Predicting Stock Prices with ARIMA Time Series Models

Pranav Ahluwalia,[1] Mike Andrei,[2]
[1]*Texas Tech University High School, 2515 15th St, Lubbock, TX 79401*
e-mail address: pranav.ahluwalia2@gmail.com
[2]*University of California, Santa Barbara, Department of Statistics and Applied Probability*

Recently, there has been widespread use of automated functions to predict price action in markets. In many of these functions, past data is used to forecast future data. In this project, an Auto-Regressive Integrated Moving Average (ARIMA) model was used to forecast the price action of various stocks. The fundamental question of the project was to see if price data alone would be sufficient to make accurate forecasts for stock prices. While quantitative trading has mostly been reserved for major financial institutions, this project aimed to develop a quantitative trading strategy accessible to the retail trader. Since the retail trader does not have an army of analysts to conduct research on various companies, models were made using the only data that is completely accessible to everyone; price data.

Keywords: ARIMA, Quantitative Finance, Forecasting, Trading, Stock Price Prediction

I. INTRODUCTION

The growing use of automated trading by institutional investors has left those with limited resources at a disadvantage. By using statistical forecasting and time series, we created a methodology to predict stock prices. Thus, the goal of this project is to explore whether or not it is possible to form a useful quantitative trading strategy using an ARIMA model, and to assess its accuracy.

A time series is a model of data using a plot of successive time intervals to showcase potential correlation. The ARIMA model is a time series with three components: an autoregressive component, a differencing component, and a moving average component. The Autoregressive component correlates past data with successive data. The differencing component makes the data "stationary" in attempt to filter randomized data from correlated data. The moving average component accounts for past forecasting errors to make the model adaptive [3].

The benefit of quantitative trading is that it is driven exclusively by price data instead of earnings reports and the news. In today's world, retail traders are constantly bombarded with a plethora of sources offering different opinions on the direction of stock prices. Consequently, the majority of retail traders have no clear filter from which to develop useable trading strategies. Because it is never wise to rely solely on one source when trading, this project will explain how the ARIMA model can be used in conjunction with other tools as well as when it should not be used at all.

What makes this project unique is the simplicity of the model used. While most ARIMA trading strategies are combined with additional linear regressions and generalized autoregressive conditional heteroskedasticity (GARCH) models, this paper exclusively focuses on the use of ARIMA without any external regressors.

II. METHODS

An ARIMA(p,d,q) model primarily deals with three parameters respectively: the lag on the autoregressive component, the order of differencing, and the lag on the moving average component. The autoregressive component (equation 1) with c as a constant, ϕ as a coefficient and e_t as white noise is combined with the moving average component (equation 2) with c as a constant, θ as a parameter and e_t as a lagged error

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + e_t \quad (1)$$

$$y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_t e_{t-q} \quad (2)$$

A lag specifies the spacing in which items are correlated with each other. In the case of this project, the lag specifies a number of days. The order of differencing specifies how many times the data is differenced in order to achieve stationarity. Additional parameters such as a drift (a linear regression of errors to specify trend direction), and a lambda value (a coefficient used to transform the data) can be applied to the model to refine its accuracy.

For this project, closing prices and daily mean prices of Facebook, Costco, Smith and Wesson, and Netflix were used to build ARIMA models and conduct forecasts. The testing data set for the stocks lasted from June 3rd to July 15th 2016. The beginning of the training data set was optimized for each stock but ended on June 2nd 2016. The accuracy of each model was assessed using three parameters: Root Mean Square Error (RMSE), Mean Absolute Percent Error (MAPE), and Price Point Accuracy (PPA); a measure of how often the model predicted exact prices.

Before an ARIMA model can be fitted, a good training data set must be selected. This process is crucial because if poor data is selected, the model's forecasts will be naïve. The key to selecting an optimal data set is an awareness of ARIMA's limitations. ARIMA models are known to function best when fitted to trending data, but don't deal well with fluctuating volatility [5]. For this reason, trending stocks with relatively even variances over time were chosen. In order to check if the stocks were trending, a charting software was used to fit a trend line to the support of the price data. The beginning of the trend was then selected as the first date of the set. An example of this technique is shown in figure 1.

