gsell (2008) compared the results of different simulation runs by including and excluding trading agents and assessed the impact of algorithmic trading on the market. It was concluded that the larger volumes executed by the algorithmic traders result in an increase in the market prices, while lower latency (delay in the amount of time that takes a client's order to be executed by the server) can result in lower market volatility. Verheggen (2017) researched the impact of automated trading on the forecast accuracy and on the dispersion (the idiosyncratic volatility). The research revealed that the increased algorithmic trading causes decrease in the idiosyncratic volatility and can reduce the errors in the forecasts made by the human analysts in the market. Boehmer et al. (2020) examined the relationship between automated trading and market liquidity, informational efficiency (how correctly the market reflects the true value of an asset), short-term volatility and the impact of automated trading on buy-side institutional investors. The study concluded that automated trading improves liquidity and informational efficiency and increases short-term volatility. It was also found that automated trading decreases execution shortfalls (the difference between the price or value of an asset when a buy or sell decision is made and the final execution price) for buy-side institutional investors. Conrad et al. (2015) provided evidence on the effect of high frequency quoting in market quality. The paper focused mainly on examining market outcomes. The relation between high-frequency quotations and the behavior of stock prices between 2009 and 2011 for the full cross-section of securities in the US was analyzed. It was found that there is a positive relation between high-frequency quotes and improvements in the efficiency of the price discovery process and reductions in the cost of trading. Stephan (2016) researched the relation between algorithmic trading and managerial disclosure decisions. It was found that there is a positive relation between algorithmic trading and the likelihood, and the quantity of guidance issued.

Performance comparison and evaluation of different algorithmic trading strategies is important for both academic purposes and for the investment management industry. Ryś and Ślepaczuk (2018) formulated and analyzed different machine learning methods and compared the efficiency of the methods in the case of moving average crossover system. Three machine learning methods (EHC, GM and DEM) were applied by using simple moving averages crossover strategy optimization problems. The value of optimization criteria was used to compare the machine learning methods. The optimization criteria values and the computation time (required to proceed the whole search process) were compared with the Exhaustive Search. It was found that the obtained results from machine learning methods and Exhaustive Search were similar, however, machine learning methods required much less execution time.

In the research prepared by Bilyk et al. (2020) the performance of VIX futures trading strategies built across different GARCH model volatility forecasting techniques was compared. It was found that using the daily data over the seven-year period (2013-2019), strategy based on the fGARCH-TGARCH and GJR-GARCH specifications outperformed those of the GARCH and EGARCH models and performed slightly below the "buy-and-hold" S&P 500 index strategy. Another finding of this research was that the classical GARCH model had better forecasting power under RMSE (Root Mean Square Error).

Kijewski and Ślepaczuk (2020) introduced a study in which the performance of investment strategies based on classical techniques (e.g., ARIMA model) and recurrent neural network model (LSTM) was compared. The trading algorithms were applied to S&P 500 index prices covering 20 years of data from 2000 to 2020. Most of the used strategies were not able to achieve better results than the benchmark (Buy and Hold strategy). Sensitivity analysis results showed that classical methods which used rolling training-testing window were significantly more robust to changes in parameters compared to LSTM model in which hyperparameters were selected heuristically.

Sporer (2020) introduced a backtesting tool for cryptocurrency markets that "allows investors to build and test algorithmic trading strategies for all major cryptocurrencies". Some investment strategies were tested with this tool and results showed that backtesting performed well for realistic applications.

Vo and Ślepaczuk (2022) constructed an algorithmic investment strategy applying the ARIMA-(S)GARCH on S&P500 stocks. The authors found that adding the GARCH component to the ARIMA model increases the precision of the forecasts and causes underlying investment strategies to outperform benchmarks. Moreover, their findings seemed not to be sensitive to a variety of parameters, including error distribution or GARCH model type (SGARCH, EGARCH). Based on the literature, we can conclude that the addition of GARCH components can potentially increase the precision of the forecasts.

There were also multiple attempts by researchers to forecast and compare VAR and ARIMA performance. Devi et al. (2013) used the historical data for NSE – Nifty Midcap50 and applied ARIMA model by using the Box Jenkins methodology to forecast the future prices. Bagshaw (1987) compared univariate and multivariate ARIMA with VAR forecasts. The forecast accuracy of these models was tested in terms of RMSE. The study also found that the method that performs the best in terms of RMSE was the multivariate ARIMA model. Javed (2013) conducted a comparative analysis of the forecasting performance of ARIMA, Regression Analysis, Vector Autoregression (VAR), Error Correction Model (ECM) and ARCH/GARCH models. Data used was Pakistan's export to United Sates and money supply. RMSE was computed for each model forecasts and compared. It was found that no single forecasting method provided a better forecast for both series. Ikechukwu and Adedoyin (2014) forecasted core inflation in Nigeria by using ARIMA and VAR models. The study found that the VAR model had smaller errors in terms of the minimum square error (MSE). Espasa et al. (2002) found that ARIMA model outperformed the VECM and dynamic factor models. Kirstin (2003) found that VAR models outperformed the autoregressive forecasting models. Zhang (2013) tested the forecasting performance of three different autoregressive models forecasting regional GDP per capita. The results showed that, AR(1) model performs better than the other two models and ARIMA model was better than the VAR models. Thomakos and Guerard (2020) compared the forecasting performance of some parametric and nonparametric models based on a training-validation sample approach. The models included naive, ARIMA, transfer function (TF) and VAR models using a variety of datasets. RMSE was computed and it was found that the bivariate models had a better performance than the univariate models. VAR and TF had the best performance. It was also found that the combined forecast was better in the longer horizon. Kumar (2010) implemented a vector autoregression (VAR) model with time-varying parameters (TVP) to predict the daily INRUSD exchange rates. The method is based on the characterization of the TVP as an optimal control problem. The out-of-sample forecasting performance of the TVP-VAR model were compared with simple VAR and ARIMA models, by employing a cross-validation process and computing popular forecast error metrics such as mean absolute error (MAE), RMSE, etc. It was found that the TVP-VAR model had better forecasting performance than the simple VAR and ARIMA models. In the other research by Anggraenia et al. (2017) ARIMAX model and VAR models were used to forecast the price of Rice and it was found that the VAR model outperformed ARIMA in terms of forecast accuracy measures.

Reviewed literature mainly introduced the algorithmic trading strategies by using classical and machine learning models and comparison of forecasting accuracy of different methods by using traditional forecast error measures such as MAE, RMSE, etc. However, they don’t examine the performance of statistical models such as ARIMA and VAR by implementing their forecast in investment strategies and compare the investment performance statistics. Additionally, the reviewed literature concerning the comparative analysis of the statistical models mainly used one example of parameters without checking the robustness of the results to the changes. To be sure that the research findings are not random results because of the initially set parameters, sensitivity analysis is needed to check how the outcomes would be affected by the changes in the parameters.

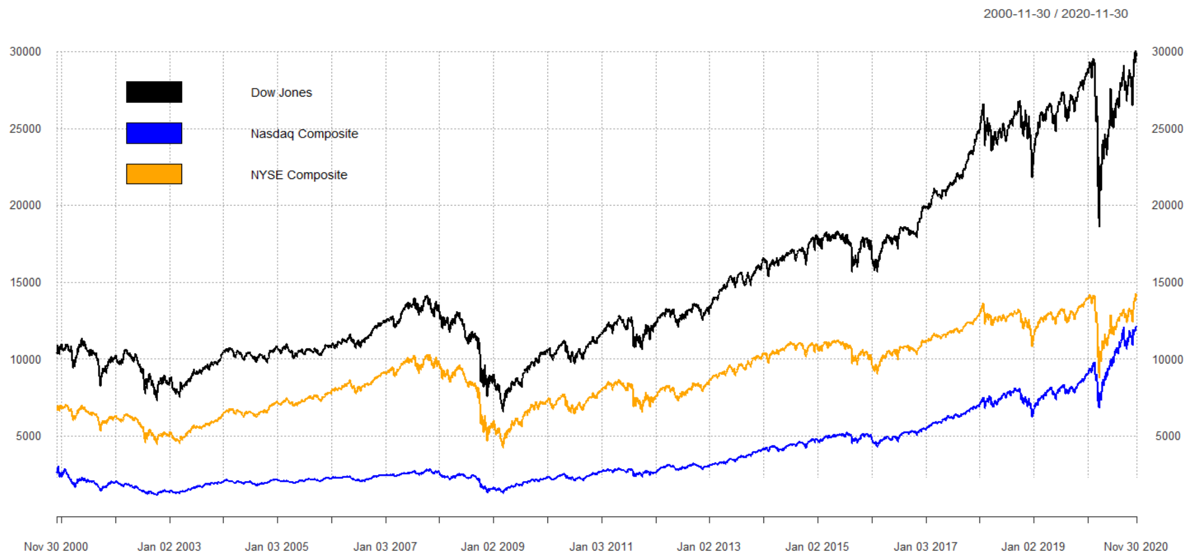
2. Data description and initial data analysis

2.1. Data description

The data used in this paper are adjusted closing prices of the three US stock market indices: NASDAQ Composite (NASDAQ), NYSE Composite (NYSE), Dow Jones Industrial Average (DJIA). The index prices cover the period from 1997-12-11 to 2020-11-30. The trading period is between 2000-11-30 and 2020-11-30. Additionally, adjusted closing prices of 10y Treasury Yield were used to calculate an average proxy for the risk-free rate (3.18%). The data source is Yahoo Finance (finance.yahoo.com).

The stock market indices used in this paper are widely followed stock indices. NYSE and NASDAQ cover a significant proportion of the global equity market. There are more than 2500 stocks in the NASDAQ and more than 2000 stock equities are in the NYSE. DJIA presents the performance of 30 big companies in the US. Stock index prices for the above-mentioned indices over the trading period are shown in Figure 1. In general, they have similar upward and downward trends over the period.

Figure 1. DJIA, NASDAQ, and NYSE indices



Note: DJIA, NASDAQ, and NYSE in USD for the period 2000-11-30 - 2020-11-30.

2.2. Initial data analysis

In this section, the initial data analysis was conducted to have a general view of the data. Basic statistical properties were computed for daily log returns of DJIA, NASDAQ and NYSE indices and the distribution of log returns were compared with the randomly generated and normally distributed vectors with the same length, standard deviation and mean.

