Comparative study between univariate and multivariate statistical time series models for stock price forecasting

Irene Simó Muñoz
irene.munoz @ estudiantat.upc.edu

April 2023

Abstract

Time series analysis and forecasting via statistical methods have been widely explored. In this work, two of the most famous methodologies are contrasted for the prediction of stock prices; Autoregressive Integrated Moving Average (ARIMA) and Vector Autoregression (VAR). This study aims to analyze whether a univariate model (ARIMA) or a multivariate (VAR) one is more suited for stock market predictions and to relate the accuracies to the fundamentals of both methodologies.

1 Introduction

Time series analysis and forecasting have evolved from statistical methods to machine learning ones in recent years. However, the latter don't always perform as well as the former; statistical methods still make better predictions under certain conditions. (Cerqueira, Torgo, and Soares 2019). Among statistical methods there also are notable differences; one of them being whether the time series to analyze is univariate or multivariate.

The goal of this study is to build a basic forecast of stock prices using both a univariate stationary time series model and a multivariate stationary time series model and contrast the obtained results. The proposed methods are Autoregressive Integrated Moving Average (ARIMA) (Box et al. 2015) for the former, and Vector Autoregressive (VAR) (Lütkepohl 2013) for the latter.
Although other comparative studies between these methodologies have been performed, most have focused on the performance contrast between ARIMA and VAR; Iwok and Okpe 2016, Sethi and Mittal 2020 or Venugopalan and Srinath 1998.
This study aims to pinpoint the critical differences that forecasts based on either ARIMA or VAR show for stock price analysis and prediction and relate them to the fundamentals of both methodologies. Other similar work has been carried out for tourism growth prediction (Du Preez and Witt 2003) or hospital patient movements (Lin 1989) among more, adapting the observations and discussion to the particularities of the dataset.

ARIMA is a commonly used linear stochastic process for stationary models in time series forecasting. The methodology is the union of the Auto-Regressive (AR), Integrated (I), and Moving Average (MA) processes. The Moving Average component takes into account the relationship between an observation and a residual error from a moving average model applied to lagged observations. The Integrated component represents the differencing of raw observations to enable the time series to become stationary. The Auto-Regressive component refers to a model that shows a changing variable that regresses on its own lagged values.
The methodology is quite popular for analyzing data typically studied with linear regressions, since ARIMA models are more flexible than other statistical models such as exponential smoothing or simple linear regression (Kinney Jr 1978).

The endogenous variables in the system are functions of the lagged values of every endogenous variable in the multivariate linear time series model known as vector autoregression (VAR). In the subject of macroeconomics, VARs are particularly well-liked (Vargas-Silva 2008, Ghysels 2016), and they are frequently employed in modeling macroeconomic shocks to the actual economy as well as in policy simulations and forecasts.
VARs are helpful tools for structural analysis in addition to forecasting since they can look at how things react to shocks. They can locate sources of variations that conventional univariate models are unable to.

The first sections 2, 3 and 4 are devoted to the case study and implementation aspects of the models. Details on dataset preparation and data characteristics as well as particular steps and engineering choices performed for both ARIMA and VAR implementations are discussed. In the results and discussion sections 5 and 6, a review is provided and more theoretical issues and aspects related to the intuition of ARIMA and VAR are discussed.

The tests to ensure some common checks for the model diagnostics are not the same for both methods; that is because of the object properties of the ARIMA and VAR models in R.

2 Data and experimental setup

The used data corresponds to the price history and trading volumes of the fifty stocks in the index NIFTY 50 from NSE (National Stock Exchange) India, extracted from Kaggle[1]. The data spans from 1st January 2000 to 30th April 2021. In particular, the file used is BAJAJFINSV, corresponding to Bajaj Finserv Ltd.; a financial services company based in Pune.
The data used for this study only takes from May 26th, 2008 to November 16th, 2019 - the last reported prices before first China's COVID-19 publicly confirmed case, (World Health Organization 2020) -. This truncation is done to avoid over-informing the predictors with pandemic data, which is not a pattern we desire to include in the analysis.