After a useable data set was selected, a Box-Cox transformation (equation 4) was applied to further stabilize variance.

$$\chi'_\lambda = \frac{\chi^\lambda - 1}{\lambda} \tag{4}$$

The data was then differenced with a lag of 1 ($y_n - y_{n-1}$) and made to be stationary. The necessary degree of differencing was determined by the ndiffs() function in R along with the Auto Dickey Fuller and KPSS tests which checked for stationarity. The steps in this process are showcased in figure 2.

Once the data was made stationary, an autocorrelation function (ACF) and a partial autocorrelation function (PACF) was run against the differenced data. Initial ARIMA models were then estimated by using lags which appeared significant as shown in figure 3.

The PACF was used to determine the lag for the autoregressive component whereas the ACF was used to determine the lag on the moving average component. Additionally, the auto.arima() function in R was ran with the trace argument to give a list of more potential models (figure 4.)

fig.1. Screenshot of Costco stock with astonishing 7-year trend and stable variance

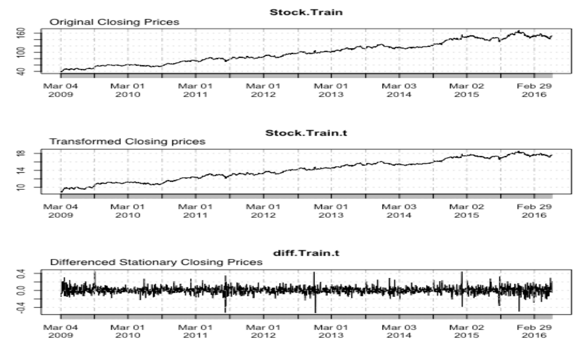

fig.2. Shows original data, transformed data, and stationary data

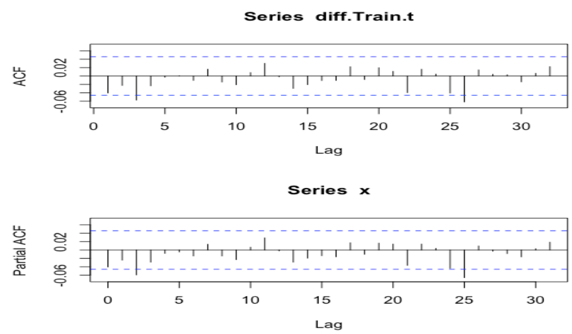

fig.3. Lags which go outside the bordered lines are significant

```
ARIMA(2,1,3) with drift         : 5361.26
ARIMA(3,1,0)                    : 5367.777
ARIMA(3,1,0) with drift         : 5363.686
ARIMA(3,1,1)                    : 5365.141
ARIMA(3,1,1) with drift         : 5359.86
ARIMA(3,1,2)                    : 5366.325
ARIMA(3,1,2) with drift         : 5361.705
ARIMA(4,1,0)                    : 5361.967
ARIMA(4,1,0) with drift         : 5356.976
ARIMA(4,1,1)                    : 5362.785
ARIMA(4,1,1) with drift         : 5358.899
ARIMA(5,1,0)                    : 5363.881
ARIMA(5,1,0) with drift         : 5358.968
```

fig.4. sample list of potential models generated by auto.arima()