Basic statistics for DJIA, NASDAQ and NYSE were shown in Table 1. Additionally, the Jarque-Bera test was conducted to test normality and the p-values were added in Table 1. Daily log returns of DJIA, NASDAQ and NYSE statistics indicate that the log returns of these market indices follow non-normal distribution. Kurtoses for all the three stock indices are considerably higher than 3 which results in taller peaks and fatter tails in the distribution of simple log returns. P-values of the Jarque-Bera tests for normality also confirm the non-normal distribution of log returns. The null hypothesis is that the data is normally distributed, while the alternative hypothesis is that the data is not normally distributed. As all p-values of the Jarque-Bera tests for log returns were equal to 0, the null hypothesis was rejected in all three cases.

Table 1. Descriptive statistics for DJIA, NASDAQ, and NYSE daily log returns

Name	Mean	SD	1st quartile	3rd quartile	Skewness	Kurtosis	Jarque-Bera test p-value
DJIA	0.02%	1.21%	-0.45%	0.55%	-0.38	13.62	0
NASDAQ	0.03%	1.51%	-0.59%	0.74%	-0.1	7.48	0
NYSE	0.01%	1.25%	-0.46%	0.57%	-0.59	12.6	0

Note: Simple descriptive statistics were calculated by using the daily log returns of DJIA, NASDAQ and NYSE indices for the period 2000-11-30 - 2020-11-30.

The density curves for DJIA, NASDAQ and NYSE indices (figures not presented in this study) also confirm the fact that the log returns are not normally distributed. The density curve of log returns has a fatter tail and a higher peak than the normal distribution. Therefore, it can be concluded that the log returns have leptokurtic distribution.

3. Methodology

Investment strategies were built based on the forecasts that were calculated by applying univariate ARIMA and three-dimensional VAR models. The results were compared with the benchmark strategy – Buy&Hold. Initial investment assumptions and rules were the same for each strategy. Initial capital was assumed as 1 mln USD. Depending on the next day forecasts and the available capital one of the three possible signals was generated on each trading day – BUY, SELL or HOLD. As the names suggest, in the case of a buy signal the algorithm will use all capital (cash) for investing in long positions. In the case of a sell signal, the algorithm will use all available capital to open short position. HOLD signal means that the previous position will be held on the next day (If HOLD signal is after SELL signal then the algorithm keeps the short position, if HOLD signal comes after BUY signal then it maintains the long position). Research assumes that the algorithm can buy or sell any portion/fraction of an index, e.g. in the form of futures contracts. The trading cost for each trade was assumed to be 0.01% of the invested capital.

3.1. Statistical models

3.1.1. ARIMA

Conventional linear regression models are usually insufficient to explain the different characteristics of a time series. Whittle (1951) introduced Autoregressive (AR) and autoregressive moving average models (ARMA). Later, Box and Jenkins (1970) added non-stationary models to the ARMA model and introduced the ARIMA modelling which means autoregressive integrated moving average.

ARIMA modelling is similar to ARMA modelling. The variable is usually differenced as many times as required to remove a trend and then an ARMA model is estimated on the differenced variables. If p is the order of the autoregressive model and q is the order of the moving average model, an ARMA (p, q) model for the variable that is differenced d times would be equal to an ARIMA (p, d, q) model on the original data. AR(p) and MA(q) processes can be respectively described in the formulas (1) and (2):

$$y_t = \mu + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + \varepsilon_t \quad (1)$$

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (2)$$

where:

μ is a constant,

p is the order of a moving average model,

q is the order of autoregressive model,

y_t is the value of the time series at time t,

ε_t is the error term at time t.

For simplicity we can remove μ constant from the equations. And ARMA (p, q) model can be presented then as:

$$y_t = \varphi_1 y_{t-1} + \ldots + \varphi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad (3)$$

For the model estimation, the values of an ARIMA (p, d, q) need to be defined. They can be found in several ways. Box and Jenkins (1976) presented a systematic approach in estimating an ARMA/ARIMA model.

Information criterion is another way of defining model orders. The model with the lowest information criterion is considered to be the best model. Information criterion method was used also in this paper to find the best model. The main reason for using information criteria was the automation process. The information criterion methods can help in avoiding subjective interpretation of ACF, PACF plots. In this paper, the chosen Information criterion was The Akaike information criterion (AIC). It was introduced by Akaike (1969). The general formula can be written as below:

$$AIC = -2ln(L) + 2K \quad (4)$$

where:

K is the number of estimated parameters in the model,

L is the maximum value of likelihood function for the model.

After defining the proper orders of the ARIMA model based on the minimum AIC values, the models were estimated with conditional-sum-of-squares method and one day ahead forecasts were calculated. Forecast precisions were assessed by forecast error measures.

The algorithm for predicting the next day forecasts for selected equity indices can be summarized as follow:

1. Select the first 500 days as a training window size.

2. Estimate the model with the conditional sum of squares method by using ARIMA values (p, d, q) that were defined with AIC. Optimal ARIMA values were identified by checking all combinations of p: 0-4, d: 0-3, q: 0-4.

3. Predict the one step ahead price.

4. If the predicted price is higher than the last real price plus trading costs - create BUY signal if the predicted price is lower than the last observed price minus transaction costs - create SELL signal. If none of the above conditions are met, then generate a hold signal.

5. Move the training window by one step forward and repeat the above-mentioned steps (1-4) till the end of the dataset.

6. Based on the generated signals construct the equity lines for the strategy and calculate the daily returns.

7. Calculate the performance statistics for the entire trading period.

3.1.2. VAR

Vector Autoregression models (VAR) are extended versions of univariate autoregressive models. The VAR model is considered to be one of the most successful models because of its suitability in describing different behaviors of the time series and its forecasting performance. VAR models were introduced by Sims (1980). They are used in modelling multivariate time series. In VAR models there are mutual dependencies between the variables.

If k is the number of variables ($k \geq 2$) the general equation for the VAR (p) model with the order of p can be written as follow:

$$Y_t = b + B_1 Y_{t-1} + B_2 Y_{t-2} + \cdots + B_p Y_{t-p} + u_t \tag{5}$$

where:

$Y_t = (y_{1t}, y_{2t}, \dots, y_{nt})'$: an $(k \times 1)$ vector of time series variables,

b - is an $(k \times 1)$ vector of intercepts,

B_i $(i = 1,2, \dots, p)$ are $(k \times k)$ coefficient matrices,

u_t - is an $(k \times 1)$ unobservable zero mean white noise vector process.

There are several methods to define the lag order for VAR models. In this paper AIC - Information Criterion was used to define proper lags of VAR (p) model. After calculating AIC values for up to 4 lags, the model with the lowest AIC was selected as the most appropriate model for forecasting purposes. Before estimating VAR models stationary tests were conducted by using Augmented Dickey Fuller test. Forecasting steps for VAR can be summarized as follow:

1. Select the first 500 days as a training window size.

2. Check stationary with Augmented Dickey Fuller test. Make the data stationary if needed.

3. Identify the optimal lag order by using AIC and estimate The VAR model by using OLS per equation.

4. Predict one step ahead price.

5. If the predicted price is higher than the last observed price plus trading costs - create BUY signal if the predicted price is lower than the last observed price minus transaction costs – create SELL signal. If none of the above conditions are met, then generate a hold signal.

6. Move the training window by one day forward and repeat the same steps (1-5) till the end of the dataset.

7. Based on the generated signals construct the equity lines for the strategy and calculate the daily returns.

8. Calculate the performance statistics for the trading period.

3.2. Performance measures

3.2.1. Forecast Error Measures

There are multiple criteria to evaluate the quality of forecasts. In this paper, three common forecast error metrics that were also reviewed by Hyndman and Koehler (2006) and Botchkarev (2018) were used to evaluate the forecast accuracy of the models.

- Mean Absolute Error:

$$MAE = \frac{1}{N}\sum_{j=1}^{n}|A_j - P_j| \tag{6}$$

- Root Mean Squared Error:

$$RMSE = \sum_{j=1}^{n}\frac{1}{N}(|A_j - P_j|)^2 \tag{7}$$

- Mean Absolute Percentage Error:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(A_j - P_j)}{A_j}\right| \tag{8}$$

For equations (6), (7) and (8) A_j denotes real values, P_j denotes forecasted values.

3.2.2. Investment performance statistics

Performance and risk measures assist investors in assessing the investment returns and risks more accurately. The chosen performance and risk metrics in this paper are widely used and also applied previously in papers analyzing algorithmic trading strategies such as Ryś and Ślepaczuk (2018) and Kijewski and Ślepaczuk (2020). Below are performance and risk metrics used in this paper:

- Annualized Return Compounded:

$$ARC = 252 * \frac{1}{N} * \Sigma_{i=1}^{N} r_t \tag{9}$$

where:

r_t - is daily logarithmic rate of returns,

N - is the number of trading days,

- Annualized Standard Deviation:

$$ASD = \sqrt{252} \times \sqrt{\frac{1}{N-1} \times \Sigma_{i=1}^{N}(r_t - \bar{r})^2} \tag{10}$$

where:

$\bar{r}$ - is the average of daily logarithmic rate of returns.

- Information Ratio (IR):

$$IR = \frac{ARC}{ASD} \tag{11}$$

- Maximum Drawdown is the highest percentage drawdown in the value of a portfolio in the trading period.

$$MD = \frac{Max\ \ value\ before\ the\ largest\ drop - Min\ value\ before\ new\ peak\ reached}{Max\ \ value\ before\ the\ largest\ drop.} \tag{12}$$

- Annual Return Compounded/Maximum Drawdown (ARCMD):

$$ARCMD = \frac{ARC}{MD} \tag{13}$$

- Sharpe Ratio (SR) can be represented as:

$$SR = \frac{ARC - R_f}{ASD} \tag{14}$$

where:

R_f - is the risk-free rate which was assumed to be the average of 10y Treasury Yield in the trading period (3.18%).

- AllRisk - the combined risk measure:

$$AllRisk = \frac{ASD \times MD}{100} \tag{15}$$

4. Empirical results

One step ahead forecasts for DJIA, NASDAQ and NYSE indices were computed for the last 20 years (2000-11-30 - 2020-11-30) by using ARIMA and VAR models. Based on the forecasts, the trading signals were generated and the equity lines during the trading period were presented for each investment strategy. The tables in this section present performance statistics for each investment strategy and forecast error metrics calculated for each model. In addition to the above-mentioned performance statistics and error measures, the number of days on which trades took place was also presented in the tables.

