Case Study 4, Unit 8 - Time Series
Daniel Davieau, Lei Jiang, Kristen Rollins
3 March 2020

Introduction

Predicting the future may seem like a hopeless exercise, but this is not far off from the goals of time series modeling and forecasting. Anticipating what will happen in the future would be invaluable to any business, or even personal applications such as making investments. While much randomness is present in cases like the stock market, for example, time series models can still identify patterns in the data to make reasonable forecasts. One of the most popular traditional time series models is ARIMA, which is made up of AutoRegressive, Integrated, and Moving Average components. As long as we believe that the future will be reflective of past behavior, these components work together to propagate that behavior so we can make powerful predictions in countless applications.

Background

In this case study, we will use the Gilead Sciences, Inc. (NASDAQ: GILD) stock price from Yahoo starting from October 1st, 2015 through February 1st, 2020. Gilead is a biopharma company based in California that focuses primarily on antiviral drugs for treating HIV, hepatitis B, hepatitis C, and influenza. Gilead's experimental drug *remdesivir* is currently under clinical trials in both the U.S. and China for treating patients with the 2019 novel coronavirus (COVID-19) infection. The preliminary results are very promising. Thus, it is interesting to examine how the stock price could be influenced by the undergoing clinical trials.

Stock Price Analysis

Intuition Model

We first wanted to create an ARIMA(p,d,q) model based on our intuition and understanding of ARIMA models. We examined the raw stock data for the GILD symbol, shown in Figure 1, and observed strong trend and wandering behavior. This behavior, along with the strong positive autocorrelations of the raw data shown in Figure 2, led us to believe that differencing was necessary. Furthermore, the Dickey-Fuller test failed to reject its null hypothesis that the series was non-stationary ($p = 0.183931$).

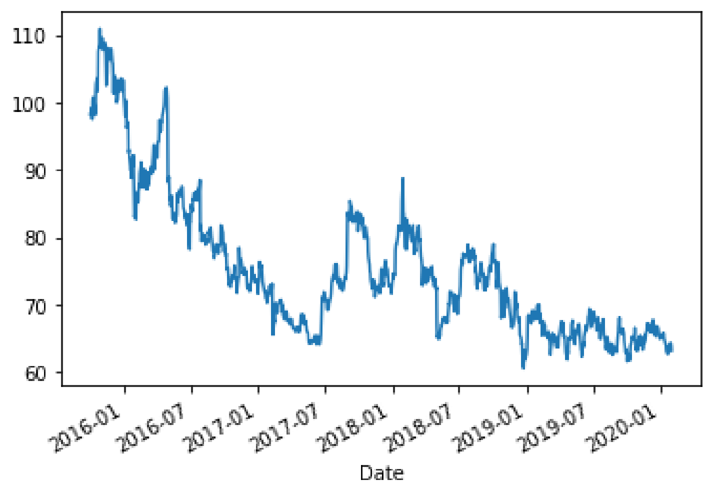

Figure 1. Raw Stock Time Series for GILD Symbol

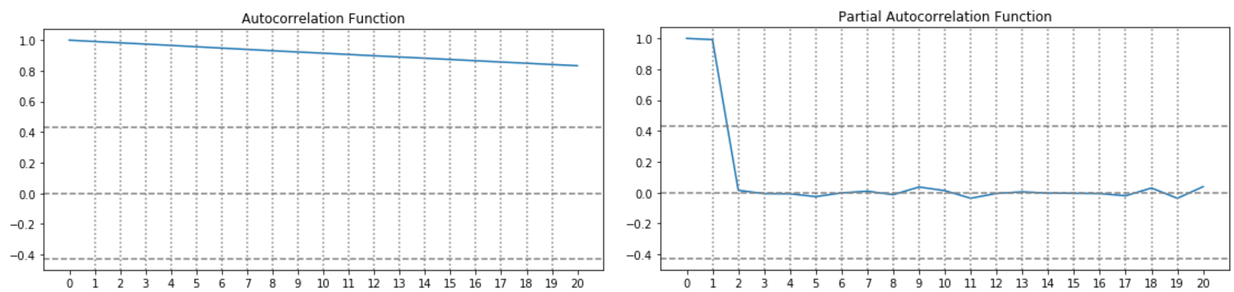

Figure 2. ACF and PACF for Raw Stock Data

Because of the strong indications of non-stationarity, we decided to difference our stock data. Figure 3 depicts the results of taking the first difference of our raw data along with a 30-point rolling mean and standard deviation. This differenced data appears to have a fairly constant mean (at 0), relatively consistent variance (though not perfectly constant, not erratic enough for concern), and does not appear to change autocorrelation structures within the time series. Thus we assume this data is stationary, as it appears to meet the three conditions for stationary (constant mean, variance, and ACF). Furthermore, the Dickey-Fuller test produced a p-value of practically 0, indicating strong evidence for stationarity of this differenced data. Since we are wary of overdifferencing as well, this evidence for stationarity after only one difference factor suggested that we should not continue differencing, and thus we selected d=1.