For ARIMA, the target variable to study is the Volume Weighted Average Price (VWAP) - traced in blue in Figure 1 -, a trading benchmark used by traders that gives the average price the stock has traded at throughout the day, based on both volume and price.
For VAR, The target variables to study are VWAP and the Open price - traced in red in Figure 1 -; which corresponds to the average price stock at opening time at NSE. Both values are highly correlated.

[1] https://www.kaggle.com/datasets/rohanrao/nifty50-stock-market-data?resource=download

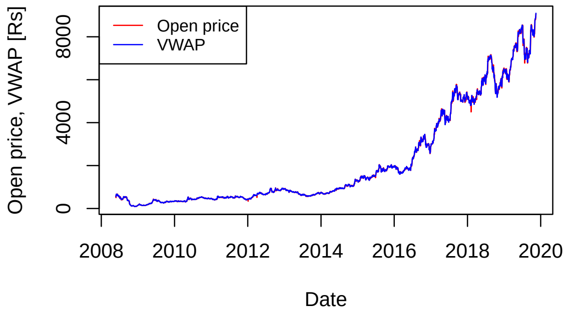

Figure 1: Time evolution for the open price and VWAP

2.1 Data preprocessing

The usability of the selected NSE dataset is qualified with the maximum ranking, and quick analysis has shown that no null nor faulty values are contained in the selected date range. Before experimental runs, two main data manipulations are done.

Daily to weekly data The given data is daily, Monday to Friday both included. To use it to forecast and obtain unbiased models, it is imperative to take into account the fact every five days there are two days of inactivity - the weekend -. As every piece of data is linked to an actual date, including all points means adding non-existent days, the weekends, for which there is no data, which implies an added difficulty for the model to predict those. Therefore, only Mondays have been selected and the present study has been carried out for weekly data on weekly forecasting.

Converting non-stationary to stationary data One of the main basic assumptions of both ARIMA and VAR (in a non-differentiated version) is that the data is stationary. To test if the selected time series can be considered stationary, Augmented Dickey-Fuller Test (ADF) is performed obtaining a p-value larger than the selected $\alpha = 0.05$, thus failing to reject the null hypothesis. This means the time series is non-stationary. In other words, it has a time-dependent structure and does not have constant variance over time.
Therefore, the data will require some manipulations to make it stationary. Both ARIMA and VAR differencing will be used, as detailed in their respective sections.

2.2 Experimental setup

All operations and analysis performed for this study are done in the R language[2], using RStudio 2022.02.1+461 "Prairie Trillium" Release.
For ARIMA and VAR-specific implementations, the functions provided by `tseries` (Time Series Analysis and Computational Finance), `tsDyn` (nonlinear dynamics for the analysis-modeling of observed time series), `aTSA` (Alternative Time Series Analysis) libraries, and in `forecasting` for forecasting using the built models are used.
ARIMA fitting model to a univariate time series is done via `arima` function in the `stats` package, and VAR fitting is done via `VAR` function in `vars` package.

As with the ADF test, all α values used for statistical tests in this work are set to 0.05.

3 ARIMA

The building of ARIMA forecast is composed by an expolaroty analysis of the data first, then the fitting of the model and a diagnostinc of the model, lastly. These phases aim to prepare the data and check the method's assumptions about the data before fitting the model, and then ensuring that the obtained result is valid for forecasting.

[2] `https://www.r-project.org`

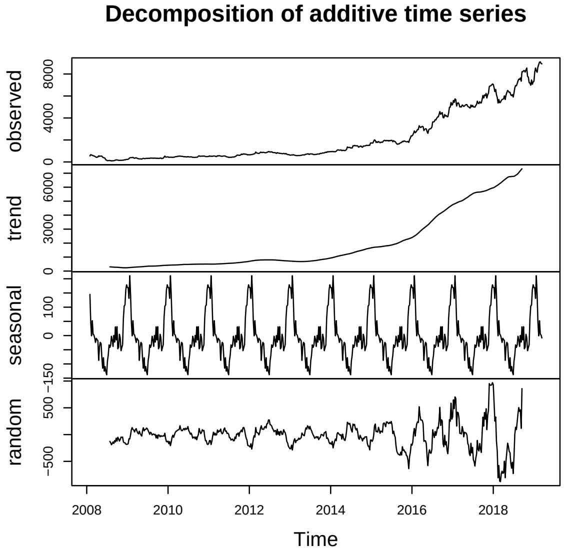

Figure 2: Components of VWAP time series

3.1 Exploratory analysis

The main assumptions that the input data must check for ARIMA are stationarity and non-seasonality. The former accounts for independency on the period in which the data is captured, and the second accounts for regular patterns ins the data through time. Already knowing that the data is non-stationary, the only needed tests to perform are in regards to the correlation and seasonality, which include decomposition as well.

Trend estimation and decomposition Although implementations of ARIMA in R are able to automatically estimate the parameters needed for the model, the data might not be able to clearly express its behavior. For this, Seasonal and Trend decomposition using Loess - Loess is a method for estimating nonlinear relationships - (STL) decomposition is performed. STL works as an additive process where the data is decomposed into trend, seasonality, and remainder and each component of the data can be taken apart for analysis.

Figure 2 shows the results of the STL decomposition. The provided graphs display the components separately; the *Observed* plot is the actual data, the *Trend* is the overall upward or downward movement of the data points, the *Seasonal* graph depicts any recurrent pattern of the data points, and *Random* one accounts for the unexplainable part of the data. Observing the graphs closely, we can find out if the data satisfies the seasonality assumption.

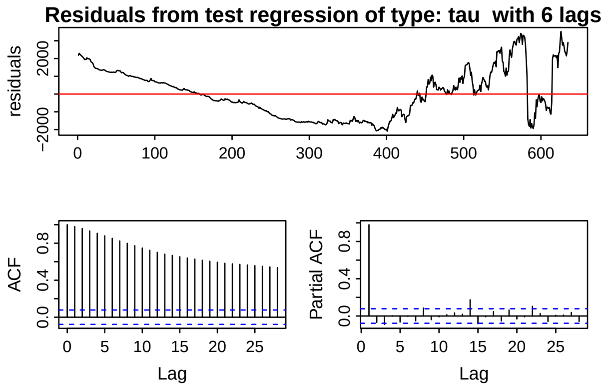

Figure 3: Plots of data before differentiating

The *Trend* and *Random* plots give information on two relevant characteristics of the data that are key for the forecasting discussion; firslty that the tendency for VWAP is to grow, and secondly that the data contains more noise in the recent periods of study, thus adding uncertainty in the model.

Making the data stationary In regards to the stationary assumption, in order to remove the non-stationary part for ARIMA we first run the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. This is a type of Unit root test and is used to find out the first difference or regression which should be used on the trending data to make it stationary. In Figure 3 we see the residuals, autocorrelation, and partial autocorrelation before differentiating. As suggested by the p-value obtained in the KPSS test, differencing is required. Figure 4 shows the values for VWAP with a differentiation of 1. Note that as the random component suggested in Figure 2, Figure 4 highlights the increasing noise or volatility in the behavior of the VWAP behavior along the years.

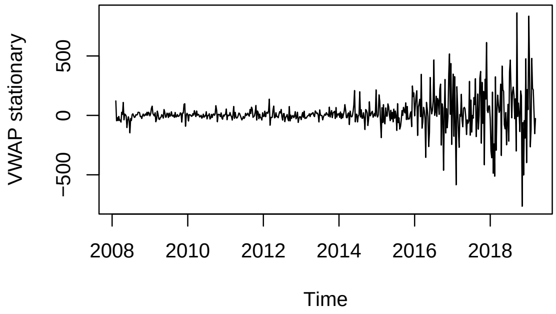

Figure 4: Data plot of VWAP after unitary differentiation

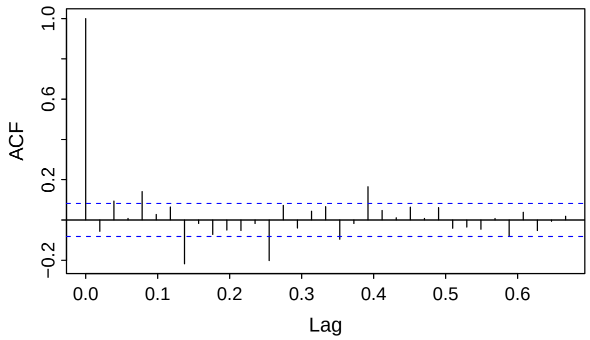

Figure 5: Autocorrelation plot for VWAP after differentiation

Atuocorrelation Autocorrelation also serves for seasonality detection and helps identify various lags worth investigating.

The autocorrelation plot in Figure 2 indicates that the autocorrelation continues to decrease as the lag increases, but still shows strong values; thus confirming that there is a linear association between observations separated by larger lags. To remove seasonality from the data, the seasonal component from the original series is substracted and then differentiated to make it stationary. Figure 5 shows the ACF test for the data already without the seasonal and stationary behaviors.

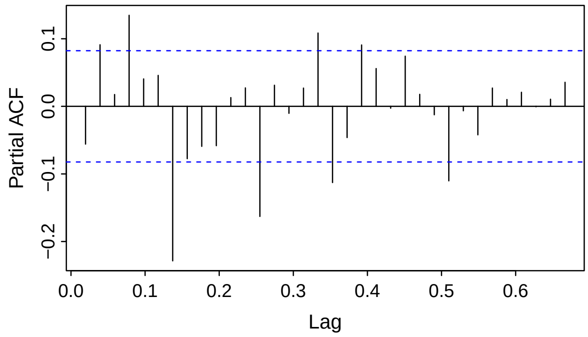

Figure 6: Partial Autocorrelation plot for VWAP after differentiation

3.2 Fitting the model

To fit the ARIMA model, an iterative process must be done to ensure the optimal values of all the ARIMA parameters are selected. These are the following:

p Autoregressive parameter (AR). The number of lag observations included in the model is also called the lag order.

d Integrated parameter (I). The number of times that the raw observations are differenced is also called the degree of differencing.

q Moving average parameter (MA). The size of the moving average window is also called the order of the moving average.

To examine which p and q values will be appropriate, the autocorrelation and partial autocorrelation tests provide useful information. For the former, Figure 5 defines values of the AR and MA parameters depending on the plot shape. Alternate positive and negative spikes decaying to 0 indicates dealing with an AR model without - or with little - MA influence; this is, set $q := 0$. The latter, partial autocorrelation analysis in Figure 6 guides in the order of the AR parameter.

After a few iterations and the results provided by `auto.arima`, the coefficients are set to $p = 1$, $q = 0$, and $d = 1$ (given the performed order 1 differentiation).

3.3 Diagnostic measures

Once the model is built, residuals are tested for various assumptions;

Autocorrelation In order for the model to be valid, the residuals must not be correlated. Top plot in Figure 7 shows the ACF results; the residuals are not correlated.

Heterostadicity Heterogeny of variance in the residuals is also needed for the model to be valid. In this case, it is tested with the Breusch Pagan test.

Normality in the distribution Finally, normality is tested via the Jarque-Bera test, the Kurtosis Test, the Skewness test - all tests passed -, and the QQplot in Figure 7 (the values are normal as they rest on a line).

As all the graphs are in support the assumption that there is no pattern in the residuals, the model is considered valid and forecasting is carried out.

4 VAR

For VAR, the phases are similar to those carried out for ARIMA, a brief exploratory analisys, a detail on the fitting of the model and the diagnostics, lastly.

4.1 Exploratory analysis

Most of the assumptions regarding the data are already tested for ARIMA.
One relevant difference is that the time series selected must be correlated and show some dependency. This is relevant because it ensures the usefulness of one variable to forecast another.

Granger Causality Granger's causality test returns p-values of less than $2.2{\cdot}10^{-6}$ indicating that the open proce time series has causality impact on the VWAP series.

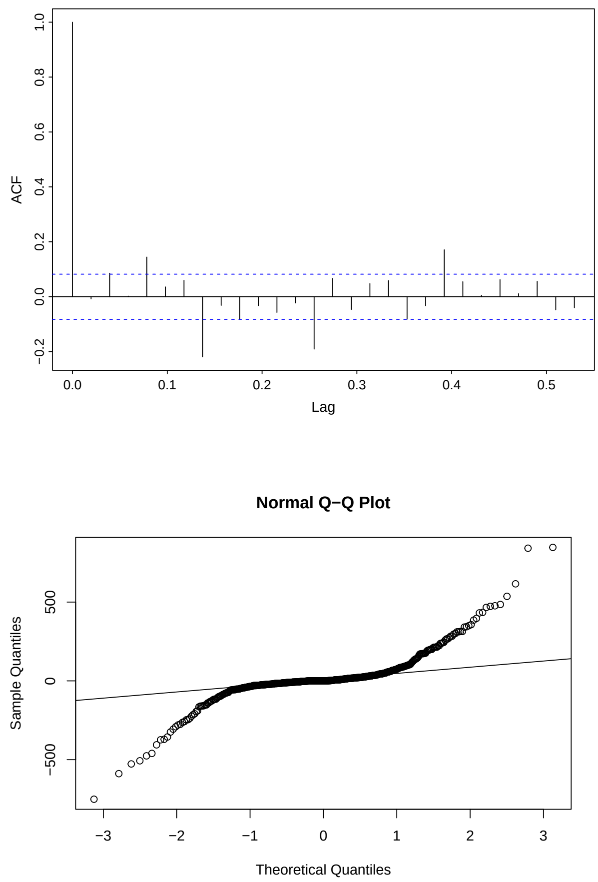

Figure 7: Autocorrelation and QQplot plots for ARIMa residuals

	VWAP	Open
VWAP	1	0.9286
Open	0.9286	1

Table 1: Correlation matrix of residuals

4.2 Fitting the model

VAR has only one parameter; the lag order. The structure of VAR is that each variable is a linear function of past lags of itself and past lags of the other variables. To select the optimal lag order behind the model the implemented `VARselect` function is used, which provides minimum values found for the Akaike information criterion (AIC), Hannan–Quinn information criterion (HQ), Schwarz criterion (SC) and Final Prediction Error (FPE).

Note that greater lag orders indicate the effects reflect greater persistence; for the data, 15 lags are selected.

The resulting model shows a multiple R-Squared of 0.9981 and an adjusted R-squared of 0.998.

In regards to the relationship between the two variables, the correlation matrix of residuals - Table 1 - shows a very strong correlation, previously suspected from data observations and the Granger causality test.

4.3 Diagnostic measures

As with ARIMA, the residuals must be tested to ensure the model is valid. For the first three tests all obtained p-values were satisfactory, no further comment is added in this report to avoid repetitions.

Autocorrelation of the residuals In this case, the Portmanteau and Breusch-Godfrey test is also performed, but this time the computation looks for serially correlated errors.

Heteroscedasticity To check the heterogeneity of variance in the VAR

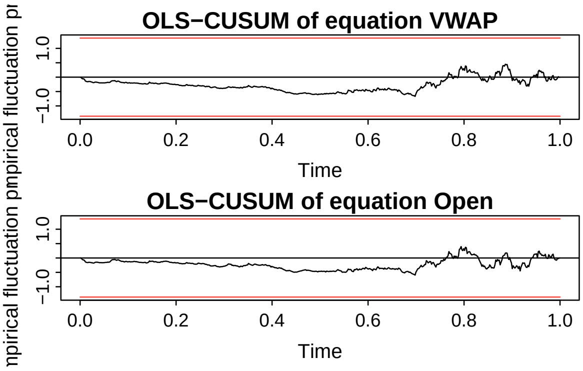

Figure 8: Stability results for the VAR model

model run ARCH Engle's Test for Residual Heteroscedasticity.

Normal distribution of residuals The Jarque-Bera test, the Kurtosis Test, and the Skewness test are performed for the residuals.

Stability test The stability test is a test for the presence of structural breaks. The inability of a model to test for structural breaks could implies that the whole estimation may be inaccurate. If VAR is unstable, the impact of the shocks will never fade (see the Impulse REsponse Functions below), unstable VAR implies that the variables entered in the system are non-stationary. See results in Figure 8.

4.4 Policy Simulations

Forecast Error Variance Decomposition To trace the development of shocks in the system the Forecast Error Variance Decomposition (FEDV) is used.

Figure 9 clearly shows in the top graph that the forecast error of the VWAP at short horizons is due to itself. This is the case because the VWAP was placed first in the ordering and no other shocks affect VWAP contemporaneously. At longer horizons such as 10 weeks, we can see that Open price accounts for about 5%.

For the Open price, its error is still mainly due to VWAP, even for the shortest horizons.

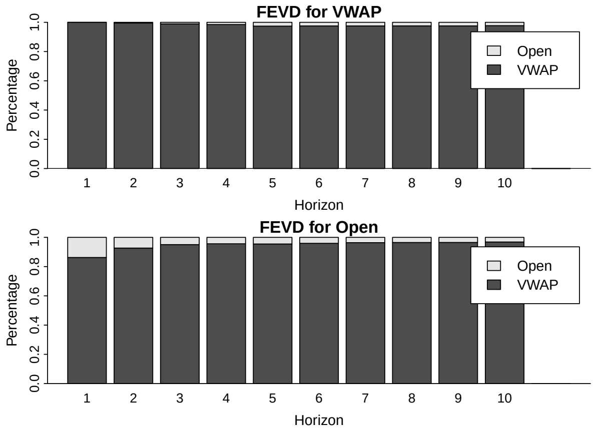

Figure 9: Forecast Error Variance Decomposition (FEDV)

Impulse Response Functions The Impulse Response Functions (IRFs) compute the response of a variable concerning another's impulse - see Figure 10 -. The most informative plots are the top right and bottom left since they portray the responses between the two studied variables. However, note that for a VWAP impulse, both VWAP's (top left) and Open's (top right) responses are quite similar. Equally, for an Open impulse, VWAP's (bottom left) and Open's (bottom right) responses are similar too. This indicates that both variables react very similarly to impulses, regardless of the type of impulse.
The results in the bottom row suggest that prices go up after an Open hike firstly have a negative impact on the prices - both measured via VWAP or Open price -, but recover. Contrarily, a hike in VWAP immediately boosts stock prices and then allows a recovery to the original values approximately.

These results help understand what kind of behavior to expect from forecasting results based on this VAR model and also serve as a control quality of the model.

5 Forecasting results

The results of the predictions and forecasting from the ARIMA and VAR models are presented in this seciton.

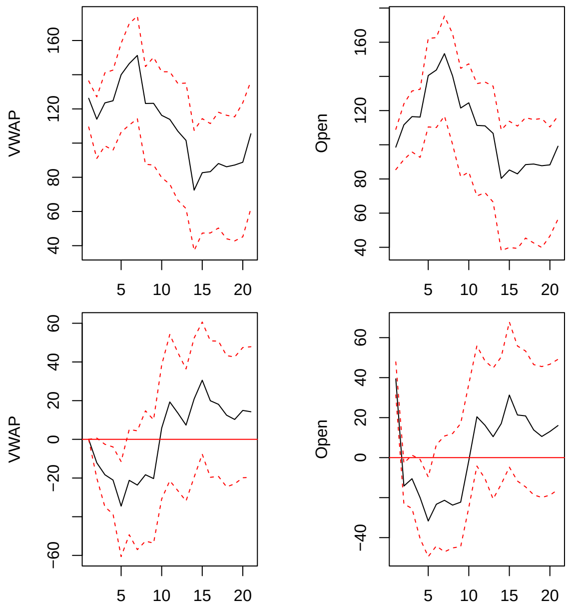

Figure 10: Impulse responses. Top left: VWAP's response to VWAP. Top right: Open's response to VWAP. Bottom left: VWAP's response to Open. Bottom right: Open's response to Open.

5.1 ARIMA forecast

The forecast provided by the ARIMA model is depicted in Figure 11 with the data used and in Figure 12 in detail. Observe that, in comparison with the data used for modeling and forecasting, the prediction is rather inconsistent. Although the forecast appears to show a contant trend in contrast with the growth that the time series data shows, the detail of the forecast reveals that the prediction is also growing, but at a dramatically lower rate. See that the prediction also includes some of the randomness and spikes seen in the data used.
The uncertainty of the prediction grows at a low rate but its magnitude is notable. This issue and the lack of trend for the actual prediction seem to indicate the inability of ARIMA to correctly forecast the data. The given prediction is quite conservative and carries a huge error margin.

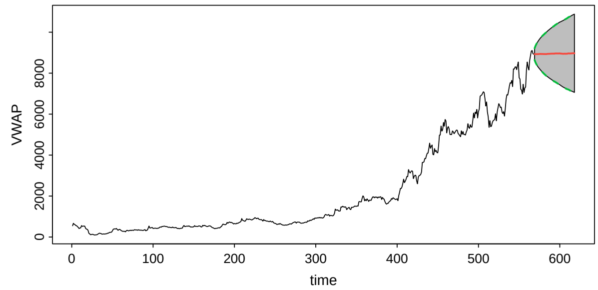

Figure 11: VWAP forecast (ARIMA)

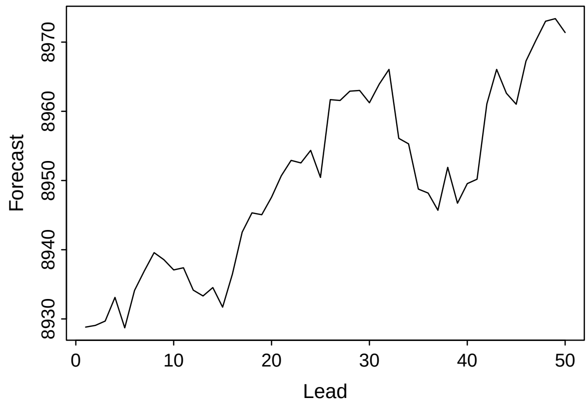

Figure 12: VWAP forecast in detail (ARIMA)

5.2 VAR forecast

The VAR forecast for VWAP can be seen in Figure 13, and for the Open data in Figure 14. For both variables, the apparent accuracy of the prediction is much larger than that of ARIMA. See also that the prediction for both variables is quite similar, which is logical since the VAR model includes both variables and both behave similarly over time.
Furthermore, the uncertainty of the predictions is much smaller that that of ARIMA. This gives indicated the estimations done by the VAR model are more confident.

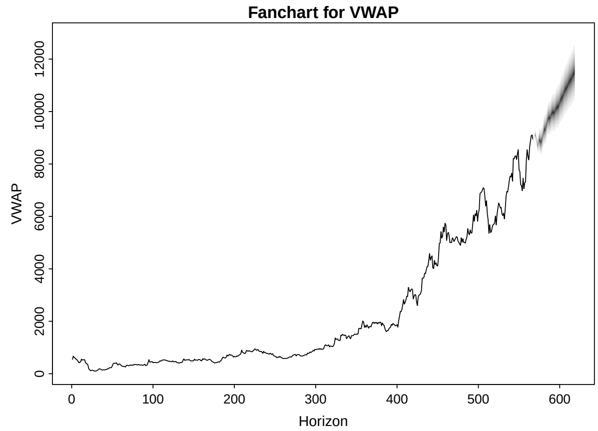

Figure 13: VWAP forecast (VAR)

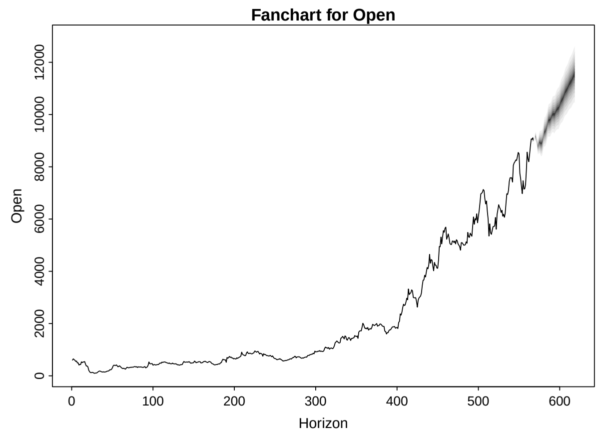

Figure 14: Open price forecast (VAR)

6 Discussion and conclusions

Stock prices are known to be dependent on a myriad of factors, hence often regarded as almost unpredictable.
Consequently, as seen in the forecasts presented

in the results section, a univariate model (ARIMA) underperforming a multivariate one (VAR) is rather expected. However, the univariate time series models present a simplicity that may be well suited for many other data types; from the clarity of trend decomposition to the lack of stability of impulse response computation and checking, ARIMA is a much more flexible method for forecasting. The lack of incorporation of other variable's impact is what makes ARIMA predictions inaccurate for complex data such as stock market. This is because ARIMA uses the own variable's lags are predictors via linear regression in the AR term, and since linear regression performs better if predictors are not correlated, the model fails to capture great part of the phenomena that stock market data contain.

In contrast, the results for predictions using VAR with only two series are surprisingly promising. The use of a lag of 15 - this is, using data from up to 15 weeks before informing the following one - suggests that stock market data is susceptible to be analyzed with long-term dependencies methodologies or long-term recurrent models. It seems fitting to try LSTMs (long short term memory models) to predict stock market prices. The research has compared LSTMs to VARs, and the findings indicate that VAR is better suited for multivariate continuous data, while LSTMs' strength in simulating long-term memory may not be as useful (Goel et al. 2016).

References

Kinney Jr, William R (1978). “ARIMA and regression in analytical review: An empirical test”. In: *Accounting Review*, pp. 48–60.

Lin, Winston T (1989). “Modeling and forecasting hospital patient movements: Univariate and multiple time series approaches”. In: *International Journal of Forecasting* 5.2, pp. 195–208.

Venugopalan, R and M Srinath (1998). “Modelling and forecasting fish catches: Comparison of regression, univariate and multivariate time series methods”. In: *Indian Journal of Fisheries* 45.3, pp. 227–237.

Du Preez, Johann and Stephen F Witt (2003). “Univariate versus multivariate time series forecasting: an application to international tourism demand”. In: *International Journal of Forecasting* 19.3, pp. 435–451.

Vargas-Silva, Carlos (2008). “Monetary policy and the US housing market: A VAR analysis imposing sign restrictions”. In: *Journal of Macroeconomics* 30.3, pp. 977–990.

Lütkepohl, Helmut (2013). “Vector autoregressive models”. In: *Handbook of research methods and applications in empirical macroeconomics*. Edward Elgar Publishing, pp. 139–164.

Box, George EP et al. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Ghysels, Eric (2016). “Macroeconomics and the reality of mixed frequency data”. In: *Journal of Econometrics* 193.2, pp. 294–314.

Goel, Hardik et al. (2016). “Multivariate aviation time series modeling: VARs vs. LSTMs”. In: *Unpublished manuscript Retrieved from https://www% 20semanticscholar% 20org/paper/Multivariate-Aviation-Time-Series-Modeling*.

Iwok, IA and AS Okpe (2016). “A comparative study between univariate and multivariate linear stationary time series models”. In: *American Journal of Mathematics and Statistics* 6.5, pp. 203–212.

Cerqueira, Vitor, Luis Torgo, and Carlos Soares (2019). “Machine learning vs statistical methods for time series forecasting: Size matters”. In: *arXiv preprint arXiv:1909.13316*.

Sethi, Jasleen Kaur and Mamta Mittal (2020). “Analysis of air quality using univariate and multivariate time series models”. In: *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*. IEEE, pp. 823–827.

World Health Organization (2020). *Listings of WHO’s response to COVID-19*. `https://www.who.int/news/item/29-06-2020-covidtimeline`. Accessed: 2023-April-12.