Each figure presented in this section has equity lines for ARIMA and VAR strategies and equity lines for Buy&Hold strategy. Buy&Hold strategy means an investor buys an asset on the first day of trading and keeps it over the trading period.

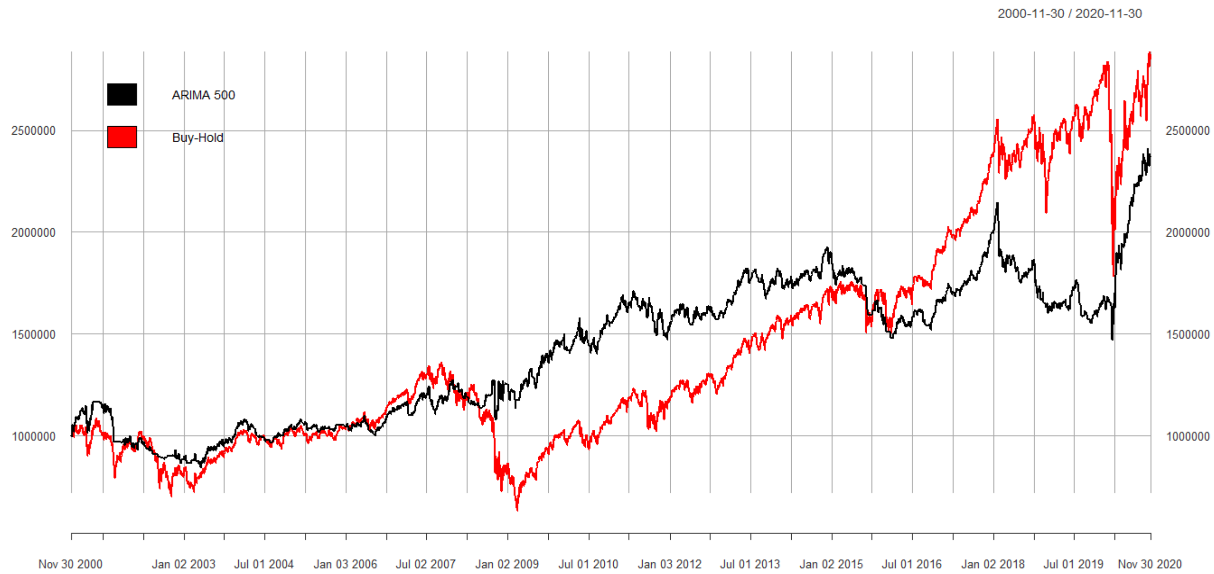
4.1. ARIMA results

4.1.1. ARIMA strategy applied for DJIA stock index

The results for the ARIMA model and Buy&Hold strategy for DJIA index were presented in Figure 2. ARIMA equity line had different levels of fluctuations, however an increasing trend was observed and reached its peak at the end of the trading period. In general, the ARIMA model had a remarkable performance during the highly volatile period and during the periods of big market crashes and downturns.

Performance and risk metrics are presented in Table 2. It shows that the ARIMA based strategy performed better than the benchmark strategy obtaining the higher Information Ratio (IR). ARIMA had a lower AllRisk measure than the benchmark which means the ARIMA based portfolio had a lower risk than the benchmark portfolio. However, the benchmark strategy had a higher Sharpe Ratio (SR) compared to the ARIMA-based strategy which means the benchmark had a better risk-adjusted performance. Considering most of the measures ARIMA model outperformed the benchmark strategy.

Figure 2. ARIMA strategy and the benchmark strategy applied for DJIA index.



Note: *ARIMA 500* indicates the equity line for ARIMA strategy applied for DJIA index by using 500 training days and identifying parameters using Akaike information criterion. *Buy-Hold* indicates equity line for Buy&Hold strategy applied for DJIA index.

Table 2. Performance statistics for ARIMA and the benchmark strategies applied for DJIA index.

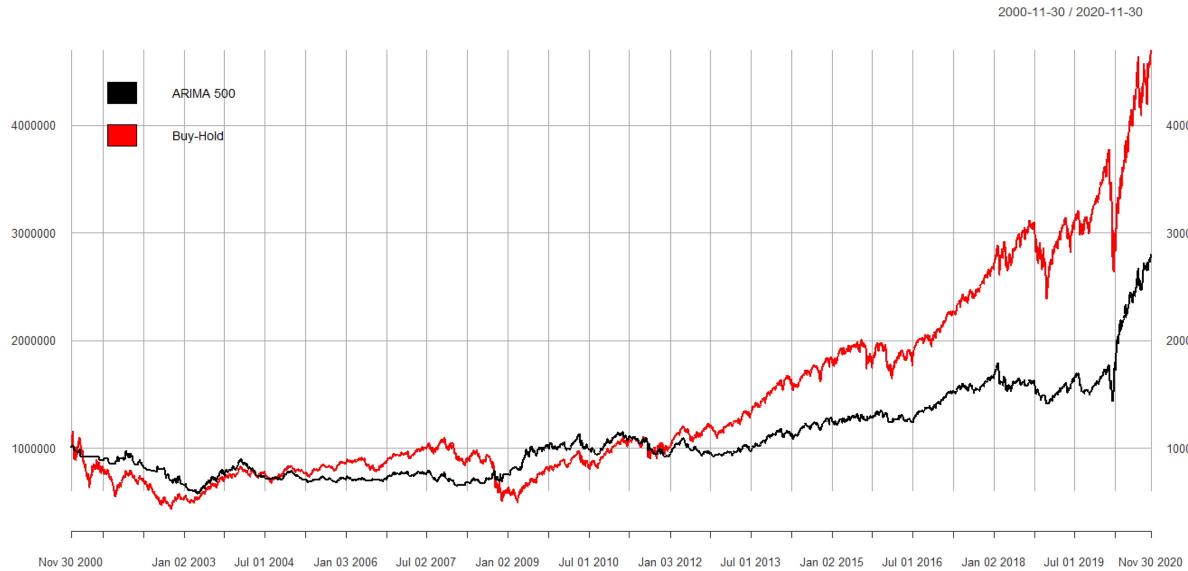
Names	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
ARIMA	5.15%	**12.74%**	**0.404**	**31.38%**	**0.164**	0.154	**4.00**	1465
BUYHOLD	**7.07%**	19.12%	0.370	53.78%	0.131	**0.203**	10.28	1

Note: ARIMA indicates the performance statistics for ARIMA strategy applied for DJIA index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for DJIA index. Bolded font indicates the best value with regards to each performance measure.

4.1.2. ARIMA strategy applied for NASDAQ index

The results for the ARIMA model for NASDAQ index were presented in Figure 3. The ARIMA equity line for NASDAQ was less volatile in comparison to Buy&Hold, however a general increasing trend was observed, and it reached its peak at the end of the trading period. Starting from the first trading day the equity line was below its starting value till 2009. The NASDAQ equity line was also influenced by the decline in the years 2001-2003. That decline reached its final low in 2003 when the US war with Iraq started. NASDAQ index equity line increased at the beginning of 2018. A considerable decrease can also be spotted in Feb 2020 when the Coronavirus Crash began. After that period there can be spotted a relatively stable increase in the equity line till the end of the trading period. In general, the ARIMA model had a remarkable performance during the highly volatile period and the periods of big market crashes.

Figure 3. ARIMA and the benchmark strategies applied for NASDAQ index.



Note: *ARIMA 500* indicates the equity line for ARIMA strategy applied for NASDAQ index by using 500 training days and identifying parameters using Akaike information criterion. *Buy-Hold* indicates equity line for Buy and Hold strategy applied for NASDAQ index.

Performance and risk metrics were presented in Table 3. It shows that the ARIMA based strategy did not outperform the benchmark strategy obtaining the lower Information Ratio (IR). However, ARIMA had a lower annualized standard deviation (ASD) and Maximum Drawdown (MD) which resulted in a lower AllRisk measure than the benchmark portfolio. That means ARIMA based portfolio had a lower risk than the benchmark portfolio. Additionally, the benchmark strategy had a higher Sharpe Ratio (SR) compared to the ARIMA based strategy which means the benchmark had a better risk-adjusted performance.

Table 3. Performance statistics for ARIMA and the benchmark strategies applied for NASDAQ index.

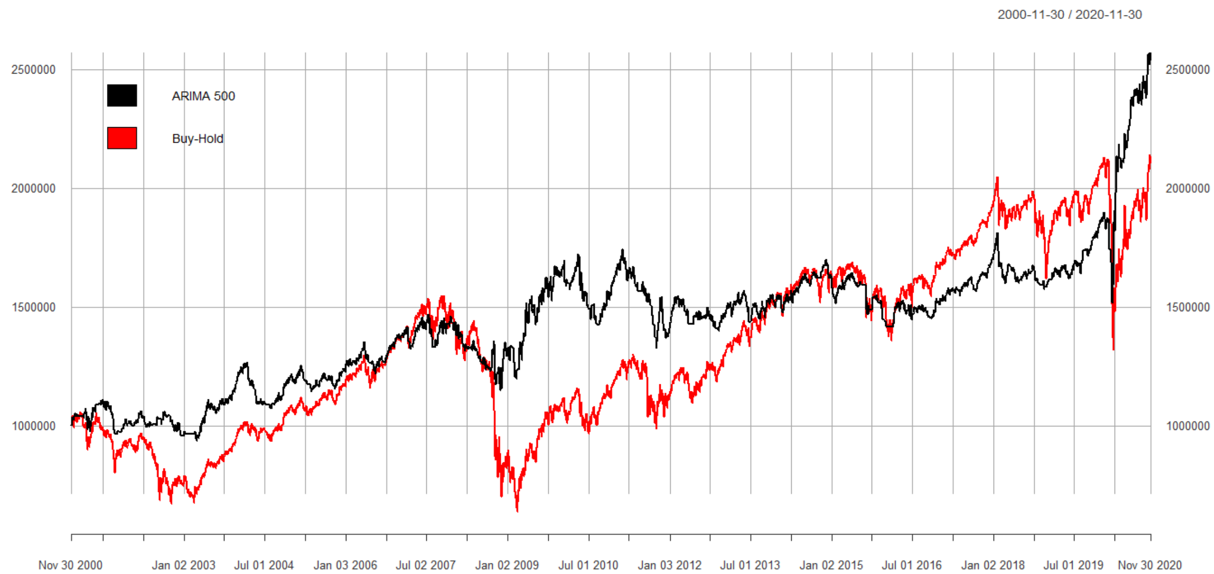
Name	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
ARIMA	6.29%	**15.07%**	0.418	**43.75%**	0.144	0.206	**6.59**	1333
BUYHOLD	**10.63%**	24.01%	**0.443**	63.05%	**0.169**	**0.310**	15.14	1

Note: ARIMA indicates the performance statistics for ARIMA strategy applied for NASDAQ index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for NASDAQ index. Bolded font indicates the best value with regards to each performance measure.

