NTNU
Kunnskap for en bedre verden

Department of Computer Science

TDT4173

Forecasting Stock Closing Price with Autoregressive Integrated Moving Average

Authors:
Olav Røed Meberg
Fridtjof Høyer

November 29, 2020

Abstract

The stock price of a company is important for many stakeholders. It works as a critical KPI for the enterprises financial and operational health, and thus predicting it carries huge benefits and competitive advantages. Although the price may fluctuate in a nonlinear fashion, predicting it linearly over short timeframe, may be sufficient for some applications. This paper will discuss how to utilize the ARIMA model forecast closing prices for the upcoming day. This paper shows that in short-term predictions the ARIMA model has significant forecasting abilities, supporting other research conducted. This paper will exclusively focus on closing prices and the end of the intraday. The data analysed is provided by the Oslo Stock Exchange form Q1 2015 to Q3 2020 on Equinor ASA. We note unoriginal cliché, that past performance is not a guarantee for future results but nevertheless, a good indication of the direction of the price and bid–ask spread.

The suggested template provided by NTNU was not to a full extent used in this article, since a section just containing data was not suitable for the given problem. For a easy understanding of our task, the topic is told chronological and the data is continuously described throughout the article.

0.1 Additional Resources

Please consider visiting our GitHub repository and web page. You will be able to find our dataset, code, some non-essential bonus explanations, and calculations.

- `https://github.com/Fridthoy/ARIMA`

For additional information please contact: olavrm@ntnu.no and/or fridthoy@ntnu.no

Table of Contents

0.1 Additional Resources i

List of Figures **ii**

List of Tables **iii**

1 Introduction **1**

1.1 The Experiment 1

2 Methods **2**

2.1 The ARIMA Model 2

2.2 Our Approach 4

3 Results and Conclusions **7**

4 Related Work and Reflection **9**

Bibliography **11**

Appendix **12**

A Equations 12

B Supplementary Figures and Charts 12

List of Figures

1 Example on patrial correlation: Consider three distinct values in a plane. The points would have some amount of correlation with one another. The calculation for autocorrelation of Lag-2, controlling for lag-1, would be the partial corelation. . 4

2 Our Approach inspired from [Hyndman and Athanasopoulos, February 2020, Ch. 9.7] 4

3 plot of the EQRN closing prices of a period of five years 5

4 plot of the EQNR closing prices of a period of five years differenced ones 6

5 Partial autocorrelation plot of closing prices 6

6 Autocorrelation plot of closing prices 7

7 Residual error 7

8 Prediction compared to actual stock prices 8

9 This is a snapshot of the predications in the data set for the month of October 2020. By observation, the two predictions P1 and P2 both succeed to meet the criteria of hitting between the boundaries of the investors for the most part. The predictions are done per intaday with 75% training data 9

10 BP layer 10

11 Results 10
12 Deviation from actual values. The red and blue line represent the total data at any given point 12

List of Tables

1 Dickey-fuller test for Equinor closing prices 5
2 Evaluating forecast accuracy 8

1 Introduction

The topic for this paper is stock marked forecasting, and the ability to make predictions based on historical observations. Thus, creating hypothetical competitive advantages. As proposed, the article will calculate future stock prices for Equinor ASA (EQNR), although the algorithm should be valid for any given non-volatile stock listed at the Oslo Stock Exchange. There are however a few reasons for choosing Equinor ASA as the target enterprise for this research. One hypothesis is that following factors allows for greater stability on results and higher predictability. Thus, choosing a high volatile stock may result in pseudoscientific nonsense due to the stocks random nature.

- Equinor has historically low volatility and β. (Not to be confused with other β in this paper)
- The Norwegian Ministry of Oil and Energy are the main shareholder, increasing enterprise stability.
- The stock is traded in high volume and high frequency.
- The enterprise value is steady.

These supporting reasons would also be true for a great deal of the Norwegian heavy-wights, but out of authors' curiosity and motivations for the EP-sector, Equinor ASA was chosen. The data frame used in this project is publicly provided here OsloBors [2020].

There is an ocean of different ways to predict a stocks future prospect. From the short term intraday technical signals such as Relative Strength Index (RSI), Bollinger Bands, to long term forecasting like Simple Moving Average (SMA: special case of ARIMA(0,0,1)) or the rather naïve "draw a straight line" principle. This project will run a ARIMA algorithm to predict the closing prices.

The reason to run an ARIMA model in this project, is the properties associated with underlying modules. ARIMA models consists of an autoregressive (AR) and moving average (MA) component. Readers may think of the AR as "the change since last time or happening", and the MA component describes the smoothed trends (must not be confused with exponential smoothing [Hyndman and Athanasopoulos, February 2020, Ch. 8.1]) in the observation set. The integrate, hence the "I", determines the level of differencing needed to force the data to be stationary. This will be discussed in detail later in section 2.1. Forecasting is generally hard and the ARIMA models occupy a mid-range area of in the forecasting toolbox.

