GameStop stock Price Prediction using ARIMA model

Alina Armijo

Hunter College CUNY

2021

Table of Contents

Introduction

GameStop Corp. is an American video game, consumer electronics, and gaming merchandise retailer. The company is headquartered in Grapevine, Texas (a suburb of Dallas), United States, and is the world's largest video game retailer, operating 5,509 retail stores throughout the United States, Canada, Australia, New Zealand, and Europe as of February 1, 2020. The company was founded in Dallas in 1984 as Babbage's, and took on its current name in 1999. The company declined during the mid-late 2010s due to the shift of video game sales to online storefronts and failed investments by GameStop in smartphone retail.

In 2021, the company's stock price skyrocketed due to a short squeeze orchestrated by users of the Internet forum r/wallstreetbets. The company received major media attention during January and February 2021 due to the volatility of its stock price. In early March 2021, the company's stock price rose significantly again, likely due to changes in its executive-level staff and prospects of a changing business model.[Ref1]

Dataset

2.1 Data Overview

The raw data set is taken from kaggle.com and includes daily GameStop information from February 2002 until January 2021. It includes seven variables, shown in Table 2.1, and 2416 observations that are in chronological order. Each observation includes price information and volume information for every 24 hour step.

Table 2.1 Variables Description

Variables	Meaning
Open_price	Price of the first trade at the current time step
High_price	Highest Price of trades at the current time step
Low_price	Lowest Price of trades at the current time step
Close_price	Price of the last trade before the next time step
Volume	The amount of stocks traded during that day
Adjclose_price	The adjusted close price
Date	The date when the prices were recorded
	*All prices are in U.S. dollars (USD)

First of all we need to prepare the data. There are no missing values which indicates that trade occurred every single day within a given period. I decided to analyse close price variable as time series so I created the data frame which consists of only date and close_price variables.

2.2 Exploratory Data Analysis

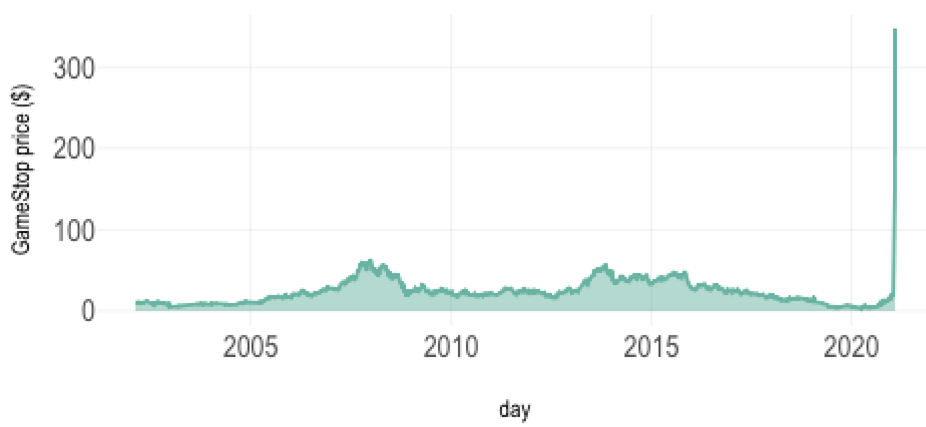
As I already mentioned that GameStop stock becomes extremely popular in 2021. Now let's see how volatile the stock is. From Table 2.2 we can see that its closing price rose from 2.80 dollars to 347.51 dollars which is 12292.9% increase. We also see a huge increase in trading volume happening probably due to attention of users of the Internet forum r/wallstreetbets and media.

Table 2.2 Raw Dataset Distribution

Variables	Min	Max	Mean	Standard Deviation	Median
Open_price	2.85	354.83	23.20	14.70	21.76
High_price	2.94	483.00	23.69	16.21	22.11
Low_price	2.57	249.00	22.71	13.63	21.43
Close_price	2.80	347.51	23.19	14.51	21.76
Adjclose_price	2.56	347.51	16.88	11.02	15.48
Volume	65000	196784300	3398255	6547035	2491800

Now, let's see when fluctuations in price have occurred. According to Figure 2.1 at the early stage price was very low and pretty stable. A huge price jump happened in January 2021.