The models with the lowest relative AIC's were chosen and fitted to the data. ACF and PACF functions of the residuals for each model were then plotted to assure there was no correlation among errors. Any models which had a significant correlation among residuals were thrown out. The remaining models were then used to conduct forecasts. A matrix was made consisting of the forecasts of each model and the actual prices of the testing data set. The accuracy() function in R was then used to measure the RMSE (equation 5). The model with the lowest RMSE was kept and used to develop a trading strategy. The forecasts given in R provided a set of confidence intervals which specified a range in which the prices would reliably trade within.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

	p.V1	d.V1	q.V1	RMSE.V1
1:	5	1	6	3.786028
2:	4	1	3	3.845111
3:	5	1	3	3.900943
4:	5	1	7	4.025470
5:	6	1	3	4.245033
6:	7	1	7	4.245403
7:	7	1	6	4.257161
8:	3	1	6	4.260969
9:	6	1	1	4.291452

fig.5. shows a list of models and their RMSE in increasing order

III. RESULTS

Additional to the RMSE, the MAPE (equation 6) was used to measure the percent error on the model. That value was then subtracted from 100 to give a percent accuracy of the model. The PPA shown in equation 7 gives a percentage of how often the model falls within the daily highs and lows of the stock. This tells us how frequently the model intersects with real prices.

$$MAPE = \left[\left(\frac{|(y_i - \hat{y}_i)|}{y_i}\right) * 100\right] \quad (6)$$

$$\text{PPA} = \frac{Lows \leq Predictions \leq Highs}{Total\ Number\ of\ Predictions} * 100 \quad (7)$$

It should be noted that RMSE is a scale dependent measure of error, therefore the RMSE should not be used to compare the accuracy of models using different data sets. The reason we have included the RMSE in our table is to aid anyone who may want to replicate our results. Our RMSE values can thus be used as a benchmark.

The charts below showcase the accuracy of the model in predicting the closing prices (Table 1) and daily mean prices (Table 2) for each stock. The closing price predictions for Costco shown in figure 6 forecast both the steepness and direction of the price action with resounding accuracy. The forecasts for the daily mean price of Facebook in figure 7 adequately predict the overall direction of the stock for the month; even predicting the price dip which occurred on day 17. It is important to note that the price data taken from Facebook was primarily in an upward trend. The testing month however, was a period of downward movement. The fact that the ARIMA model was able to make accurate downward predictions in an overall upward trend is further testimony to its effectiveness. The data taken from Smith and Wesson and Netflix had a far less stable variance than that of Facebook and Costco. This could explain why there is slightly less accuracy in predicting these two companies

Stock	Model	RMSE	MAPE	Percent Accuracy (%)	PPA (%)
Facebook	(3,1,3)	3.69	2.61	97.39	30
Costco	(5,1,6) with drift	3.79	1.64	98.36	40
Smith and Wesson	(3,1,4) with drift	2.14	7.65	92.35	20
Netflix	(3,1,4) with drift	5.43	4.90	95.11	10

Table 1. Results for closing prices on all stocks

Stock	Model	RMSE	MAPE	Percent Accuracy (%)	PPA (%)
Facebook	(7,1,7)	3.44	2.39	97.61	30
Costco	(7,1,8) with drift	4.78	2.24	97.76	20
Smith and Wesson	(7,1,7) with drift	2.14	7.65	92.9	16.67
Netflix	(6,1,4) with drift	7.12	5.03	94.96	43.33

Table 2. Results for daily mean prices on all stocks

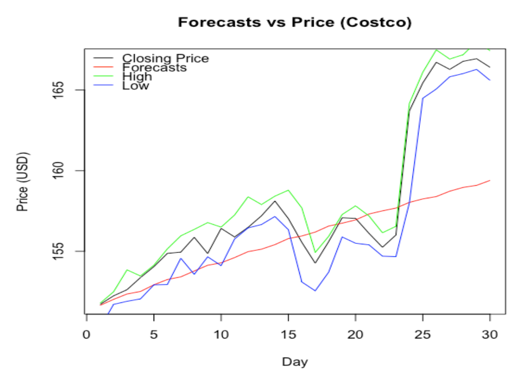

fig.6 Forecasts for Costco show strong overall predictions

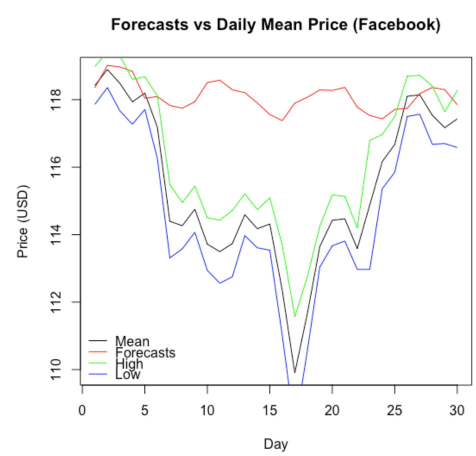

fig.7. Forecasts for daily mean price of Facebook

There are numerous ways in which forecasts from an ARIMA model can be incorporated into a trading strategy. The strategy we developed makes use of Support and Resistance lines combined with Exponential moving averages and chart patterns. After a trader has made forecasts for the next thirty days, he should make a trading plan for the month. The first step will be to examine the direction of the ARIMA forecast. If the overall direction is upwards, the trader should then consult his 20 and 50 day exponential moving averages and locate their nearest crossover. If the crossover is upwards as well, then there is a greater likelihood that the ARIMA model will correctly predict the direction. Next, the trader should search for any chart patterns that may indicate trend continuation or breakout. If the chart patterns also support the claim of the model, the trader can then use support and resistance lines to plan his entry and exit on the trade. We suggest that a buying order be placed on the price of the first significant line of support that intersects with the forecast plus the last known RMSE of the model. A selling order can be placed at an intersected line of resistance minus the last known RMSE of the model. This allows the trader to account for the error of the model.

It is important to practice risk management while trading. stop loss orders should be placed at appropriate lines of resistance/support depending on whether the trader is taking up a long or short position.

IV. CONCLUSIONS

Based on an evaluation of forecasts, we can confidently say that ARIMA models are highly effective trading indicators. Being able to have a general idea of where a stock will be at the end of a thirty-day period is an extremely powerful advantage when investing. That being said, the tests done in this study are largely theoretical. Future projects should involve use of the ARIMA model to predict stock prices on a rolling basis rather than on a predefined set of data. The trading strategy suggested in this paper should also be tested further and confirmed for effectiveness.

ACKNOWLEDGEMENTS

I would like to show my gratitude to Dr. Kim and Mr. Kerr for managing the RMP program and creating a profoundly impactful experience. I would like to thank Mr. Melczer for constantly giving me constructive feedback on this project. Lastly, I would like to sincerely thank my mentor, Mike Andrei for being so generous with his time and guiding me throughout the research process.

APPENDIX

1. Training Data

a. Costco

b. Facebook

c. Netflix

d. Smith and Wesson

REFERENCES

[**1**] Upadhyay, Roopam. "Step-by-Step Graphic Guide to Forecasting through ARIMA." *You Canalytics*. N.p., 28 June 2015. Web. 09 July 2016. <http://ucanalytics.com/blogs/step-by-step-graphic-guide-to-forecasting-through-arima-modeling-in-r-manufacturing-case-study-example/>

[**2**] Moore, Michael Halls. "Autoregressive Integrated Moving Average ARIMA(p, D, Q) Models for Time Series Analysis." *QuantStart*. N.p., 15 Sept. 2015. Web. 09 July 2016. <https://www.quantstart.com/articles/Autoregressive-Integrated-Moving-Average-ARIMA-p-d-q-Models-for-Time-Series-Analysis>.

[**3**] Hyndman, Rob J., and George Athanasopoulos. "Chapter 8 Arima Models." *Forecasting: Principles and Practice*. Heathmont: OTexts, 2016. N. pag. Print.

[**4**] Mondal, Prapanna, Labani Shit, and Saptarsi Goswami. *Study on the Effectiveness of Time Series (ARIMA) in Forecasting Stock Prices*. Institute of Engineering and Management, n.d. Web.

[**5**] Miswan, Nor Hamizah, Nor Azazi Ngatiman, Khairum Hamzah, and Zaminor Zamzamir Zamzamin. "Comparative Performance of ARIMA and GARCH Models in Modelling and Forecasting Volatility of Malaysia Market Properties and Shares." *Applied Mathematical Sciences* (2014): n. pag. Web.