4.1.3. ARIMA strategy applied for NYSE index

The results for the ARIMA model for NYSE index values were presented in Figure 4. ARIMA equity line for NYSE index was less volatile than the equity lines for Buy&Hold strategy. It also reached its peak at the end of the trading period. The NYSE equity line went downward in the years 2001-2003 and that decline reached its final low in 2003 when the US war with Iraq started. In the years 2008-2012 there can be spotted dramatic fluctuations which also cover the period of the financial crisis of 2008. NYSE index equity line also increased sharply at the beginning of 2018. A considerable decrease can also be spotted in Feb 2020 when the Coronavirus Crash began. After that period there can be spotted a relatively stable increase in the equity line till the end of the trading period. In general, the ARIMA model had a good performance during the highly volatile period and the periods of big market crashes and downturns.

Figure 4. ARIMA and the benchmark strategies applied for NYSE index.



Note: *ARIMA 500* indicates the equity line for ARIMA strategy applied for NYSE index by using 500 training days and identifying parameters using Akaike information criterion. *Buy-Hold* indicates equity line for Buy and Hold strategy applied for NYSE index.

Performance and risk metrics were presented in Table 4. It shows that the ARIMA based strategy outperformed the benchmark strategy with the higher IR compared to IR obtained by the benchmark. ARIMA had a lower annualized standard deviation (ASD) and Maximum

Drawdown (MD) than the benchmark ASD and MD which resulted in the ARIMA based portfolio having a lower AllRisk metric than the benchmark portfolio. Moreover, the benchmark strategy had a lower Sharpe Ratio (SR) compared to the ARIMA based strategy which means the ARIMA had a better risk-adjusted performance.

Table 4. Performance statistics for ARIMA and the benchmark strategies applied for NYSE index.

Name	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
ARIMA	5.49%	**12.92%**	**0.425**	**23.82%**	**0.231**	**0.179**	**3.08**	1405
BUYHOLD	**5.70%**	19.86%	0.287	59.01%	0.097	0.127	11.72	1

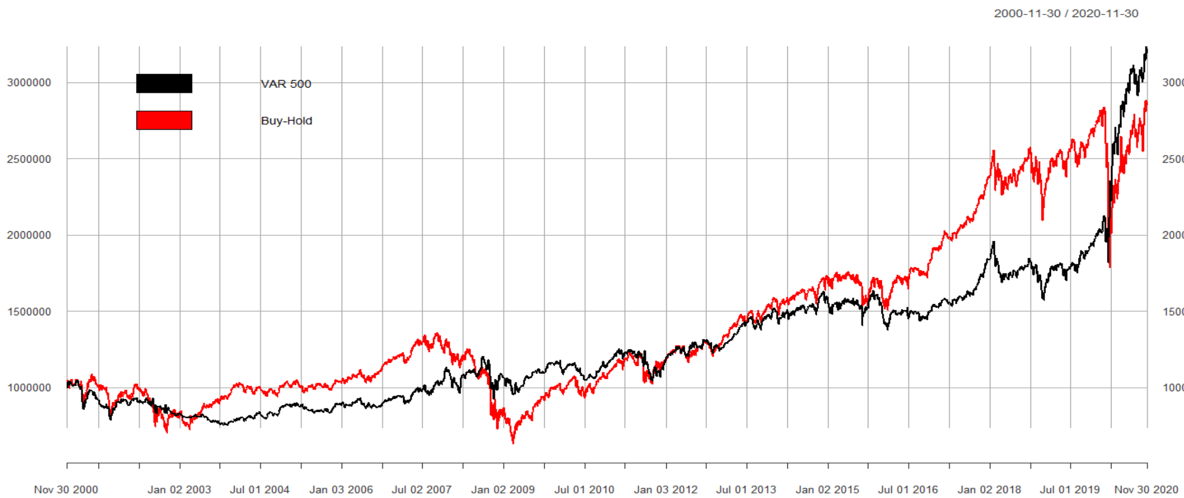
Note: ARIMA indicates the performance statistics for ARIMA strategy applied for NYSE index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for NYSE index. Bolded font indicates the best value with regards to each performance measure.

4.2. VAR results

4.2.1. VAR strategy applied on DJIA index values

The results for the VAR model by using DJIA index values were presented in Figure 5. VAR equity line has volatile fluctuations, however a general increasing trend was observed and reached its peak at the end of the trading period. In general, VAR performed well during the highly volatile periods and during the periods of big market crashes and downturns.

Figure 5. VAR and the benchmark strategies applied for DJIA index.



Note: VAR 500 indicates the equity line for VAR strategy applied for DJIA index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates equity line for Buy and Hold strategy applied for DJIA index.

Performance and risk metrics were presented in Table 5. We can see that VAR based strategy performed better than the benchmark strategy with the higher Information Ratio (IR) compared to IR obtained by the benchmark. VAR had a lower annualized standard deviation (ASD) and Maximum Drawdown (MD) than the benchmark which resulted in the VAR based portfolio having a lower AllRisk measure. Moreover, the benchmark strategy had a lower Sharpe

Ratio (SR) compared to the VAR based strategy which means VAR had a better risk-adjusted performance.

Table 5. Performance statistics for VAR and benchmark strategies applied for DJIA index.

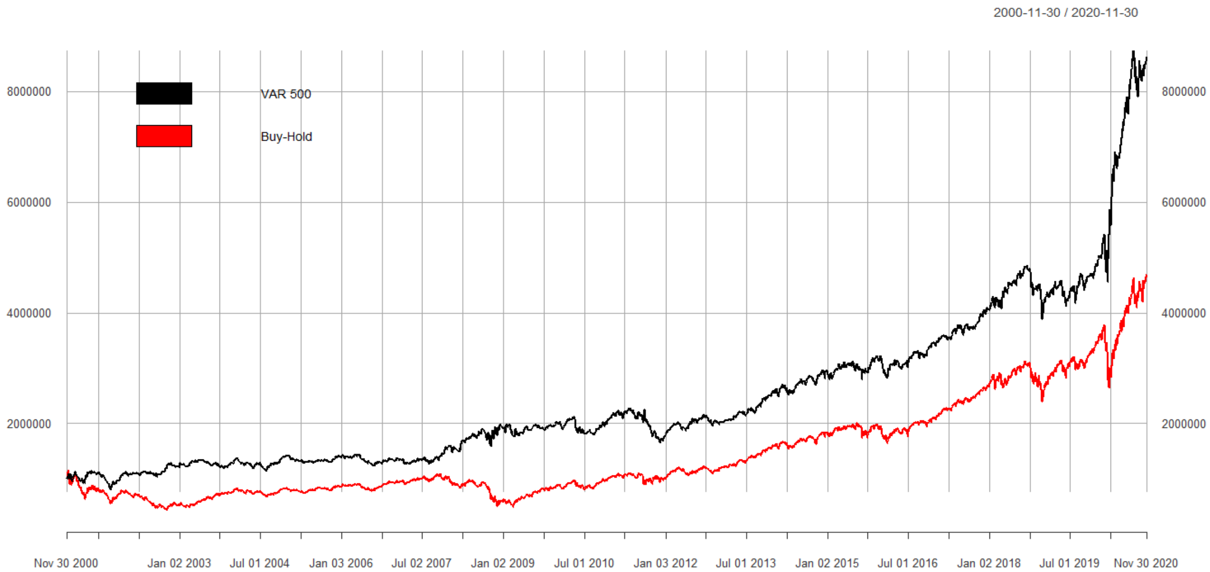
Name	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
VAR	6.76%	**13.82%**	**0.489**	**28.51%**	**0.237**	**0.259**	**3.94**	2085
BUYHOLD	**7.07%**	19.12%	0.370	53.78%	0.131	0.203	10.28	1

Note: VAR indicates performance statistics for VAR strategy applied for DJIA index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for DJIA index. Bolded font indicates the best value with regards to each performance measure.

4.2.2. VAR strategy applied on NASDAQ index values

The results of the VAR model for NASDAQ index were presented in Figure 6. VAR equity line for NASDAQ also showed a strong increasing trend and reached its peak at the end of the trading period.

Figure 6. Equity lines for VAR and the benchmark strategies applied for NASDAQ index.



Note: VAR 500 indicates the equity line for VAR strategy applied for NASDAQ index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates equity line for Buy and Hold strategy applied for NASDAQ index.

Performance and risk metrics were presented in Table 6. It shows that the VAR-based strategy performed better than the benchmark strategy obtaining the higher Information Ratio (IR). VAR had a lower annualized standard deviation (ASD) and Maximum Drawdown (MD) than the benchmark which resulted in the VAR based portfolio having a lower risk than the benchmark portfolio. Moreover, the benchmark strategy had a lower Sharpe Ratio (SR) compared to the VAR based strategy which means VAR had a better risk-adjusted performance.

Table 6. Performance statistics for VAR and benchmark strategies applied for NASDAQ index.