Figure 2.1: Price of GameStop



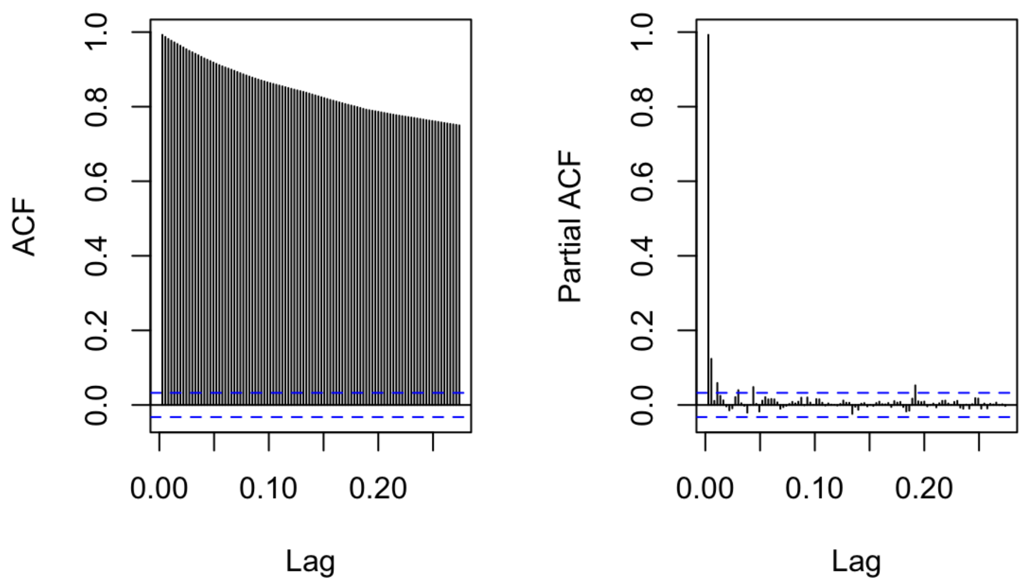
In January 2021, a short squeeze of the stock, causing major financial consequences for certain hedge funds and large losses for short sellers. Approximately 140 percent of GameStop's public float had been sold short, and the rush to buy shares to cover those positions as the price rose caused it to rise even farther. The short squeeze was initially and primarily triggered by users of the subreddit r/wallstreetbets, an Internet forum on the social news website Reddit, although a number of hedge funds also participated. Short selling is a finance practice in which an investor, known as the short-seller, borrows shares and immediately sells them, hoping to buy them back later ("covering") at a lower price, return the borrowed shares (plus interest) to the lender and profit off the difference. The practice carries an unlimited risk of losses, because there is no inherent limit to how high a stock's price can rise.

On January 28, some brokerages, particularly app-based brokerage services such as Robinhood, halted the buying of GameStop and other securities, citing the next day their inability to post sufficient collateral at clearinghouses to execute their clients' orders. This decision attracted criticism and accusations of market manipulation from prominent politicians and businesspeople from across the political spectrum. Dozens of class action lawsuits have been filed against Robinhood in U.S. courts, and the U.S. House Committee on Financial Services held a congressional hearing on the incident. In reaction to brokerages halting the buying of GameStop and other securities, the total market capitalization of cryptocurrencies and metal futures increased as well. [Ref2]

2.3. Data Preprocessing

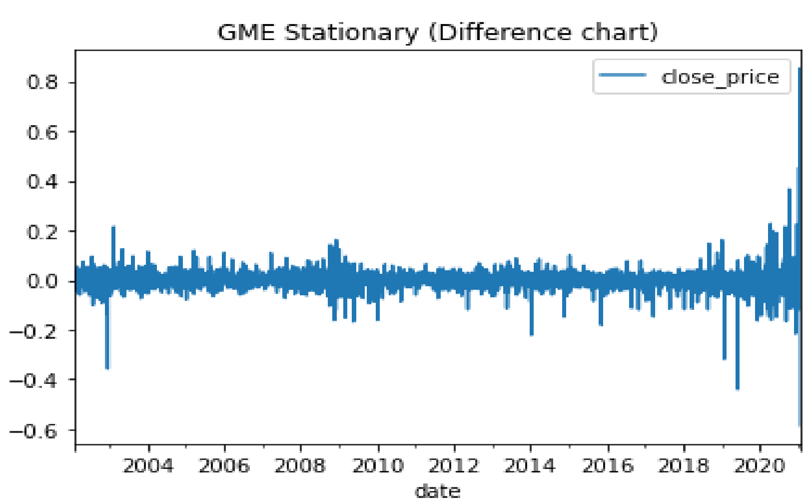
The ACF of the data on Figure 2.2 decays gradually, whereas the PACF cuts off abruptly. As suggested in [Ref3] if ACF dies down, PACF cuts off after lag n: ARIMA(n,d,0) is appropriate to consider. Also the gradual decay of the ACF tells us that we should difference the data, which would also address the issue of the upward trend. Also we are going to use log transformation of the data.