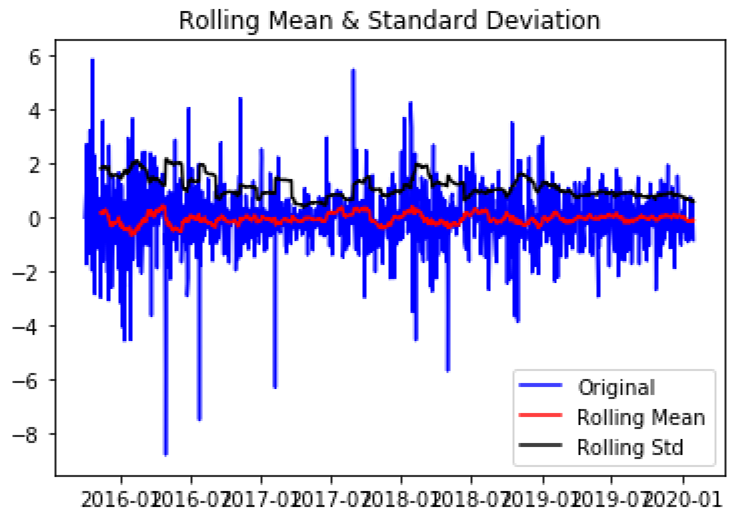

Figure 3. Differenced Time Series Data

After determining that a d=1 factor was necessary, we continued on to estimate the p and q parameters for our ARIMA(p,d,q) model based on our intuition. For the AR (autoregressive) parameter p, we definitely have dependence on the past, therefore we are at a minimum of p=1. It is also very likely that we are dependent on more than one lag, since intuitively when stocks are trending up or downward investors react to it. Selling or buying the stock presumably affects the price at closing. Furthermore our partial autocorrelation plot showed a sharp cutoff and lag 1 is positive (Rule 6). With that said we assume at least p=2. For the MA (moving average) parameter q, we intuitively believe that the moving average has an influence on the stock price. However if we have p=2 then we probably do not want q to cancel it out (Rule 8). For now we will make q=1 based on this rule of thumb. Therefore, our model we would choose based on intuition is an **ARIMA(2,1,1)** model.

Note that we did not readily observe strong seasonality in the data, but this could be investigated further (as we have the hypothesis that stocks behave differently at the end of fiscal quarters), and models besides ARIMA that can incorporate seasonality could be considered.

Grid Search

Although we could create an ARIMA model based on our intuition, we wanted to see if this was reflected in model accuracy when many combinations of parameters were considered. Therefore we ran a grid search of parameters, where we allowed p to range from 0 to 3, and d and q could range from 0 to 2 (inclusive). We performed this grid search on both our 1st differenced data as well as the raw stock data to compare results. The metric we used to compare the various models was RSS, which we believe denotes Root Sum of Squares based on the code. It is calculated by squaring each of the errors (i.e. predicted - actual), summing those values up, and taking the square root of that quantity.

When we first ran a grid search on our manually differenced data, the best parameters were p=2, d=0, and q=2. However since we had already differenced the data, this should be equivalent to an **ARIMA(2,1,2)** model on the original data. This model resulted in an RSS of 39.73. We are cautious of this model because we do not want p and q to cancel each other out.

Next we ran a grid search on our raw stock data. We hoped that this would produce equivalent results to our first grid search, i.e. an ARIMA(2,1,2), but it actually chose an **ARIMA(3,0,1)** model and obtained an RSS of 45.08. We are not convinced of this model either, because we are confident that we require differencing based on our analysis above.

Final Model

The competing models we gathered based on our intuition and grid searches are ARIMA(2,1,1), ARIMA(2,1,2), and ARIMA(3,0,2) models. First, we chose to eliminate the ARIMA(3,0,2) model because we are confident that our raw data requires differencing because of its high autocorrelations for many lags (Figure 2). Next we considered the ARIMA(2,1,2) model chosen by our first grid search. We are suspicious of this model because the p and q terms could cancel, so we examined output for the AR and MA terms that were estimated for these parameters, which are shown in Table 1. The columns we are concerned with are the real and imaginary components; if an AR and an MA factor have identical real and imaginary values then those terms will cancel each other out. Here we see that the 2nd AR and 2nd MA terms are extremely close to each other, only 0.0002 apart. We believe this is close enough for the behavior of the model to act as if the terms are actually canceling out. Thus our suspicions are confirmed for this model, so we will follow the rule of thumb that an AR parameter greater than 1 should only be combined with a 0 or 1 MA parameter (or vice versa).