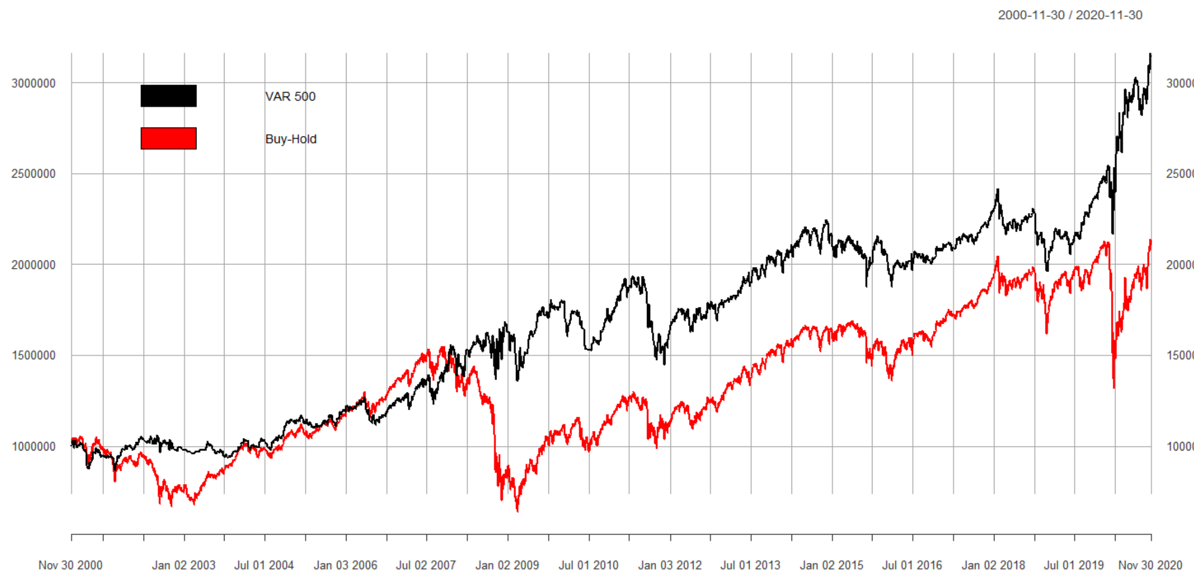
Names	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
VAR	**12.25%**	**17.14%**	**0.715**	**29.57%**	**0.414**	**0.529**	**5.07**	1923
BUYHOLD	10.63%	24.01%	0.443	63.05%	0.169	0.310	15.14	1

Note: VAR indicates the performance statistics for VAR strategy applied for NASDAQ index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for NASDAQ index. Bolded font indicates the best value with regards to each performance measure.

4.2.3. VAR strategy applied for NYSE index values

The results of the VAR model for NYSE index were presented in Figure 7. VAR equity line for NYSE index was less volatile than Buy&Hold strategy, however, a general increasing trend was observed and reached its peak at the end of the trading period.

Figure 7. VAR and the benchmark strategies applied for NYSE index



Note: VAR 500 indicates equity line for VAR strategy applied for NYSE index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates equity line for Buy and Hold strategy applied for NYSE index.

Performance and risk metrics were presented in Table 7. It shows that the VAR based strategy performed better than the benchmark strategy obtaining the higher Information Ratio (IR) compared to IR obtained by the benchmark. VAR had a lower annualized standard deviation (ASD) and Maximum Drawdown (MD) than the benchmark's ASD and MD which resulted in the VAR based portfolio having a lower All Risk metrics than the benchmark portfolio. Moreover, the benchmark strategy had a lower Sharpe Ratio (SR) compared to the VAR based strategy which means VAR had a better risk-adjusted performance.

Table 7. Performance statistics for VAR and the benchmark strategies applied for NYSE index.

Name	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
VAR	**6.74%**	**14.14%**	**0.477**	**25.19%**	**0.268**	**0.252**	**3.56**	2067
BUYHOLD	5.70%	19.86%	0.287	59.01%	0.097	0.127	11.72	1

Note: VAR indicates the statistics for VAR strategy applied for NYSE index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for NYSE index. Bolded font indicates the best value with regards to each performance measure.

4.3. Comparison of ARIMA and VAR based investment strategies

In this section performance and risk measures for ARIMA and VAR based strategies were compared. Table 8 shows the performance statistics for ARIMA, VAR and benchmark strategies applied for DJIA index. The VAR model had the best performance due to the highest IR and SR. Although ASD was lower in case of ARIMA portfolio, VAR based portfolio had still the lowest AllRisk metric. Regarding the annualized return compounded, the BuyHold strategy had the highest ARC, while ARIMA had the lowest ARC value. Considering most of the measures the VAR model outperformed ARIMA and benchmark strategy.

Table 8. Performance statistics for ARIMA, VAR and the Benchmark strategies applied for DJIA index.

Names	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
ARIMA	5.15%	**12.74%**	0.404	31.38%	0.164	0.154	4.00	1465
VAR	6.76%	13.82%	**0.489**	**28.51%**	**0.237**	**0.259**	**3.94**	2085
BUYHOLD	**7.07%**	19.12%	0.370	53.78%	0.131	0.203	10.28	1

Note: VAR/ARIMA indicates the performance statistics for VAR/ARIMA strategy applied for DJIA index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for DJIA index. Bolded font indicates the best value with regards to each performance measure.

Table 9 shows the results obtained from ARIMA and VAR models for NASDAQ index. The VAR model had again the best risk-adjusted performance due to the highest IR and SR. VAR based portfolio had the lowest AllRisk metric although its ASD value was slightly higher than ASD for ARIMA strategy. Considering the reviewed measures, it can be concluded that the VAR model performed better than both ARIMA and the benchmark strategy.

Table 9. Performance statistics for ARIMA, VAR and the Benchmark strategies applied for NASDAQ index.

Name	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
ARIMA	6.29%	**15.07%**	0.418	43.75%	0.144	0.206	6.59	1333
VAR	**12.25%**	17.14%	**0.715**	**29.57%**	**0.414**	**0.529**	**5.07**	1923
BUYHOLD	10.63%	24.01%	0.443	63.05%	0.169	0.310	15.14	1

Note: VAR/ARIMA indicates the performance statistics for VAR/ARIMA strategy applied for NASDAQ index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for NASDAQ index. Bolded font indicates the best value with regards to each performance measure.

Table 10 shows the results obtained from ARIMA and VAR models for NYSE index. The VAR model had once again the highest risk-adjusted performance due to the highest IR and SR. ARIMA based portfolio had a lower AllRisk metric. ARIMA had the lowest annualized return compounded with the second best SR. Considering its best risk-adjusted performance, highest information ratio, and highest annualized return compounded, it can be concluded that the VAR model still outperformed the ARIMA and benchmark strategies.

Table 10. Performance statistics for ARIMA, VAR and the Benchmark strategies applied for NYSE

Name	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
ARIMA	5.49%	**12.92%**	0.425	**23.82%**	0.231	0.179	**3.08%**	1405
VAR	**6.74%**	14.14%	**0.477**	25.19%	**0.268**	**0.252**	3.56%	2067
BUYHOLD	5.70%	19.86%	0.287	59.01%	0.097	0.127	11.72%	1

Note: VAR/ARIMA indicates the performance statistics for VAR/ARIMA strategy applied for NYSE index by using 500 training days and identifying parameters using Akaike information criterion. Buy-Hold indicates performance statistics for Buy and Hold strategy applied for NYSE index. Bolded font indicates the best value with regards to each performance measure.

4.4. Comparison of ARIMA and VAR model forecast accuracy measures

The forecast error measures of ARIMA and VAR models for DJIA index were presented in Table 11. ARIMA model has lower RMSE, MAE and MAPE than VAR model. Therefore, ARIMA can be considered as a better model in terms of forecasting accuracy.

Table 11. Error Measures of ARIMA and VAR models forecasting DJIA index prices.

Names	RMSE	MAE	MAPE
ARIMA	**185.716**	**111.386**	**0.787**
VAR	190.068	112.331	0.789

Note: ARIMA/VAR indicates ARIMA/VAR error measures forecasting DJIA index prices by using 500 days rolling training window. Bolded font indicates the lowest value with regards to each error measure.

The forecast error measures of ARIMA and VAR models for NASDAQ index was presented in Table 12. ARIMA model had a lower RMSE, MAE, and MAPE than VAR model. Therefore, ARIMA can be considered as a better model in terms of forecasting accuracy.

Table 12. Error Measures of ARIMA and VAR models forecasting NASDAQ index prices.

Name	RMSE	MAE	MAPE
ARIMA	**62.634**	**35.544**	**1.029**
VAR	62.736	35.699	1.032

Note: ARIMA/VAR indicates ARIMA/VAR error measures forecasting NASDAQ index prices by using 500 days rolling training window. Bolded font indicates the lowest value with regards to each error measure.

The forecast error measures of ARIMA and VAR models for NYSE index was presented in Table 13. ARIMA model has a lower RMSE, MAE than VAR model. They have equal MAPE metrics. Therefore, ARIMA can be considered as a better model in terms of forecasting accuracy.

Table 13. Error Measures of ARIMA and VAR models forecasting NYSE index prices.

Name	RMSE	MAE	MAPE
ARIMA	**103.863**	**68.039**	0.819
VAR	104.826	68.207	0.819

Note: ARIMA/VAR indicates ARIMA/VAR error measures forecasting NASDAQ index prices by using 500 days rolling training window. Bolded font indicates the lowest value with regards to each error measure.

5. Sensitivity analysis

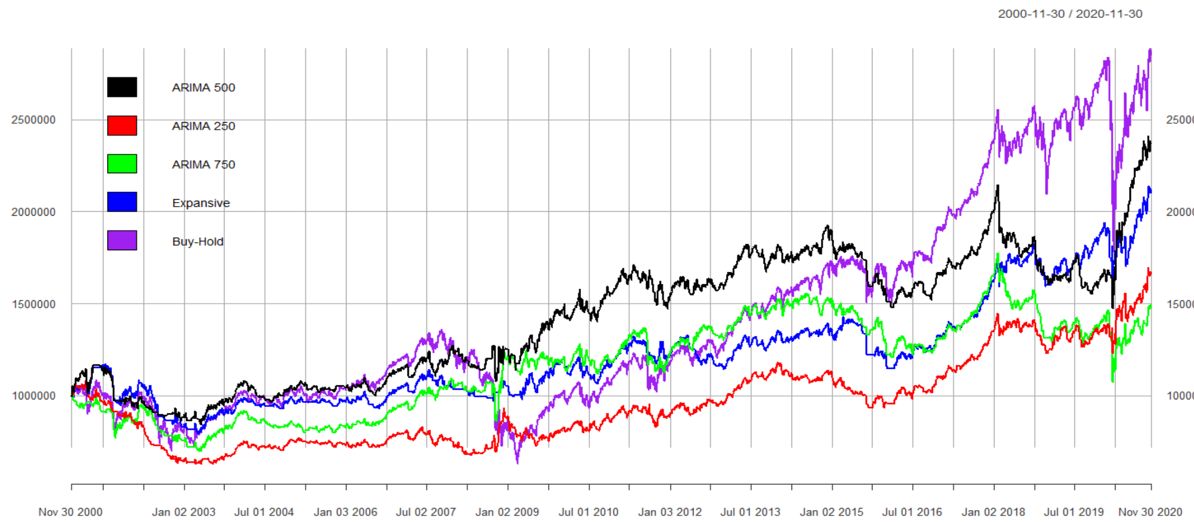
To check the robustness of the models the parameters should be changed to see how the results would be affected by the changes. In this section the results were computed for additionally selected 250 days rolling window, 750 days rolling window and 500 days rolling window with the fixed starting point (fixed starting point means that each new training window increases by one more step with the fixed starting point). Then the results were compared to the results obtained by using the initially set parameters (500 days rolling window). For each additionally selected new parameter the other default parameters that were described in the methodology section did not change.

