THE MODERN WARFARE WITH THE GAMESTOP STOCK:
FORECASTING THE GAMESTOP FUTURE PRICES

A Project
Presented to the
Faculty of
California State Polytechnic University, Pomona

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science
In
Economics

By
Lance Cao
2021
SIGNATURE PAGE

PROJECT: THE MODERN WARFARE WITH THE GAMESTOP STOCK: FORECASTING THE GAMESTOP FUTURE PRICES

AUTHOR: Lance Cao

DATE SUBMITTED: Spring 2021

Department of Economics

Dr. Craig Kerr
Project Committee Chair
Economics

Dr. Bruce Brown
Economics

Dr. Carsten Lange
Economics
ACKNOWLEDGMENTS

I would like to express my gratitude to Dr. Craig Kerr, chair of our Economics Department Project Committee, for all the patience and help he gives me.
ABSTRACT

The primary purpose of this research is to study the activity of GameStop stock for the recent five calendar years. The goal is to analyze and forecast the future price trend based on the historical data of GameStop by using ARIMA and GARCH. First, the researcher examines the timeline to see how the Twitter of elites would affect the close price. An ARIMA model is estimated to explain a given time series based on GameStop’s past values, and these are in use for forecasting GameStop’s future prices. Additionally, a GARCH model is estimated to predict the future prices of GameStop stock with a minimized volatility effect. The predictions are used to forecast the future prices’ movement of GameStop Stock.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature Page</td>
<td>ii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>iii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>1 Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Literature Review</td>
<td>1</td>
</tr>
<tr>
<td>2 Data</td>
<td>4</td>
</tr>
<tr>
<td>3 Methodology</td>
<td>5</td>
</tr>
<tr>
<td>3.1 ARIMA Forecasting</td>
<td>7</td>
</tr>
<tr>
<td>3.2 GARCH Forecasting</td>
<td>14</td>
</tr>
<tr>
<td>4 Result</td>
<td>20</td>
</tr>
<tr>
<td>5 Conclusion</td>
<td>21</td>
</tr>
</tbody>
</table>
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>GME Stock Summary</td>
<td>7</td>
</tr>
<tr>
<td>3.2</td>
<td>Augmented Dickey-Fuller Test of GME close price</td>
<td>8</td>
</tr>
<tr>
<td>3.3</td>
<td>Auto ARIMA</td>
<td>10</td>
</tr>
<tr>
<td>3.4</td>
<td>AIC and BIC</td>
<td>17</td>
</tr>
<tr>
<td>3.5</td>
<td>Ljung-Box test of the squared standardized residuals</td>
<td>18</td>
</tr>
</tbody>
</table>
List of Figures

3.1 Stock Timeline with Twitter ................................. 6
3.2 GME Stock ................................................. 8
3.3 ACF ......................................................... 9
3.4 PACF ....................................................... 11
3.5 ARIMA Residuals ............................................ 12
3.6 Histogram Zoom Out ........................................ 12
3.7 Histogram ................................................... 13
3.8 Residuals ..................................................... 13
3.9 ARIMA Forecast ............................................. 14
3.10 ARIMA Forecast Zoom In .................................. 15
3.11 Mean Tendency ............................................. 15
3.12 Volatility Plot .............................................. 16
3.13 Normal Residuals .......................................... 17
3.14 Normal Q plot ............................................. 18
3.15 GARCH forecast .......................................... 19
Chapter 1

Introduction

Currently, GameStop has become a talk that has taken over all the news. Based on what is said about the GameStop stock, it creates particular feeling that it would be an interesting attempt studying their stock. Most economists and financial analysts are curious about the stock’s future prices and struggles with the idea on what the stock will look like after many hedge companies short-sells the stock and causing many social elites to express their opinions on social media encouraging individual investors to go against the wolves of Wall Street. Many scholars have conducted studies and given their ideas and opinions regarding the issue of stock exchange, stock market and stock prices. Below is a literature review on some of the relevant information by different scholars concerning the topic.