Table 1. AR and MA Parameters for ARIMA(2,1,2) Model

Roots

	Real	Imaginary	Modulus	Frequency
AR.1	1.0340	+0.0000j	1.0340	0.0000
AR.2	-1.0384	+0.0000j	1.0384	0.5000
MA.1	1.0156	+0.0000j	1.0156	0.0000
MA.2	-1.0386	+0.0000j	1.0386	0.5000

These findings led us to choose our initial intuitive model of ARIMA(2,1,1) as our final parameters. The performance from this model was very close with an RSS of 39.79 (versus 39.73 for the q=2 model), and it does not break any of the rules for ARIMA models. Figure 4 depicts the forecasting behavior from this model after manual differencing (blue is differenced data, red is forecasts). While the model does not appear to capture much of the variance in the data, we should not expect perfect results because there is so much randomness present in the data.

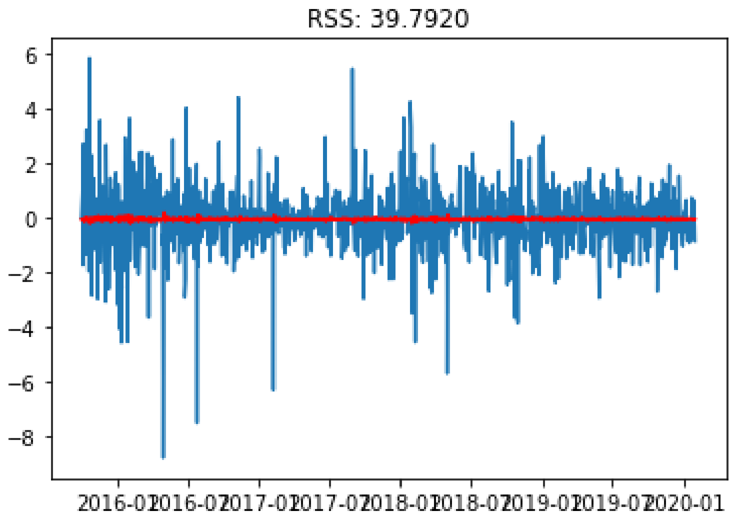

Figure 4. Forecasts for ARIMA(2,1,1) Model (Manually Differenced First)

Conclusions

In this study we found that estimating the parameters for ARIMA models can be done in a number of ways, but using our intuition seems to be a safer method than letting a brute force method decide. This is because some combinations of parameters may not actually fit the characteristics of our data or may lead to models with unfavorable characteristics (such as terms canceling out). The intuitive ARIMA(2,1,1) model that we chose was very close to the parameters and performance of the best model produced by a grid search, and it did not have any undesirable attributes.

Something we also noticed was that manually differencing the data impacted the performance of the models greatly. In theory, differencing the data introduces a d=1 term, and thus an ARIMA(2,0,1) model fit to the already differenced data should be equivalent to an ARIMA(2,1,1) fit to the raw data. However, as shown in Figure 6 in the Appendix, this was not the case, since the ARIMA(2,1,1) performed very poorly (as did all models with a d>0 parameter, when fit to the original data). It is unclear to us why the theoretically equivalent models performed so differently when fit to manually differenced and raw data, and we wonder if this is due to some nature of the implementation of the ARIMA functions in Python. Furthermore we are curious as to why the ARIMA(3,0,1) shown in Figure 7 seems to visually capture the trend of the data better than our chosen model, even though it was a theoretically inappropriate model.

Finally, although the forecasts from ARIMA were not overly impressive in the case of our stock data, this model is the foundation for other types of models that may incorporate additional information. For example, ARUMA or SARIMA models have similar structures but can incorporate seasonality, which we would like to investigate since we have the hypothesis that stock prices have different behavior at the end of fiscal quarters. Another extension of ARIMA are VAR models, which can incorporate multiple concurrent time series to provide even more information to attain more accurate predictions. Thus we were glad to get the experience working with ARIMA models because they have many extensions that can be very powerful in numerous applications.