5.1. Sensitivity analysis results for VAR and ARIMA based investment strategies

5.1.1. Sensitivity analysis results for VAR and ARIMA based strategies applied for DJIA index

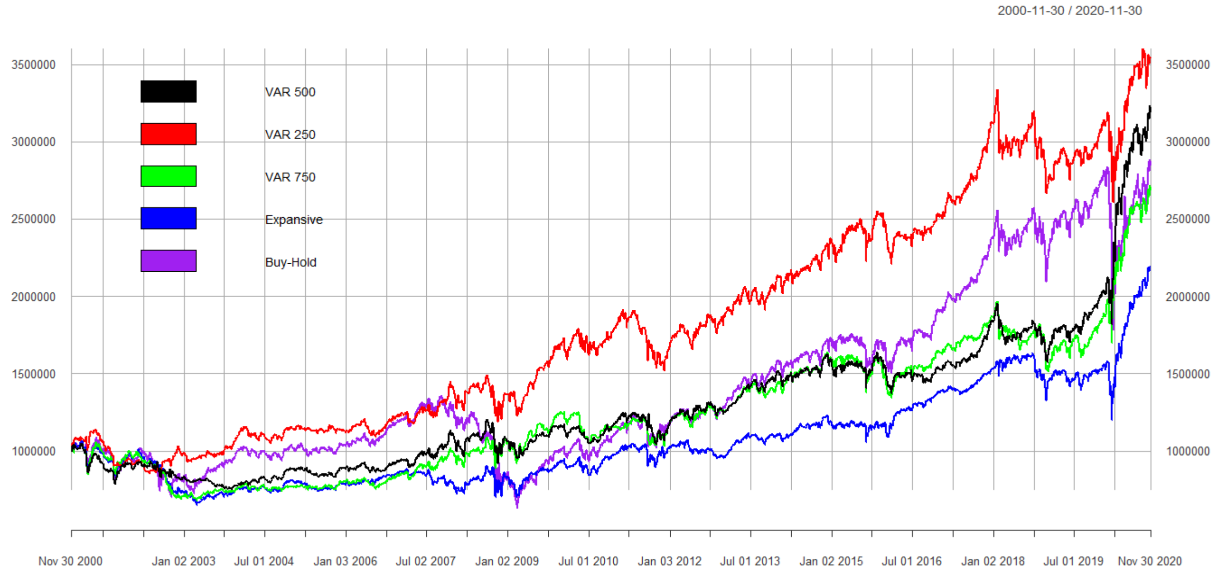
Figure 8, 9 and Table 14 present the sensitivity analysis for ARIMA and VAR based strategies applied for DJIA index. An increase and decrease of the ARIMA training window affected its performance statistics negatively. The VAR model performed better when the training window size decreased. The VAR model was less robust to the changes with a higher standard deviation of information ratios (8.85%) compared to the ARIMA model (8.72%). Compared to the benchmark strategy the VAR model can be considered as better because of the mainly higher IR and lower risk measures.

Figure 8. Sensitivity analysis for ARIMA based strategy applied for DJIA index.



Note: Sensitivity analysis for ARIMA strategy applied for DJIA index. Changes were made in the parameters as follows: 250 training window days, 750 training window days, fixed beginning point for training window.

Figure 9. Sensitivity analysis for VAR based strategy applied for DJIA index.



Note: Sensitivity analysis for VAR strategy applied for DJIA index. Changes were made in parameters as follows: 250 training window days, 750 training window days, fixed beginning point for training window.

Table 14. Performance statistics for all ARIMA and VAR based strategies applied for DJIA index

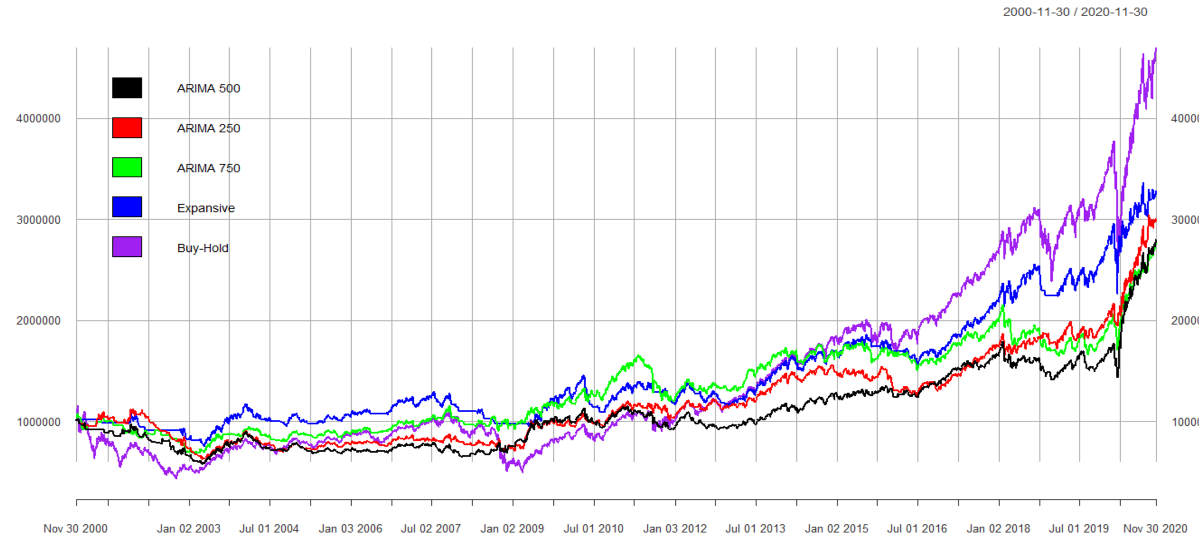
Names	ARC	ASD	IR	MD	ARCMD	SR	AllRisk	Trades
ARIMA	5.15%	12.74%	0.404	31.38%	0.164	0.154	4.00	1465
ARIMA 250	3.24%	11.99%	0.271	42.57%	0.076	0.005	5.10	1553
ARIMA 750	2.82%	13.06%	0.216	39.36%	0.072	-0.028	5.14	1427
Expansive ARIMA	4.43%	**11.94%**	0.371	32.46%	0.136	0.104	3.87	739
VAR	6.76%	13.82%	0.489	28.51%	0.237	0.259	3.94	2085
VAR 250	**7.22%**	13.21%	**0.546**	**24.82%**	**0.291**	**0.305**	**3.28**	1975
VAR 750	5.96%	14.29%	0.417	35.93%	0.166	0.194	5.14	2127
Expansive VAR	5.00%	14.60%	0.342	39.38%	0.127	0.124	5.75	2431
BUYHOLD	7.07%	19.12%	0.370	53.78%	0.131	0.203	10.28	1

Note: Performance statistics for all ARIMA and VAR strategies applied for DJIA index. Changes were made in the parameters as follows: 250 training window days, 750 training window days, fixed beginning point for training window. Bolded font indicates the best value with regards to each performance measure.

5.1.2. Sensitivity analysis results for VAR and ARIMA based strategies applied for NASDAQ index

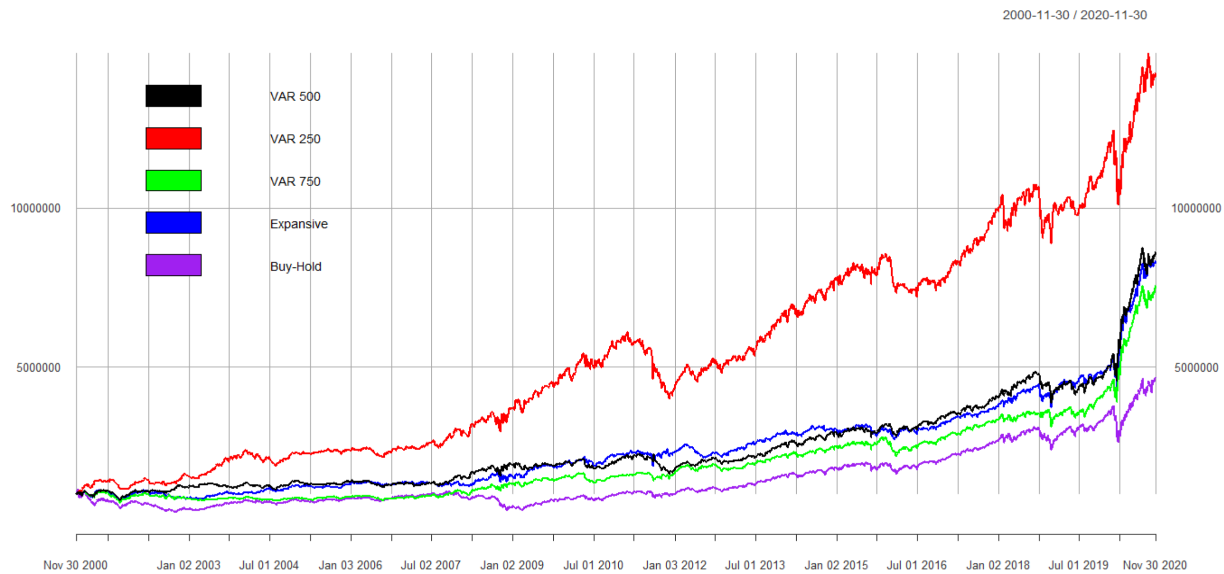
Figure 10, 11 and Table 16 presents the sensitivity analysis for ARIMA and VAR based strategy applied for NASDAQ index. ARIMA strategy performance statistics indicates that ARIMA based portfolio had stable results with regards to changing trading window, while the VAR model was less robust to this parameter. VAR based strategy had the highest Information ratios (IR), the highest annualized returns compounded (ARC) and the highest Sharpe Ratios (SR). The VAR model was also less robust to the changes with the higher standard deviation of information ratios (10.31%) compared to the ARIMA model (3.37%). Comparing above mentioned risk and performance metrics, the VAR based strategy can be considered to be better than the ARIMA and the benchmark strategies.

Figure 10. Sensitivity analysis for ARIMA based strategy applied for NASDAQ index



Note: Sensitivity analysis for ARIMA strategy applied for NASDAQ index. Changes were made in parameters as follows: 250 training window days, 750 training window days, fixed beginning point for training window.