1.1 The Experiment

The experiment the team were interested to run was for forecasting on future closing prices for the upcoming intraday. The experiment has been conducted before by others like for example in 4 with other data sets and parameters. The target objective was on low volatility, openly traded, capital stocks. Every intraday has a high (h_t) and a low (l_t) where $h_t \geq closing \geq l_t$. The Experiment is deemed successful if and only if the prediction P_t and the actual value A_t meets:

1.1.1

$P_t \geq l_t \wedge P_t \leq h_t$ is true for at least 51% of the predicitons

1.1.2

The error estimator $MRE : 1,90\% > \sum_{t=0}^{t} \frac{|P_t - A_t|}{A_t}$

1.1.1 is a non-scientific version of the squeeze theorem. By believing that the investors have a fairly universal understanding of the true value of the stock price, the method should aim to fit within the same upper and lower boundaries as the investors. This goal is measured as the sum binomial successes s, with respect to the number of predictions n. Hence, $\frac{1}{n}\Sigma s_n$
1.1.2 The MRE goal is derived from the sum of relative deviation of the mean, from closing to low and closing to high A. In other words, 1,9% was chosen because this represent the average stock fluctuating magnitude in both directions for a given random day. 1,9% is equivalent to approximately kr 3,0 NOK on average. The experiment has failed if one or more the gaols are not met.

2 Methods

2.1 The ARIMA Model

In short, the main idea of the ARIMA model is to capture autocorrelations in the series by modelling it directly [Hyndman and Athanasopoulos, February 2020, Ch. 4.2] and [Hyndman and Athanasopoulos, February 2020, Ch. 9.0] . The ARIMA models provides another useful method for forecasting [Hyndman and Athanasopoulos, February 2020, Ch. 9.0]. The model has a strong underlying foundation of mathematical and statistical theory, making it quite solid for many applications. The ARIMA model is very flexible and are able to capture lots of different patterns in the series. [Hyndman and Athanasopoulos, February 2020, Ch. 9.1]

2.1.1 Differencing and Stationarity

Prior to discussing the models in detail, let us consider the limitations the ARIMA model (and autoregressive family) is bound by. The ARIMA model assumes that the series is stationery. Thus implying; no trend or seasonality, a constant level and variance, and autocorrelation that remains steady throughout the series [Hyndman and Athanasopoulos, February 2020, Ch. 9.1]. In other words, a stationary time series is a series that is not constrained by time at which the time series are observed [Hyndman and Athanasopoulos, February 2020, Ch. 9.1]. It is also important to note that ARIMA models extrapolate patterns form past observation into the future. Therefore, assumption is that the future will remain similar, or comparable to what ever have been observed prior to the given point in time.
In this case study [2.2] of stock prices, a non-stationary dataset is likely to appear. Therefore, the paper wishes to present one the of techniques of transforming a series to its stationary form by *differencing*. Differencing is the act of calculating the difference between consecutive points in the series [Hyndman and Athanasopoulos, February 2020, Ch. 9.1] and stabilizing the mean, level, and variance (might be necessary with second-order differencing for some series. Please consider reading [Hyndman and Athanasopoulos, February 2020, Ch. 9.1] for closer details on this topic).

2.1.2 AR(p) – Autoregressive

The autoregressive model, AR(p) forecast using a linear combination of predictors of past observations [Hyndman and Athanasopoulos, February 2020, Ch. 9.3] (in this example β). The AR model order $p \in \mathbb{N}_0$ may be given by:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \ldots + \beta_p Y_{t-p} + \epsilon_t$$

Given a predictor Y_t, β is the autocorrelation, $t, p \in \mathbb{N}_0$, and ϵ_t denote the white noise. [Hyndman and Athanasopoulos, February 2020, Ch. 9.3]

2.1.3 MA(q) – Moving average

In contrast to the AR models, the MA model uses past errors ϵ_t in a "regression" [Hyndman and Athanasopoulos, February 2020, Ch. 9.4]. However, the MA model of order $q \in \mathbb{N}_0$ share some similarities to the regression like denotation:

$$Y_t = C + \epsilon_t \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

Additionally, note that θ is the parameter of the moving average and Y_t may be tough of as the weighted moving average [Hyndman and Athanasopoulos, February 2020, Ch. 9.4].