References

- Python code provided by professor Blanchard
- Rules of thumb from professor Blanchard's slides

Appendix

Supplemental Figures/Tables

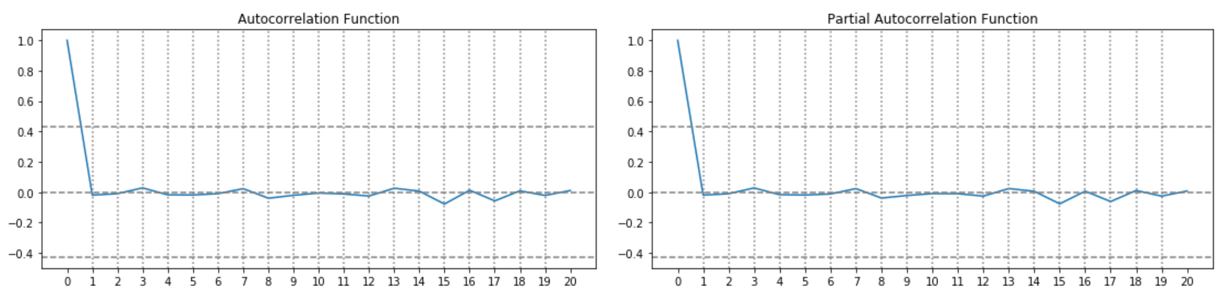

Figure 5. ACF and PACF for Differenced Data

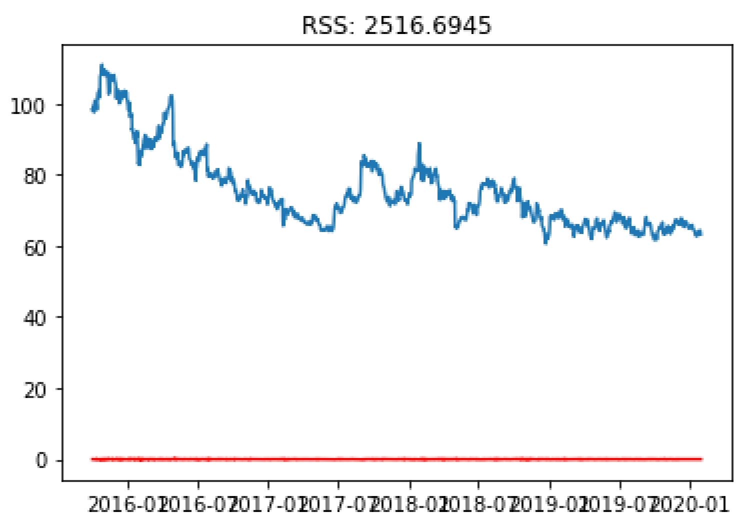

Figure 6. ARIMA(2,1,1) Fit to Original Data (Not Manually Differenced)

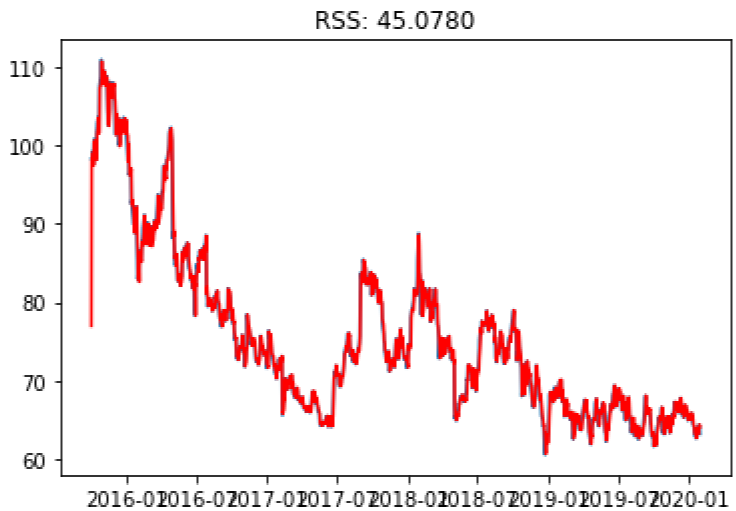

Figure 7. ARIMA(3,0,1) Fit to Original Data

Codebase

Our code has been included in a zipped folder called lJiang_kRollins_dDavieauCase4_Code.zip. The files in this folder include:

- lJiang_kRollins_dDavieauCaseStudy4.ipynb - our primary code base
- modifiedinputs17and18.ipynb - primarily a copy of our code base with a couple plots added

Assignment

- Pick a stock and get 4 years worth of data
- Try and estimate the parameters p, d, q using techniques discussed in class
- Do a grid search for parameters
- What is your final decision on parameters and why