Figure 2.2 ACF and PACF



Since one of the assumptions when using the ARIMA model is having constant variance, differencing will help us with that. We will take the first difference between one period of time(t) and the period of time before that (t-1). Figure 2.3 shows that there is no obvious trend in the time series plot after first differencing. However there is a huge volatility in price change of GameStop in the beginning of 2021.

Figure 2.3 Time Series plot of the first difference

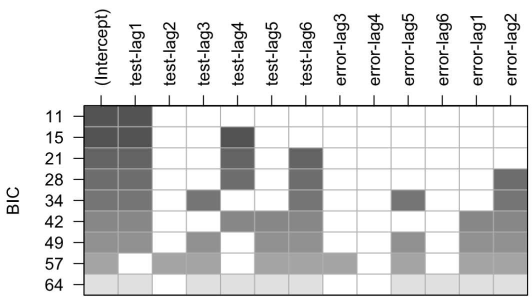


Implementation

3.1 ARIMA Model Selection

From Figure 3.4 we see the set of possible models from BIC and EACF is {ARIMA(0,1,0), ARIMA(1,1,0), ARIMA(2,1,0), ARIMA(3,1,1), ARIMA(3,1,2), ARIMA(9,1,10), ARIMA(8,1,4)}

Figure 3.4 BIC and EACF tables of transformed data




```
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x o o o o o o o o o o  o  o  o
1 x o o o o o o o o x o  o  o  o
2 x o o o o o o o o x o  o  o  o
3 x x x o o o o o o o o  o  o  o
4 x x x x o o o o o o o  o  o  o
5 x x x x o o o o o o o  o  o  o
6 o x x x x o o o o o o  o  o  o
7 o o x x x x o o o o o  o  o  o
```

3.2 ARIMA Modeling

After comparing different models from the set of seven suggested models I concluded that ARIMA(3,1,1) is the most appropriate based on the lowest AIC criteria Figure 3.5.

Model	**AIC**
(0,1,0)	-6734.85
(1,1,0)	-7885.99
(2,1,0)	-8335.00
(3,1,1)	-9376.71
(3,1,2)	-9374.76
(9,1,1)	-9374.15
(8,1,4)	-9371.59

On Figure 3.6 we can see a summary of the chosen model ARIMA(3,1,1) with

AIC=-9376.71 and Sigma^2 is 0.00445.

Figure 3.6 ARIMA(3,1,1)

```
Call:
arima(x = diff.data.t, order = c(3, 1, 1), method = "ML")

Coefficients:
          ar1     ar2     ar3      ma1
      -0.0349  0.0068  0.0044  -1.0000
s.e.   0.0166  0.0166  0.0166   0.0014

sigma^2 estimated as 0.00445:  log likelihood = 4692.35,  aic = -9376.71
```

So we have the ARIMA(3,1,1) model

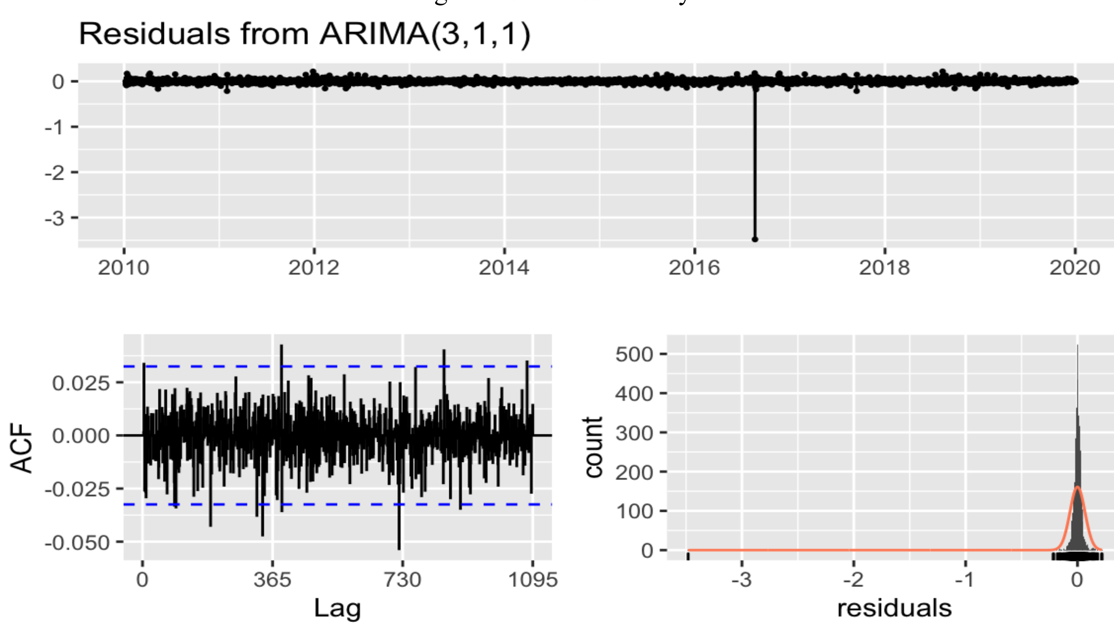
$$X_t + 0.0349X_{t-1} - 0.0068X_{t-2} - 0.0044X_{t-3} = Z_t + 1.0000Z_{t-1}$$

where Z_t should be approximately white noise with mean 0 and variance σ_Z^2.