2.1.4 ARIMA(p, d, q)

The ARIMA model's main idea to capture all forms of autocorrelation by including lags of the series, and of the forecast errors model [Hyndman and Athanasopoulos, February 2020, Ch. 9.0]. This model includes predictors, in addition to the p -lagged series also q-lagged versions of the forecast errors, and thus we continue to assume that the series is stationary. By adding a differencing step called integrate (I) we get the full Autoregressive Integrated Moving Average model. This differencing operator removes trend and, or the seasonality. The ARIMA models may be done in the two following steps. First apply differencing (of order d, and typically lag-1 for a moving trend), and then fit the ARIMA(p, d, q) model for the different series. The models ARIMA(p, d, q) may be written as the following:

$$Y'_t = c + \beta_1 Y'_{t-1} + \beta_2 Y'_{t-2} + \ldots + \beta_p Y'_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

Where d denotes the order of differencing where $d \in \mathbb{N}_0$ (ch. 9.5). The parameter d in ARIMA (p, d, q) tells how many time to apply the lag-1 differencing as for instance:

- $d = 0$ no differencing.
- $d = 1$ linear trend where differencing is preformed once.
- $d = 2$ is double differencing. After differencing the original series at lag-1, it is again differenced the difference series at lag-1. [Hyndman and Athanasopoulos, February 2020, Ch. 9.1]

2.1.5 Order Selection of the Parameters p and q using ACF and PACF

The parameters p and q in an ARIMA(p, d, q) can often be hard to calculate without specialized software [Hyndman and Athanasopoulos, February 2020, Ch. 9.5]. Thus, the process of selecting the parameters p, q manually involve numerous rounds of differencing, reviewing of plots, comparison, and experience. For many applications the ARIMA model is sufficient with either an MA(q) or an AR(p) component, like for instance ARIMA($p = 0, d, q$) or ARIMA($p, d, q = 0$)[Hyndman and Athanasopoulos, February 2020, Ch. 9.5]. Models with parameters of $p > 0 \wedge q > 0$ are less common and experience show that they may be tricky to work with, as demonstrated in section 2.2 and 3. When looking for candidates for the parameters the ACF plots might be useful If the positive autocorrelation at lag-1, the AR(p) terms are probably the best for forecasting applications. And similarly, if there is negative autocorrelation at lag-1, MA(q) terms must be considered [Hyndman and Athanasopoulos, February 2020, Ch. 9.5]. The decay of the autocorrelation function (ACF) will for some series give of small clues of what terms and orders to use. It is not rare to find correlations issues in the ACF, or rather lack of new information in the data [Hyndman and Athanasopoulos, February 2020, Ch. 9.5]. For those cases, the PACF must be considered. Partial correlation is the correlation between two distinct variables regulating for the values of another set of variables as seen in figure 1 [Dégerine and Lambert-Lacroix [October 2003]].

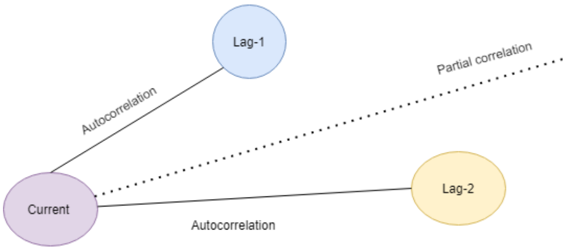

Figure 1: Example on patrial correlation: Consider three distinct values in a plane. The points would have some amount of correlation with one another. The calculation for autocorrelation of Lag-2, controlling for lag-1, would be the partial corelation.

Source: Authors

Assume a partial autocorrelation function plot with a sizable spike at lag-j. This means that all autocorrelation is largely described by lag-j [Hyndman and Athanasopoulos, February 2020, Ch. 9.5]. The partial autocorrelation function may also decide what parameter p to use in the AR(p). If the PACF cuts or drops of at lag-k, generally means that AR($p = k$) is a good explanation of the pattern in the series. If it is a more gradual descent, suggest a more weighted q in a MA(q) model [Hyndman and Athanasopoulos, February 2020, Ch. 9.5].
If the parameters $p \wedge q > 0$ the ACF and PACF functions are not so useful. The Akaike's and Bayesian Information Criterion might be useful for finding the correct orders to use for the model [Hyndman and Athanasopoulos, February 2020, Ch. 9.6]. These models are described more careful in 3.

2.2 Our Approach

The stock prices are dependent on many factors, such as supply and demand, market trends, popularity, the global economy, and much more Seneviratna and jun Wang [October 2013]. In this experiment the focus is upon forecasting the close price on the Equinor ASA stock for the upcoming day. The steps for forecasting the model may be seen in figure 2. In this section the first three steps is described, in section 3 the remaining subsequent steps will be described in detail.

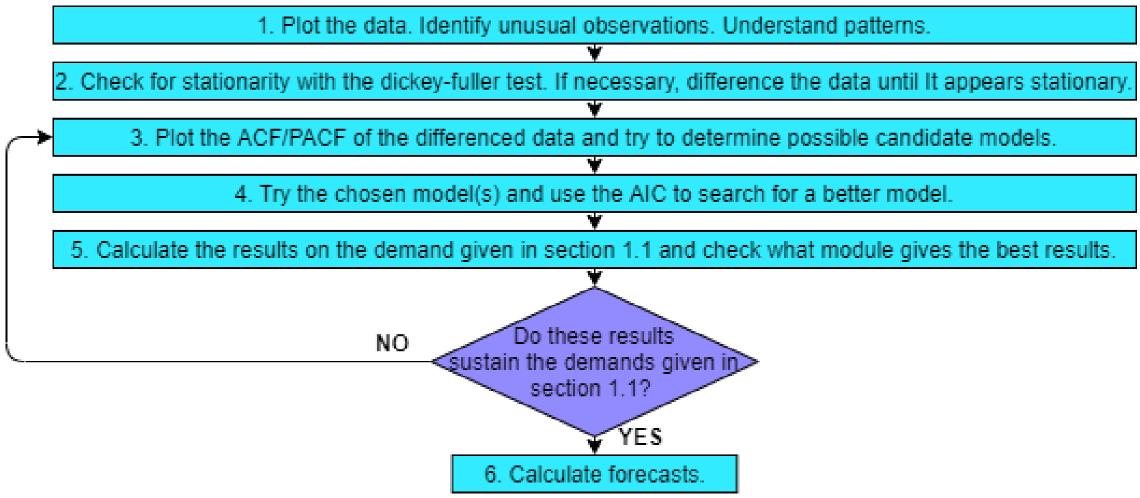

Figure 2: Our Approach inspired from [Hyndman and Athanasopoulos, February 2020, Ch. 9.7]

Source: Authors

The first step in our analysis was to create exploratory data analysis, here visualization of the stock was created to summarize the main characteristics [Hyndman and Athanasopoulos, February 2020, Ch. 9.7]. The Equinor stock prices gathered from Oslo Stock Exchange needed to be 'cleaned up', to provide Python with the information about what type the different columns from the CSV file contains. The plot in figure 3 shows the plot of the closing prices of Equinor over a period of five years.

Figure 3: plot of the EQRN closing prices of a period of five years

Source: Authors

As described in section 2.1.1 the ARIMA model must be stationary. By looking at the graph 3 one can easily see that the provided graph for the closing prices is not stationary. Therefore, the closing prices needs to be converted into a stationary series. A common test for testing if a series is stationary, is the Dickey-Fuller test Ajewole K.P [2020]. This test where integrated for testing stationarity of the Equinor closing prices. The result from the test can be seen in table 1.

Test statistic	-1,496875
Critical value at 1%	-3.43561
Critical value at 5%	-2.86386
Critical value at 10%	-2.56801

Table 1: Dickey-fuller test for Equinor closing prices

For the Dickey-Fuller test there are a null hypothesis that the time series is Non-stationary, and the alternate hypothesis that the time series is stationary. If the test statistic is less than the critical value one can reject the null hypotheses. As seen in Table 1, one can see that the test Statistic is greater than the critical values, therefore the time series is non-stationary.

There are many approaches for converting a series from non-stationary to a stationary series. One simple and optimal approach that gave the correct results in our case was to calculate the discrete difference along the closing price series. The plot of this is given in figure 4, one can see that the data is stationary by only looking at this plot [Hyndman and Athanasopoulos, February 2020, Ch. 9.7].

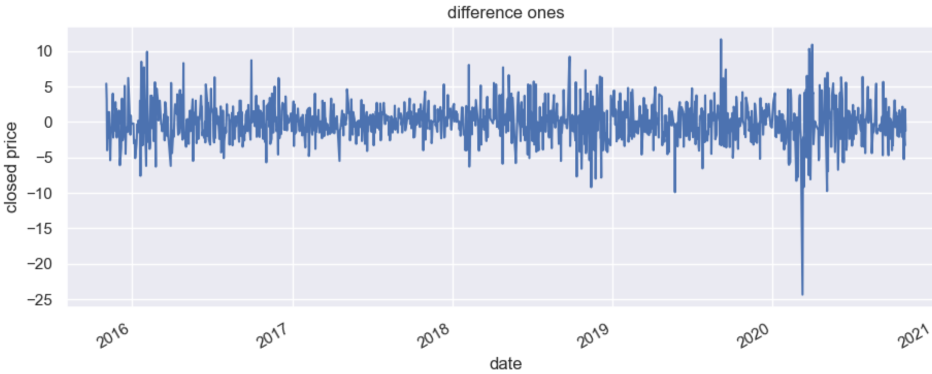

Figure 4: plot of the EQNR closing prices of a period of five years differenced ones

Source: Authors

After differencing the series ones, the test statistics was calculated to be $-12,627$. The test statistics is less than critical values seen in table 1, therefore the series is now stationary.

As described in section 2.1 the ARIMA model is dependent on three parameters, namely AR, I and MA. These are generally denoted as p, d and q and are important for creating an optimal forecast. Calculating stationarity is how the ARIMA model calculates the parameter d. This parameter is associated with the integrated part of the model, and in our case this is 1.

The next step is to calculate both the p and q parameters to be implemented in the ARIMA model. There are several ways to calculate this, but one way to identify the number of AR terms is to inspect the partial autocorrelation plot (PACF) [Hyndman and Athanasopoulos, February 2020, Ch. 9.7]. This is a plot that looks at the correlation between the series and its lag. With the help of PACF one can know if a lag is needed in the AR term or not. The order of AR terms is equal to the number of lags that crosses the significance limit in the PACF plot given a positive PACF. These are the lags that stands out and we want to investigate.

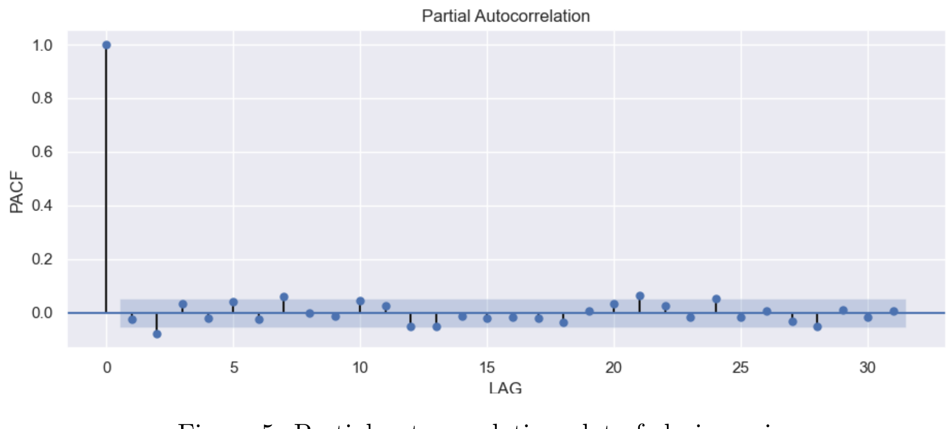

Figure 5: Partial autocorrelation plot of closing prices

Source: Authors

For calculating the MA follow the same patterns as described in 2.1.5 [Hyndman and Athanasopoulos, February 2020, Ch. 9.7]. The ACF plot tells how many MA terms that are required to remove autocorrelation in the stationarized series. Just as finding the p value finding the lags above the significance line will determine what the q value is.

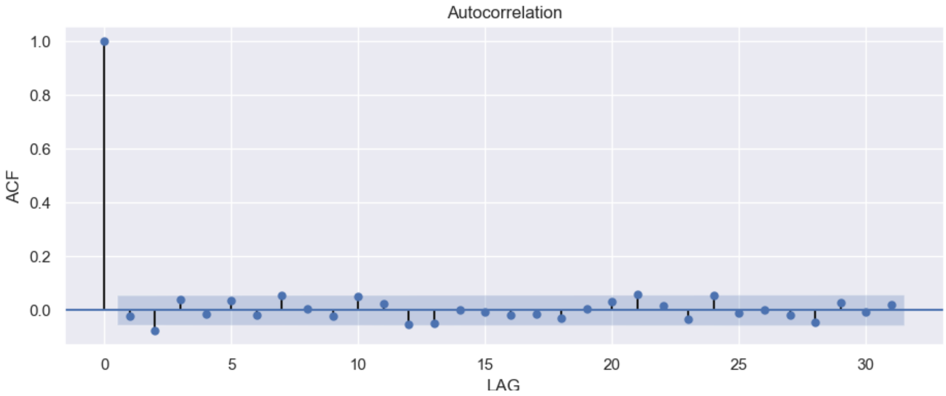

Figure 6: Autocorrelation plot of closing prices

Source: Authors

From the ACF and PACF plots (6 , 5) the lag with the most significance are the values $p = q = 7$ and are therefore the most interesting to investigate further [Hyndman and Athanasopoulos, February 2020, Ch. 9.7].

3 Results and Conclusions

One common way to look at the accuracy in a prediction is to analyze the residual errors created when one compares the forecast to the actual values [Hyndman and Athanasopoulos, February 2020, Ch. 5.4]. In figure 7 it's possible to observe that the mean of the residual errors is approximately zero. In addition to this, the residual is normally distributed. Verifying these properties is important when checking if the method is using all available information, but it does not tell if the parameters in the ARIMA model is the best fit for the given time series [Hyndman and Athanasopoulos, February 2020, chapter 3.3]. There are many methods that can satisfy these properties, therefore it's important to verify weather or not the selected order is the most accurate parameters for the ARIMA model.

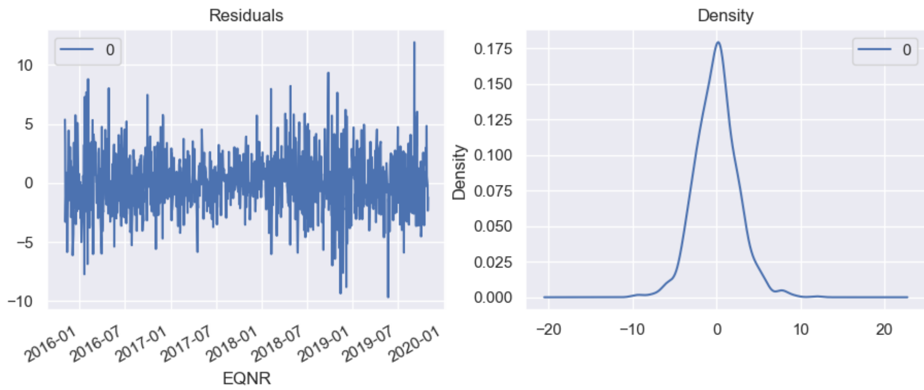

Figure 7: Residual error

Source: Authors

In figure 7 one can visualize the precision of the results. The forecast with the parameters $p = q = 7$ seems accurate by looking at the graph. In addition to this, one can look at a given interval in figure 8 where the prediction from the model is compared to the actual values. The predictions seems accurate from the comparison.

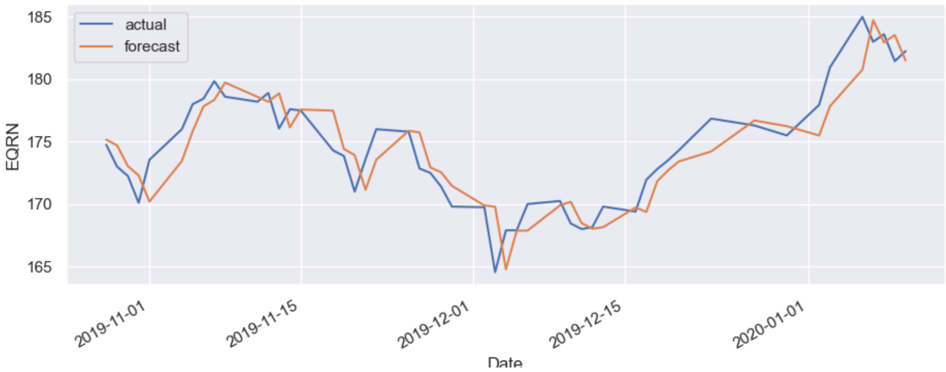

Figure 8: Prediction compared to actual stock prices

Source: Authors

another useful approach for selecting predictors for regression and for determining the order of the ARIMA model is the Akaike's Information Criterion (AIC) (given $p \wedge q > 0$). AIC is used to compare different models and determine what model that fits the data with most precision [Hyndman and Athanasopoulos, February 2020, Ch. 9.6]. The AIC is calculated from how well the model reproduce the data, as well as checking the number of independent variables to build the model. AIC also considers the risk of overfitting and underfitting the model [Hyndman and Athanasopoulos, February 2020, Ch. 9.6]. As our data-frame is over a time period of 5 years the chance of overfitting is high. Therefore, the implementation of AIC is suitable for the dataset as it will punish overfitting. The AIC formula is given by:

$$AIC = -2log(L) + 2(p + q + k + 1)$$

L denotes the likelihood of the data. c is the parameters $\beta_1, \beta_2, ..., \beta_p, \theta_1, \theta_2, ..., \theta_q$, $k = 1$ if $c \neq 0$, and if $c = 0$ then $k = 0$. The last term in the formula punish the module if the sum of p and q is high [Hyndman and Athanasopoulos, February 2020, Ch. 9.6]. As seen a lower AIC implies a better model.

By iterating over every combination of p and q from $p = q = 0$ to $p + q \leq 16$ and calculating the AIC for every module, the best result was given with the paramters $p = 6$ and $q = 7$. The AIC value for $p = 6$ and $q = 7$ was 4109, compared to $p = q = 7$ with an AIC value equal 4111. The results are promising as the values do politely agree with our ACF and PACF interpretation.

Even if the AIC value for $p = 6$ and $q = 7$ is the lowest, the candidate module predicted from the PAC and PACF graphs may have a better prediction for achieving our experiment target 1.1.1. The main objective as described in section 1.1.1 is to have at least 51% of the predicted data between the boundaries of high and low prices during the same day. In addition to this we wanted to have an MRE less than 1,9% 1.1.2.

	P1; ARIMA(7,1,7)	P2; ARIMA(6,1,7)	P1 is best
MAE:	2,0263	2,031902	True
MAD:	23,7799	23,7800	True
MRE:	1,3150%	1,3185%	True
Pred in intervall:	56,12%	58,27%	False
Corr:	0,99551195	0,995510441	True
std.div:	2,754160836	2,754602844	True

Table 2: Evaluating forecast accuracy

Table 2 shows a combination of different results for calculating the accuracy of the two models discussed. As seen prediction 2 (denoted: P2) has more predictions inside the interval compared

to prediction 1 (denoted: P1). However, this does not mean P2 is the best forecast. This is a non-scientific version of the squeeze theorem as described in 1.1.1. The results from P1 and P2 where very equal in all error estimations, though P1 was the best for all but one of the tests. On that basis, the paper found that P2 is better than P1 by chance. To strengthen this argument, the paper also includes an absolute deviation chart in the appendix B, were it is possible to see that P1 is much smoother. The formulas used to calculate the errors can be found in the formula section of the appendix A. P1 and P2 both pasted the main objective, namely 1.1.1 and 1.1.2 with a clear margin, and thus the experiment was deemed successful. It is natural to ponder on why these test results were by far greater than the one described in table 11. It might be caused by non-ideal orders of the model, EQNRs data set was easier to forecast*, or the share unexplainable complexity of the stock market and random events. Further work should include improving the model by adding a neural network layer to be able to forecast not only piecewise, but also non-linearly, as this is shown in application 4, this should improve the accuracy. It should also include forecasting and examining correlating factors such as competitors, supply chain, alliance partners, and for the price of oil and gas
*[deemed as a low chance for the reason that indexes are usually less volatile then standalone stocks]

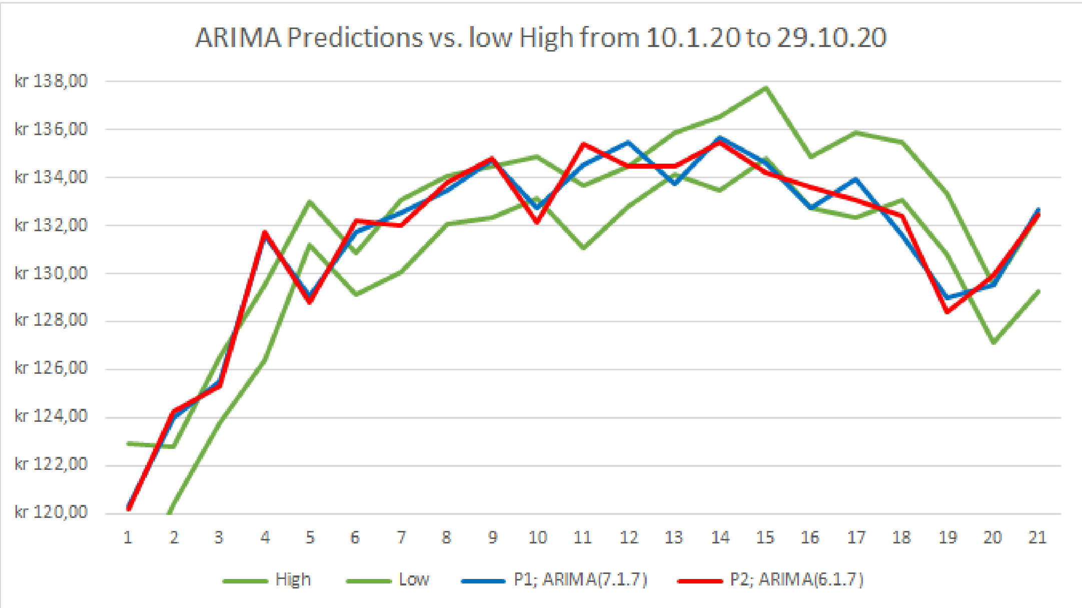

Figure 9: This is a snapshot of the predications in the data set for the month of October 2020. By observation, the two predictions P1 and P2 both succeed to meet the criteria of hitting between the boundaries of the investors for the most part. The predictions are done per intaday with 75% training data

Source: Authors

4 Related Work and Reflection

This section is dedicated to introduce the paper "Application and analysis of forecasting stock price index based on combination of ARIMA model and BP neural network" (the paper (not to be confused with; this paper)) by Yulin[a] and Du[b] form the Fudan University ,Shanghai, China[a], and East China University of Political and Science of Law, Shanghai[b].

The paper describes a very comparable problem as described earlier in this paper. Is it possible to estimate the Shanghai securities composition index by ARIMA, and furthermore, with a layer of neural network? This paper will mainly focus on the ARIMA with the additional Back Propagation (BP) neural network layer, hence the ARIMA-BP model. It is a vast amount different forecasting methods out in the wild. The main differences between these methods are that they can be split up into linear and nonlinear models. The experiment runed by Yulin and Du uses the ARIMA model to conduct a linear forecast whilst the BP neural network would represent a nonlinear forecast. This combination of ARIMA and BP (ARIMA-BP) is more accurate than ARIMA or BP neural network alone, suggest the study.

By assuming that all stock prices indexes follow a linear structure, the ARIMA model would be great alone. However, the stock price follows no such system and has a random walk by nature. Therefore, adding a nonlinear forecasting method seems like a good concept when the ARIMA model alone is insufficient. The Back-Propagation network is a multilayer feed forward neural network. In short, the Model takes the data $X_1, X_2, X_3, \ldots, X_n$ in as a input in the input layer and returns $Y_1, Y_2, Y_2, \ldots, Y_n$ form the output layer, through the weights of ω_{ij} and ω_{jk} as shown in figure X.

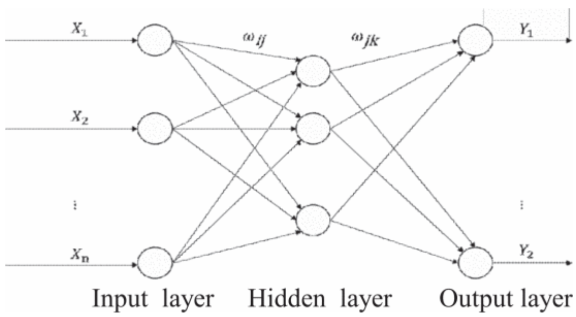

Figure 10: BP layer

Source: Yulin Du

The ARIMA-BP model synthesises ARIMA models and the BP neural network. The results produced by the linear ARIMA model $(X_1, X_2, X_3, \ldots, X_n)$ is used as input parameters for the nonlinear BP model, subsequently giving a nonlinear result. The results are as shown more precise by a factor of 5 and 7 with regards to BP and ARIMA, respectively.

Method	MRE	RMSE
ARIMA	7.45%	89.21
BP	5.61%	65.73
ARIMA-BP	1.03%	18.02

Figure 11: Results

Source: Yulin Du

Although the results form this paper is quite decent, the financial marked is complex and no forecast method alone will be accurate over a considerable timespan, due to its random nature, happenings, and dependencies. The paper still shows how the integration of two models may be powerful tool compered to using only one.

* * *

Bibliography

Jemilohun V.G Ajewole K.P, Adejuwon S.O. Test for stationarity on inflation rates in nigeria using augmented dickey fuller test and phillips-persons test. *Journal of Mathematics*, 2020.

Serge Dégerine and Sophie Lambert-Lacroix. Characterization of the partial autocorrelation function of nonstationary time series. *Journal of Multivariate Analysis*, 81:46–59, October 2003.

Rob J Hyndman and George Athanasopoulos. *Forecasting Principle and Practice 2nd edition*. February 2020.

OsloBors. Oslo børs. `https://www.oslobors.no/markedsaktivitet/#/details/EQNR.OSE/overview`, 2020.

R. M. K. T. Rathnayaka D. M. K. N. Seneviratna and Z. jun Wang. An investigation of statistical behaviors of the stock market fluctuations in the colombo stock market: Arma pca approach. *Journal of Multivariate Analysis*, October 2013.

Appendix

A Equations

- $RE_i := \sum_{j=1}^{n_i} \frac{\omega_{ij}}{\omega_i} \frac{|p_{ij} - d_{ij}|}{|d_{ij}|}$ for $\omega_i = \sum_{j=1}^{n_i} > 0$
- $MRE := \frac{1}{n} \sum_{i=1}^{n} RE_i$
- $MAD := \frac{1}{n} \sum_{i=1}^{n} |x_i - \overline{X}|$
- $MAE := \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$
- $\sigma := \sqrt{\frac{1}{n} \sum_{i=1}^{n} (x_i -)^2}$
- $Corr[X, Y] = \frac{Cov[X,Y]}{\sqrt{Var[X] \cdot Var[Y]}}$

B Supplementary Figures and Charts

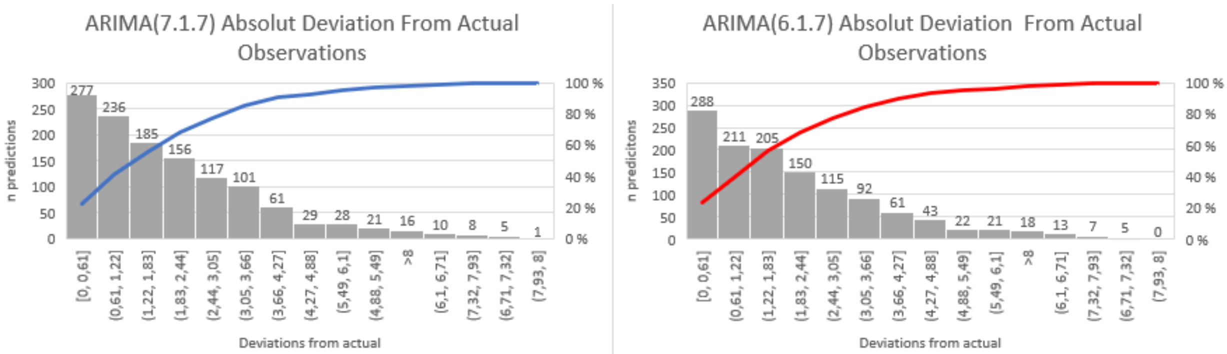

Figure 12: Deviation from actual values. The red and blue line represent the total data at any given point

Source: Authors