Figure 11. Sensitivity analysis for VAR based strategy applied for NASDAQ index.



Note: Sensitivity analysis for VAR strategy for NASDAQ index. Changes were made in parameters as follows: 250 training window days, 750 training window days and fixed beginning point for training window.

Table 16. Performance statistics for all ARIMA and VAR based strategies applied for NASDAQ index.

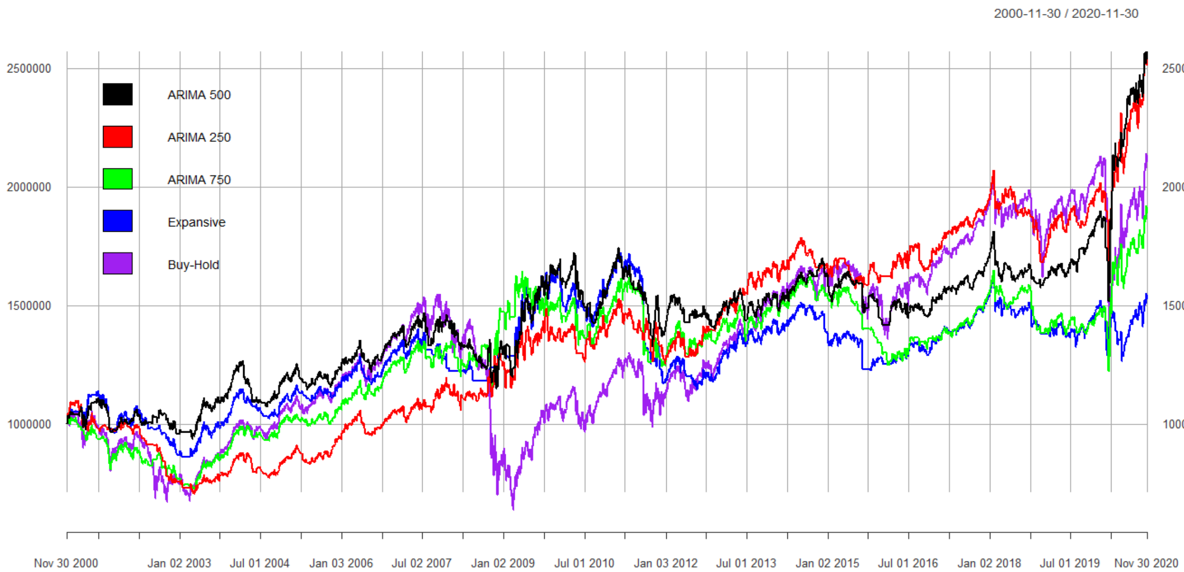
Name	ARC	ASD	IR	MD	ARC MD	SR	All Risk	Trades
ARIMA 500	6.29%	15.07%	0.418	43.75%	0.144	0.206	6.59	1333
ARIMA 250	6.54%	14.26%	0.459	44.39%	0.147	0.236	6.33	1377
ARIMA 750	6.18%	14.54%	0.425	36.62%	0.169	0.206	5.32	1269
Expansive ARIMA	6.95%	**14.15%**	0.491	29.76%	0.234	0.266	**4.21**	267
VAR 500	12.25%	17.14%	0.715	**29.57%**	0.414	0.529	5.07	1923
VAR 250	**14.67%**	16.51%	**0.889**	34.44%	**0.426**	**0.696**	5.69	1877
VAR 750	11.70%	17.72%	0.660	34.71%	0.337	0.481	6.15	1993
Expansive VAR	12.20%	17.74%	0.688	30.77%	0.397	0.508	5.46	2057
BUYHOLD	10.63%	24.01%	0.443	63.05%	0.169	0.310	15.14	1

Note: Performance statistics for all ARIMA and VAR strategies applied to NASDAQ. Changes were made in parameters as follows: 250 training window days, 750 training window days and training window with fixed starting point. Bolded font indicates the best value with regards to each performance measure.

5.1.3. Sensitivity analysis results for VAR and ARIMA based strategies applied for NYSE index

Figure 12, 13 and Table 17 present the sensitivity analysis for ARIMA and VAR based strategies applied for NYSE index.

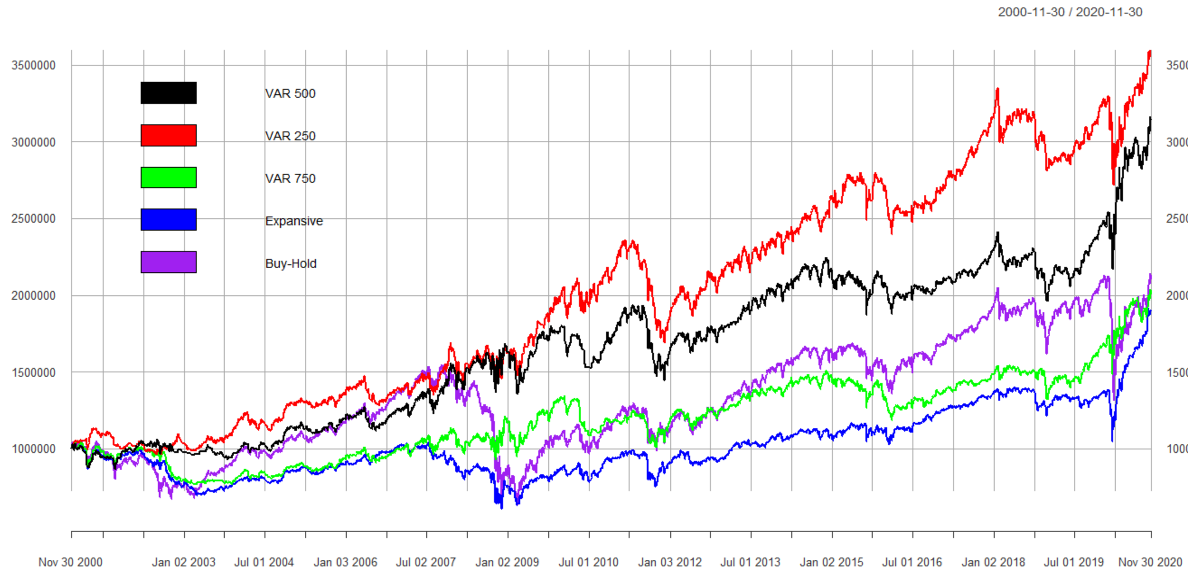
Figure 12. Sensitivity analysis for ARIMA based strategy applied for NYSE index.



Note: Sensitivity analysis for ARIMA strategy for NYSE index. Changes were made in parameters as follows: 250 training window days, 750 training window days, fixed beginning point for training window.

Changes in the parameters resulted in the increase of the AllRisk metrics of both ARIMA and VAR based strategies. ARIMA based strategy had lower AllRisk metrics for Expansive training window and 500 days training window. However, the VAR model had higher annualized return compounded (ARC), higher information ratios (IR) and higher sharpe ratios (SR) than the ARIMA based strategies. The VAR model was also less robust to the changes with the higher standard deviation of information ratios (12.05%) compared to the ARIMA model (9.45%). However, based on the results from Table 17 we can conclude that the VAR model outperformed the ARIMA model. Compared to the benchmark strategy the VAR model can also be considered as better because of the mainly higher IR, Sharpe Ratio, and mainly lower risk measures.

Figure 13. Sensitivity analysis for VAR based strategy applied for NYSE index.



Note: Sensitivity analysis for VAR strategy for NYSE index. Changes were made in parameters as follows: 250 training window days, 750 training window days and stable beginning point for training window.

Table 17. Performance statistics for all ARIMA and VAR based strategies applied for NYSE index

Name	ARC	ASD	IR	MD	ARC MD	SR	All Risk	Trades
ARIMA 500	5.49%	12.92%	0.425	**23.82%**	0.231	0.179	**3.08**	1405
ARIMA 250	5.39%	12.52%	0.431	35.92%	0.150	0.177	4.50	1515
ARIMA 750	4.05%	13.20%	0.307	29.93%	0.135	0.066	3.95	1325
Expansive ARIMA	2.81%	**11.85%**	0.237	33.92%	0.083	-0.031	4.02	835
VAR 500	6.74%	14.14%	0.477	25.19%	**0.268**	0.252	3.56	2067
VAR 250	**7.36%**	13.77%	**0.534**	28.54%	0.258	**0.303**	3.93	2021
VAR 750	4.52%	14.42%	0.314	26.59%	0.170	0.093	3.83	2155
Expansive VAR	4.36%	15.09%	0.289	41.99%	0.104	0.078	6.34	2391
BUYHOLD	5.70%	19.86%	0.287	59.01%	0.097	0.127	11.72	1

Note: Performance statistics for ARIMA and VAR strategies for NYSE. Changes were made in the parameters as follows: 250 training window days, 750 training window days, stable beginning point for training window. Bolded font indicates the best value with regards to each performance measure.

5.2. Sensitivity analysis results for VAR and ARIMA model forecast accuracy measures

Table 18 shows the sensitivity analysis results for the forecast error metrics of ARIMA and VAR models estimated on DJIA index. For 500 days and 250 days rolling windows, ARIMA forecasts had mainly lower errors than VAR model. For 750 days rolling window and the extensive window the VAR model had lower errors than ARIMA model.

Table 18. Error measures for all ARIMA and VAR models estimated on DJIA index prices.

Names	RMSE	MAE	MAPE
ARIMA	**185.716**	**111.386**	**0.787**
ARIMA 250	**188.176**	**111.8**	0.792
ARIMA 750	189.569	112.261	0.791
ARIMA Expansive	185.215	110.313	0.781
VAR	190.068	112.331	0.789
VAR 250	192.474	112.094	**0.789**
VAR 750	**187.559**	**111.217**	**0.785**
VAR Expansive	**184.305**	**110.104**	**0.78**

Note: Error Measures were calculated for all ARIMA and VAR models used in this paper forecasting DJIA index prices. Changes were made in the parameters as follows: 250 training window days, 750 training window days, stable beginning point for training window. Bolded font indicates the lowest values with regards to each error measure (the results were compared as per corresponding training windows).

Table 19 shows the sensitivity analysis results for forecast error metrics of ARIMA and VAR models estimated on NASDAQ index. For 500 days and 250 days rolling training windows ARIMA forecasts had lower errors than VAR model. For 750 days rolling training window and the extensive training window, the VAR model had mainly lower errors than ARIMA model.

Table 19. Error measures for ARIMA and VAR models forecasting NASDAQ index prices.