1.1 Literature Review

Most people focus on the Long-Short term memory on forecasting where they need to find the data to conduct their studies. Some of them believed that an extended short-term memory model such as the one-year model would give better results than either three
years or five years when compared to the ARIMA model (Manurung, Budiharto, and Prabowo, 2018). They used a five years data from Yahoo finance for the ARIMA model as long-term and one year for the LSTM as the short term where they found out that the five years ARIMA model gives a better analysis. It is out of this that the researcher learns on retrieving data for their ARIMA model within the most suitable time frame of the historical data that they probably need for the research. Most of the changes on GameStop happened within a year, hence maintaining five years’ data would be more efficient and accurate for analysis which is not only for the ARIMA model.

Others believes that the artificial neural network would be the best method to predict the stock price. The stock price is a time-variant in a nonlinear one pattern; hence, it is always difficult to predict the future price (Patel and Yalamalle, 2014). These individuals suggest that using algorithms with neural networks such as the Feedforward MLP neural network is the best technique for predicting the stock price of those companies listed under the LIX15 index of the National Stock Exchange. The Artificial neural network is a similar method however, it is an alternative that people would use in forecasting the stock price to ARIMA. In this case, ARIMA is the best option since it gives a more direct answer compared to the Artificial neural network which could generally be viewed as something beyond this level of study. GME does not fall in the suggested category.

While focusing on the neural network, some researchers looked into time delay, recurrent, and probabilistic. The prediction would limit the false alarm in stock to prevent losses (Saad, Prokhorov, and Wunsch, 1998). The research guides the researcher in focusing on some specific factors other than just the future price alone. It would be the lags in the researcher’s ARIMA model; hence lags would be tested by checking the p-values to see the significance, along with the heteroscedasticity.

In the study of stock prices, the dataset is considered as high-frequency data. Besides
focusing on the probabilistic setting and estimation methods, the researchers introduced Geometric Brownian Motion. It is capturing of the stock price movement that would help in capturing the additivity on a log scale (Mykland and Zhang, 2012). GBM is a logarithm for studying the continuous-time stochastic process. It is an excellent method for estimating the close price. In here the researcher can use GARCH after performing the ARIMA, which gives a great idea about considering the volatility in the model.

When the stock price abruptly increased at one particular moment due to a specific social factor, it was observed that it remained substantially low for the rest of the time this means that the average could go higher than the median. Hence, to analyze the dataset, regression is key. It is important to describe the relationship and use a statistical interface (Angrist and Pischke, 2008). This allows the researcher to determine the influence of social factors on close prices which in this case, is elite’ Twitters on drifting the stock price. The purpose is to testify whether the social factor has a significant influence on the close price.
Chapter 2

Data

The data used in the research was directly retrieved from Yahoo finance for the past five calendar years. Despite the fact that the GameStop situation happened within a period of a year, the outcomes would be more accurate once the data for a longer time frame is employed. The historical data of the GameStop stock contains the daily prices for high, low, open, close, and adjusted. The dataset gives precise details concerning the price movement and behavior of the GameStop stock. I investigated tweets about GameStop by elites, which caused the price shift and considered it as the dummy variable 'Event' and added it to the dataset. Those elites are the ones who are well known in different industries and have apparent influence according to what they say and do. The value is zero if there was no such activity on that day and if it was there, the value is one.
Chapter 3

Methodology

After retrieving and editing the historical data from Yahoo finance, it was evident how the elites on Twitter was manipulating the stock price as in Figure 3.1 shows the stock price of GameStop and the timing of tweets by Elon Musk-CEO of Tesla Motors, Ted Cruz-United States Senator, Stephen Colbert-American comedian, Jon Stewart-American comedian, and Alexandria Ocasio-Cortez-United States Representative. They all expressed their opinions about the GameStop stock on Twitter. It seems Elon Musk, Jon Stewart, and Stephen Colbert had a significant influence on booming the price. In contrast, Alexandria Ocasio-Cortez and Ted Cruz influenced easing the situation.

The next step is moving to R to further observe and this involves several steps. The first step being having a brief check of the data with a summary after formatting the data every time as shown in Table 3.1. It is important to know the state of the data before moving on to analysis. The mean is higher than the median, which represents that the spike is extremely high and sudden.

The stock price movement and volume are displayed in Figure 3.2. It is a common method observing the movement of the stock price. These lines indicate the stock perfor-
Figure 3.1: Stock Timeline with Twitter

![Stock Timeline with Twitter](image)

**A lexandria Ocasio-Cortez**

I am happy to work with Republicans on this issue where there’s common ground, but you almost had me murdered 3 weeks ago so you can sit this one out.

Happy to work w/ almost any other GOP that aren’t trying to get me killed.

In the meantime if you want to help, you can resign.

**Jon Stewart**

This is bullshit. The Redditors aren’t cheating, they’re joining a party Wall Street insiders have been enjoying for years. Don’t shut them down...maybe sue them for copyright infringement instead!

We’ve learned nothing from 2008.

Love,

StewBeef

**Elon Musk**

u can’t sell houses u don’t own
u can’t sell cars u don’t own
but
u “can” sell stock u don’t own?
this is bs – shorting is a scam
legal only for vestigial reasons

**Stephen Colbert**

Hedge funds are complaining about losing billions on GameStop, which means it’s a good time to invest in whatever company makes the world’s tiniest violins.
Table 3.1: GME Stock Summary

<table>
<thead>
<tr>
<th>Index</th>
<th>Open</th>
<th>High</th>
<th>Low</th>
<th>Close</th>
<th>Adj.Close</th>
<th>Volume</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>2016-05-18</td>
<td>2.85</td>
<td>2.94</td>
<td>2.57</td>
<td>2.8</td>
<td>2.8</td>
<td>972900</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>2017-08-16</td>
<td>6.65</td>
<td>6.963</td>
<td>6.285</td>
<td>6.68</td>
<td>6.68</td>
<td>2231100</td>
</tr>
<tr>
<td>Median</td>
<td>2018-11-13</td>
<td>15.19</td>
<td>15.46</td>
<td>14.86</td>
<td>15.21</td>
<td>13.97</td>
<td>3205350</td>
</tr>
<tr>
<td>Mean</td>
<td>2018-11-16</td>
<td>23.73</td>
<td>25.136</td>
<td>22.248</td>
<td>23.49</td>
<td>21.82</td>
<td>6530206</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2020-02-17</td>
<td>22.74</td>
<td>23.168</td>
<td>22.438</td>
<td>22.78</td>
<td>18.9</td>
<td>5347575</td>
</tr>
<tr>
<td>Max.</td>
<td>2021-05-17</td>
<td>379.71</td>
<td>483</td>
<td>262.27</td>
<td>347.51</td>
<td>347.51</td>
<td>197157900</td>
</tr>
</tbody>
</table>

3.1 ARIMA Forecasting

It is clear that the ‘Event’ would have a positive influence on the daily close price while ending at a higher level. Based on this, I move onto the ARIMA model and start performing the ADF test for the close price as shown in Table 3.2. The ADF test or the Augmented Dickey-Fuller test is a test used to test whether the given time series is stationary or not. In this case, a stationary time series gives a somehow nonsmall Dickey-Fuller value and a p-value bigger than 0.05. Then it would need to run more tests.

Later on, I perform the ACF and PACF tests on the dataset as shown in Figure 3.3 and Figure 3.4. The ACF test is a correlogram known as the Autocorrelation Function. It shows the serial correlation change over time in time series data. In this case, based on the correlation between the points from the plot, the autocorrelations seemed significant.
Figure 3.2: GME Stock

Table 3.2: Augmented Dickey-Fuller Test of GME close price

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dickey Fuller</td>
<td>(1.853)</td>
</tr>
<tr>
<td>Lag order</td>
<td>10</td>
</tr>
<tr>
<td>p-value</td>
<td>0.641</td>
</tr>
</tbody>
</table>

alternative hypothesis: stationary
for a large number of lags. Then continue the observation by running a PACF test.

The PACF test or the Partial Autocorrelation Function test is ACF with a shorter lag length as in Figure 3.4. The beginning lags all exceed the significant level. In this study, the spikes on the plot determine that the first couple of lag autocorrelations explain the higher-order autocorrelations.

Next step involves the use of Auto ARIMA in finding the best ARIMA model for the time series. Use the Auto ARIMA or the Auto-Regressive Integrated Moving Averages to minimize Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values to determine the parameters. The model factors, along with AIC and BIC, are clearly shown in Table 3.3.

Based on the Auto ARIMA summary, the formula for the ARIMA model is as follows.

\[ \Delta y_t = \sum_{s=1}^{5} \Delta a_s y_{t-s} + \sum_{s=1}^{4} b_s \epsilon_{t-s} \]

Then I check the model residuals and the ARIMA parameters are already selected,
Table 3.3: Auto ARIMA

<table>
<thead>
<tr>
<th>ARIMA(5,1,4)</th>
<th>ar1</th>
<th>ar2</th>
<th>ar3</th>
<th>ar4</th>
<th>ar5</th>
<th>ma1</th>
<th>ma2</th>
<th>ma3</th>
<th>ma4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.472</td>
<td>0.130</td>
<td>(0.136)</td>
<td>0.008</td>
<td>0.136</td>
<td>(0.771)</td>
<td>0.388</td>
<td>0.104</td>
<td>(-0.570)</td>
</tr>
<tr>
<td>s.e.</td>
<td>0.104</td>
<td>0.111</td>
<td>0.072</td>
<td>0.052</td>
<td>0.036</td>
<td>0.103</td>
<td>0.138</td>
<td>0.132</td>
<td>0.083</td>
</tr>
</tbody>
</table>

\[ \sigma^2 \] estimated as 76.79: \text{log likelihood}=4508.76

<table>
<thead>
<tr>
<th>AIC</th>
<th>9037.52</th>
</tr>
</thead>
<tbody>
<tr>
<td>AICc</td>
<td>9037.69</td>
</tr>
<tr>
<td>BIC</td>
<td>9088.88</td>
</tr>
</tbody>
</table>

Training set error measures:

<table>
<thead>
<tr>
<th>ME</th>
<th>RMSE</th>
<th>MAE</th>
<th>MPE</th>
<th>MAPE</th>
<th>MASE</th>
<th>ACF1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.284</td>
<td>8.728</td>
<td>1.823</td>
<td>-0.185</td>
<td>5.671</td>
<td>1.130</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Figure 3.4: PACF

Series close_price

![Plot of Partial Autocorrelation Function (PACF)](image)

as in Figure 3.5. According to the time series model, the residuals are the difference between the observed values and the fitted values. From the figure, there is heteroscedasticity, in which the line is varying compared to the earlier time.

I check the residuals over a normal line as shown in Figure 3.6. The residuals plot will have an extreme curve peak when residuals are close to 0 with great density. When I zoom in the figure as in Figure 3.7, and the residuals curve seems somehow decent when it is not equal to zero.

Then I make the last residuals plot to look at the standardized residuals, ACF of residuals, and p-values for Ljung-Box statistics as in Figure 3.8 with the main focus being on the p-values. The hypothesis of the research is that the dataset points are distributed independently. The p-values indicates that the null hypothesis cannot be rejected, and the dataset points are not correlated at all since all the p-values for all lags are greater than the 0.05 significance level.

At this moment, I have the ARIMA model fitted and applied so that they could pro-
Figure 3.5: ARIMA Residuals

Figure 3.6: Histogram Zoom Out

Histogram of ARIMA.Residuals
Figure 3.7: Histogram

Histogram of ARIMA.Residuals

Figure 3.8: Residuals

Standardized Residuals

ACF of Residuals

p values for Ljung–Box statistic
ceed to predict the future price for 30 days over the real train set of the stock price as in Figure 3.9 and a better look by zooming in on Figure 3.10. The blue line represents the prediction.

The mean tendency as in Figure 3.11, is the mean prediction tendency over the real close price. Based on the figure, it shows a future direction of the future prices.

3.2 GARCH Forecasting

Irrespective of the fact that the ARIMA model gives a reasonable forecast, the volatility effect is not minimized yet, which is a stock with higher volatility would remain uncertain to the forecasted price and higher risk for the future investment. The volatility is measured by using logarithmic returns standard deviation. It is to determine the degree of the variation in its trading price over a given time. So, the GARCH is the best ideal approach. It is known as the Generalized Autoregressive Conditional Heteroskedasticity and is a statistical modeling method that predicts future prices, including volatility.
Figure 3.10: ARIMA Forecast Zoom In

Forecasts from ARIMA(5,1,4)

Figure 3.11: Mean Tendency
Here, I need to apply the model to the close price, then find the Auto Regressive Fractionally Integrated Moving Average model (ARFIMA) parameters. ARFIMA are the parameters within the Auto Regressive Fractionally Integrated Moving Average model, that many people know as the long-memory model. ARFIMA incorporateS them into the GARCH model and have the model fitted for a volatility plot as in Figure 3.12.

It is clear that there is an unstable peak; hence, take a closer look at the Akaike and Bayes of the model as in Table 3.4. Akaike and Bayes are the selection criteria that address the model selection. Akaike measures how well the estimated statistical model fits, while Bayes has different model selection parameters among all parametric models. AIC models are not true to the model, but the BIC comes across true models only, and they are constantly being used for consistent estimation.

Based on the information, I plot the normal residuals as in Figure 3.13. The residuals are distributed at different levels when the Index is 1200, which is the time order.

Then I calculates the standardized residuals and give the plot as in Figure 3.14. There
Table 3.4: AIC and BIC

<table>
<thead>
<tr>
<th>Method</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akaike</td>
<td>2.146</td>
</tr>
<tr>
<td>Bayes</td>
<td>2.174</td>
</tr>
</tbody>
</table>

Figure 3.13: Normal Residuals
Figure 3.14: Normal Q plot

Table 3.5: Ljung-Box test of the squared standardized residuals

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X-squared</td>
<td>0.840</td>
</tr>
<tr>
<td>df</td>
<td>1</td>
</tr>
<tr>
<td>p-value</td>
<td>0.360</td>
</tr>
</tbody>
</table>

are some extreme values that do not follow normal distribution, and this is acceptable as long as most of the data are in the line.

Then I would perform a Ljung-Box test on the squared standardized residuals based on the standardized residuals.

I move on and performs a Ljung-Box test on the squared standardized residuals based on the standardized residuals. Based on the p-value of the Ljung-Box test in Table 3.5, it is clear that the squared standardized residuals do not reject the null hypothesis, which means that there is no autocorrelation. Hence the volatility and residuals are found, the
Figure 3.15: GARCH forecast

The researcher forecasts the future price for the following 30 days using the GARCH model as shown in Figure 3.15.
Chapter 4

Result

The timeline in methodology shows that the elites have social power to influence the market on shifting the stock price and with the help of R by applying ARIMA and GARCH model to the dataset, both predictions show that the future price will decrease but in different patterns, considering the volatility and heteroscedasticity.
Chapter 5

Conclusion

From the timeline graph, it is evident that the Event would drift the daily close prices. Based on the prediction from the ARIMA and GARCH, the close price would decrease inevitably.

Most analysts and market investors usually perform ARIMA and GARCH while forecasting future prices. It is crucial to always look into the actual cause that manipulates the stock price and the direction that the price would go and it would be great if researchers would dig more into the GameStop stock. Other than ARIMA and GARCH, they could compare the Artificial neural network model with ARIMA and compare the geometric Brownian motion model with GARCH. They could also perform the Prophet Forecasting, or KNN regression time-series forecasting. For the Prophet Forecasting, it would forecast the time series data based on an additive model where the nonlinear trends fit daily. The KNN regression is an algorithm used to estimate the relationship between the independent variable and the continuous outcome by averaging the observations in a non-parametric method. These two regressions would give a forecast from some other angles.
References