3.3 Diagnostics of the Model

The model that fits the data should be tested to determine the quality of the fit. Figure 3.7 shows Residuals Analysis for ARIMA(3,1,1). Thus we can confirm that the residuals are not distinguishable from a white noise series as the results are not significant.

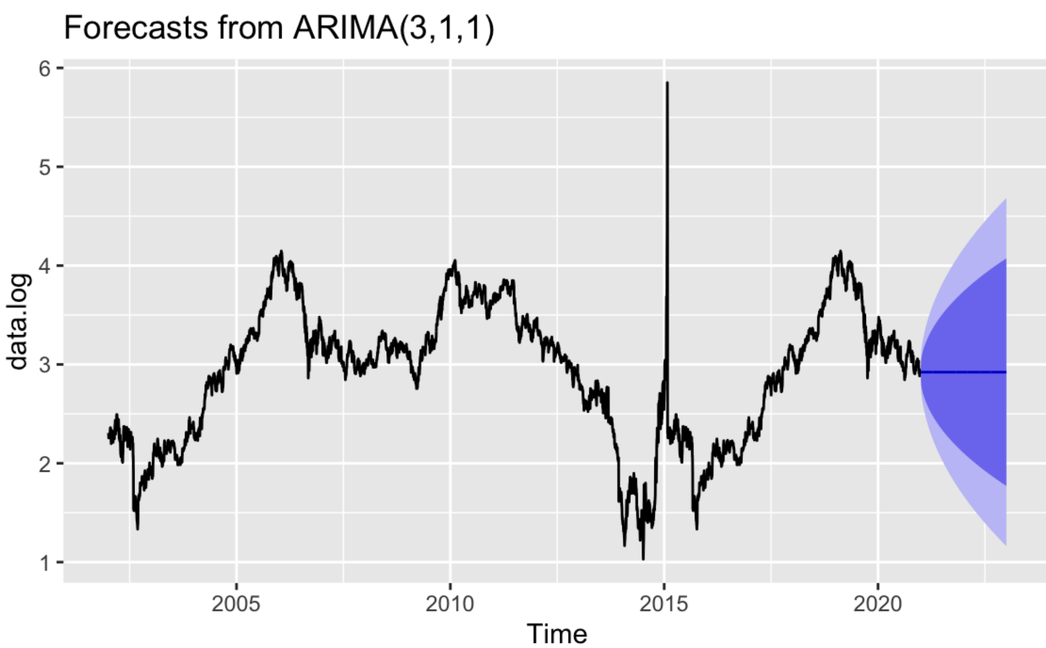
Figure 3.7 Residual Analysis



3.4 Forecasting

Figure 3.8 shows forecasting of ARIMA(3,1,1) with 80% and 95% prediction intervals. We do observe that while the forecasts predict a straight line, which the true values obviously did not follow, at the very least the confidence intervals of the predictions capture the true values.

Figure 3.8



Conclusion

The goal of this project was to find a suitable ARIMA model to predict the closing price of GAmeStop stock. The process included exploratory analysis, data visualization, model specification, model fitting and selection, followed by diagnostic checking and forecasting.

Ultimately, my lowest AICc model produced predictions that were relatively close to the true data, with the ARIMA(3,1,1) model. This model passed model diagnostics and residuals considered to be white noise. The biggest challenge for me was when I tried to create the difference plot and forecast. For some reason coding it in R was creating repeated portions of graphs, in other words my data was ordered wrong and repeated as well. I tried to reorder and check the code and even reordered in Excel. When I coded these plots in Python there were no problems. Therefore my goal in the future is to better learn Python.

If I would continue working to accurately forecast on this dataset, I would try to utilize a GARCH model instead, which better takes into account the non-constant variance. GARCH is a typical model for financial data.

References

[Ref1] https://en.wikipedia.org/wiki/GameStop

[Ref2] https://en.wikipedia.org/wiki/GameStop_short_squeeze

[Ref3] "Time Series analysis with ARIMA-ARCH/GARCH model in R" by L-Stern Group, Ly Pham