Name	RMSE	MAE	MAPE
ARIMA	**62.634**	**35.544**	**1.029**
ARIMA 250	**62.945**	**35.622**	**1.032**
ARIMA 750	62.444	35.675	1.045
ARIMA Expansive	63.19	35.351	**1.02**
VAR	62.736	35.699	1.032
VAR 250	63.61	35.739	1.033
VAR 750	**61.788**	**35.424**	**1.03**
VAR Expansive	**61.242**	**35.162**	1.023

Note: Error Measures were calculated for all ARIMA and VAR models estimated on NASDAQ index prices. Changes were made in the parameters as follows: 250 training window days, 750 training window days, stable beginning point for training window. Bolded font indicates the lowest values with regards to each error measure (the results were compared as per corresponding training windows).

Table 20 shows the sensitivity analysis results for the forecast error measures of ARIMA and VAR models estimated on NYSE index. For 500 days and 250 days rolling windows ARIMA forecasts had mainly lower error measures than VAR model. For 750 days window and Extensive window, the VAR model had mainly lower errors than ARIMA model.

Table 20. Error Measures for ARIMA and VAR models forecasting NYSE index prices.

Name	RMSE	MAE	MAPE
ARIMA 500	**103.863**	**68.039**	0.819
ARIMA 250	**104.979**	68.374	0.824
ARIMA 750	**103.466**	67.935	0.819
ARIMA Expansive	103.022	67.374	0.812
VAR 500	104.826	68.207	0.819
VAR 250	105.756	**68.253**	**0.82**
VAR 750	103.708	**67.703**	**0.815**
VAR Expansive	**102.603**	**67.148**	**0.809**

Note: Error Measures for all ARIMA and VAR models forecasting NASDAQ index prices. Changes were made in the parameters as follows: 250 training window days, 750 training window days, stable beginning point for training window. Bolded font indicates the lowest values with regards to each error measure (the results were compared as per corresponding training windows).

Conclusions

Algorithmic Trading is used in stock, commodity, and many other markets with the help of computers by using various computer software and investment systems. The most common problem people face when trading is being influenced by human feelings in their decision making. By nature, emotions can distract people from rational decision-making. Algorithmic trading enables people to make rational buy-sell decisions instead of emotional buy-sell decisions.

The use of algorithms in the markets began with the introduction of computers into our lives. The bankers and other fund managers who manage billions of dollars can face losses due to the emotional decisions that are always a risk factor. The fact that computers can perform many operations faster than humans, compare many strategies at the same time, and can choose the most suitable one, makes algorithms much more advantageous compared to human trading.

This paper showed the procedure of preparing algorithmic trading strategies by using three US stock market indices - DJIA index, NASDAQ index, and NYSE index. The trading period covers the last 20 years. Applied investment strategies used two statistical time series models – Autoregressive Integrated Moving Average (ARIMA) and Vector autoregression (VAR) models. Performance statistics (Information Ratio, Annualized Standard Deviation, Maximum Drawdown, Annualized Compounded Return, Sharpe Ratio) and Forecast Error Measures (RMSE, MAE, MAPE) were computed by using the results obtained from each method. Sensitivity analysis was also conducted for each method to check their robustness to the changes in parameters.

In this paper, Buy and Hold strategy was used as a benchmark. VAR based strategy mainly outperformed the Buy and Hold strategy and ARIMA strategy in terms of its lower portfolio risk and higher performance measures.

Three hypotheses were stated at the beginning of this paper. The first one was: *ARIMA and VAR models have similar forecasting power*. The second one stated: *VAR models should be more robust to the changes than ARIMA models*. The third one: *The model with more accurate forecasts might not perform better when applied to an algorithmic investment strategy*. Based on the delivered results, the first hypothesis is rejected as ARIMA model had lower forecasting errors than VAR model. The second hypothesis can also be rejected as the results showed that the VAR model was less robust to the changes and obtained a higher standard deviation of information ratios compared to the ARIMA model. Although ARIMA performed remarkably well during volatile periods, VAR based strategies mainly outperformed ARIMA based strategies in terms of its lower portfolio risk and higher risk-adjusted return measures. Regarding the third hypothesis, the obtained results seem to be consistent with this hypothesis. ARIMA had lower forecasting errors while the performance statistics showed that VAR based investment strategies outperformed ARIMA based investment strategies. It can be concluded that we failed to reject the third hypothesis, meaning that there is not enough evidence to reject the statement that the error metrics may not be a reliable measure to evaluate the performances of models.

Similar research as in this paper can be applied to the other statistical and machine learning models. Different data, signal generating methods can be implemented and different parameters can be changed to check the robustness of the models. So far, the research papers focusing on the forecasting model comparisons have paid more attention to the forecast accuracy of the models. However, comparing the investment performances of forecasting models can create a clearer picture of the models.

References:

Akaike, H., 1969. Fitting autoregressive models for prediction. Ann. Inst. Stat. Math., 21, 243-247,

Anggraenia W., Andri K., Sumaryanto, Mahanantoa F., 2017. The Performance of ARIMAX Model and Vector Autoregressive (VAR) Model in Forecasting Strategic Commodity Price in Indonesia. Procedia Computer Science. Volume 124. pp. 189-196,

Bagshaw M.L., 1987. Univariate and multivariate ARIMA versus vector autoregression forecasting. Working Papers (Old Series) 8706, Federal Reserve Bank of Cleveland, revised 1987,

Bilyk O., Sakowski P., Ślepaczuk R., 2020. Investing in VIX futures based on rolling GARCH models forecasts, Working Papers of Faculty of Economic Sciences, University of Warsaw, WP 10/2020 (316).

Boehmer E., Kingsley F., Wu J., 2020. Algorithmic Trading and Market Quality: International Evidence. Available at SSRN: https://ssrn.com/abstract=2022034 or http://dx.doi.org/10.2139/ssrn.2022034,

Botchkarev A., 2018. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology,

Box, G. E., & Jenkins, G. M., 1970, Time series analysis: Forecasting and control HoldenDay. San Francisco, p. 498,

Cesari R., Zheng H., 2020. Optimal Liquidation in a Mean-reverting Portfolio. Available at: https://www.researchgate.net/publication/344505583_Optimal_Liquidation_in_a_Mean reverting_Portfolio,

Conrad, J. Sunil W., Jin X. 2015. High-Frequency Quoting, Trading, and the Efficiency of Prices. Journal of Financial Economics 116 (2): 271–91, Devi B., Sundar D., Alli P., 2013, An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50, International Journal of Data Mining & Knowledge Management Process (IJDKP), 3, 1, pp. 65-78,

Dickey, D. A., Fuller, W. A., 1979. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*. **74** (366): 427–431. doi:10.1080/01621459.1979.10482531. JSTOR 2286348,

Espasa, A., Poncela, P. and Senra, E. 2002. Forecasting monthly US consumer price indexes through a disaggregated I(2) Analysis, Statistics and Economic Series 1, Universidad Carlos III de Madrid working paper 02-03,

Gsell M., 2008. Assessing the Impact of Algorithmic Trading on Markets: A Simulation Approach. Available at SSRN: https://ssrn.com/abstract=1134834,

Hyndman, R., Koehler, A., 2005. Another look at measures of forecast accuracy. International Journal of Forecasting. 22 (4): pp. 679 – 688,

Ikechukwu K., and Adedoyin S., 2014. Time Series Modeling and Forecasting Inflation: Evidence from Nigeria. The International Journal of Business and Finance Research, Vol. 8 (2) pp. 41-51, 2014, Available at SSRN: https://ssrn.com/abstract=2322918,

Javed I., 2001. Forecasting methods: a comparative analysis. Published in: Proceedings Eighth Statistics Seminar, Vol. 8, (July 2001): pp. 189-197,

Johansen, S., 1991. Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. Econometrica. Vol. 59, No. 6 (Nov., 1991), pp. 1551-1580,

Kijewski M., Ślepaczuk R., 2020. Predicting prices of S&P 500 index using classical methods and recurrent neural networks, Working Papers of Faculty of Economic Sciences, University of Warsaw, WP 27/2020 (333).

Kirstin H., 2003. Forecasting Euro area inflation: Does aggregating forecasts by HICP component improve forecast accuracy?. International journal of Forecasting, Vol. 21(1), p. 119-136,

Kumar M., 2010. A Time-Varying Parameter Vector Autoregression Model for Forecasting Emerging Market Exchange Rates. International Journal of Business and Economic Sciences Applied Research (IJBESAR), International Hellenic University (IHU), Kavala Campus, Greece (formerly Eastern Macedonia and Thrace Institute of Technology - EMaTTech), vol. 3(2), pages 21-39, December,

Ryś P., Ślepaczuk R., 2019, *Machine learning in algorithmic trading strategy optimization – design and time efficiency*, Central European Economic Journal, 5(1), pp. 206-229, https://doi.org/10.1515/ceej-2018-0021

Sims, A., 1980. Macroeconomics and reality. Econometrica, 48(1):1–48,

Sporer, J., 2020. Backtesting of Algorithmic Cryptocurrency Trading Strategies. Available at SSRN: https://ssrn.com/abstract=3620154 or http://dx.doi.org/10.2139/ssrn.3620154,

Stephan, A., 2016. The Effect of Algorithmic Trading on Voluntary disclosure. Available at SSRN: https://ssrn.com/abstract=2836558 or http://dx.doi.org/10.2139/ssrn.2836558,

Thomakos, Dimitrios D., and John B. Guerard Jr.2020. Naive, ARIMA, Nonparametric, Transfer Function and VAR Models: A Comparison of Forecasting Performance. *International Journal of Forecasting*, 2004, vol 20,

Verheggen R., 2017. The rise of Algorithmic Trading and its effects on Return Dispersion and Market Predictability Rick Verheggen, 20, issue 1, 53-67,

Vo, N., and Ślepaczuk, R. (2022). Applying hybrid ARIMA-SGARCH in algorithmic investment strategies on S&P500 Index. Entropy, 24(2), 158. doi: https://doi.org/10.3390/e24020158

Whittle, P., 1951. Hypothesis Testing in Time Series Analysis. Almquist and Wicksell. Whittle, P. (1963). Prediction and Regulation. English Universities Press. *ISBN 0-8166-1147-5*,

Zhang H., 2013. Modeling and forecasting regional GDP in Sweden using autoregressive models Author: Haonan Zhang. Business Intelligence Program School for Technology and Business Studies Dalarna University.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl