*Utilizes time-series methods such as short term forecasting, long term forecasting, and various regression methods like matrices and leverages.*

**Project**

**3**

P3

ALY6050 Intro to Enterprise Analytics

Project 3 – Forecasting Financial Time Series

**PREPERATION:**

By: John DiSessa

For: Professor Behboudi

On: March 13th, 2022

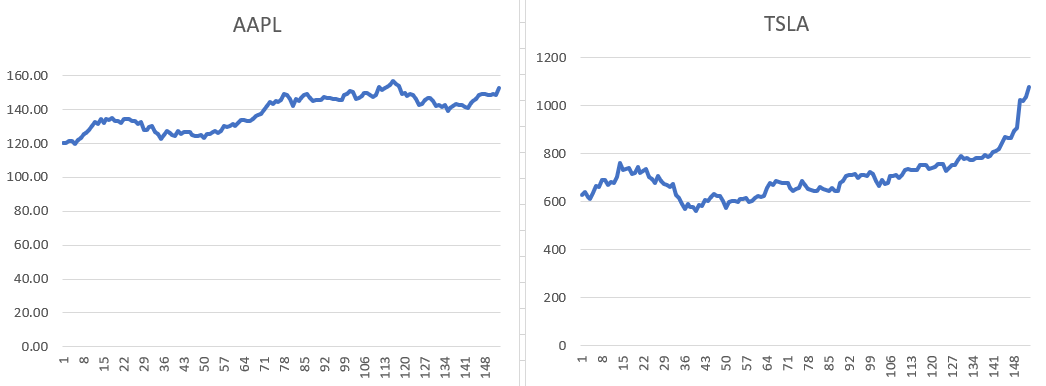
Introduction

We were given historical stock data for AAPL and TSLA for 153 days between March and October 2021 with the goal of forecasting each stock price in the future. Not only did we create accurate short term and long term forecasts for each stock, but identified correlations between the two companies and between our predicted model values to ensure accuracy and be transparent about our model’s short comings. Based on our analysis, we developed portfolio recommendations.

Analysis

**Short Term Forecasting**

Before analyzing these time series, we plotted our data in order to detect any seasonal, irregular, or trend behaviors. Both stocks had a similar first 2 months from March to May with relative peaks in April before dipping back down in May and June. This could potentially indicate seasonality. In the summer, the stocks exhibit slightly different patterns. AAPL tends to trend slightly upward with small volatility. TSLA seems to trend upward before exploding in the fall which shows irregularity.



Since we now understand the data, we created short term forecasts to predict each stock price for 10/29/2021 (day 154), which is the first day for which we do not have data. We performed adjusted exponential smoothing with various smoothing (α) and trend factors (β) in order to find the forecast with the least amount of error. We checked high and low smoothing factors to show high or low sensitivity in price changes (higher smoothing factor allows for larger spikes). We also checked high or low trend factors to show lower or higher resistance to short term prices changes (higher trend factor emphasizes recent values more). We tested the following combinations of smoothing and trend factors on both AAPL and TSLA stock forecasts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **α** | .15 | .25 | .75 | .75 |
| **β** | .15 | .75 | .25 | .75 |

After running 8 forecasts (4 combinations of α and β for both stocks), we calculated the Mean Absolute Percent Error (MAPE) of each forecast to see how far each forecast deviated from their actual values.

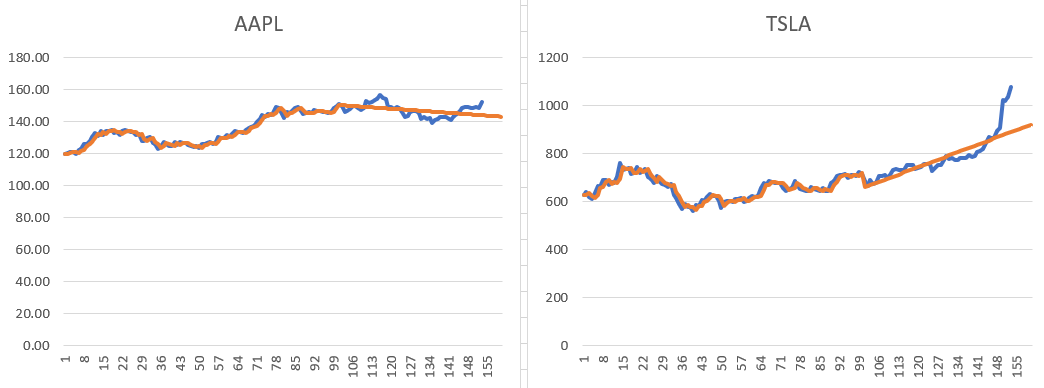
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **AAPL** | 1.9% | 1.3% | *1.0%* | 1.1% |
| **TSLA** | 3.6% | 2.5% | *1.9%* | 2.4% |

From the table above, we determined that an α of .75 and a β of .25 created the most accurate forecast since that combination yielded the lowest percent error for both stocks. The high alpha of .75 makes sense because our historical data indicated many day-to-day changes and our forecast would need to account for these daily peaks and troughs. The stock market is naturally volatile on a day-to-day basis no matter which stock is analyzed. On the flip side, a low beta of .25 also makes sense because we don’t want the trend of the data to overpower the natural day-to-day volatility. However, we can still see over a long enough period of time that all of the volatility averages out and leads to a general uptrend, which is why we still need to, at least a little bit, account for that. With more time or more advanced excel functions, we could test every possible combination of smoothing and trend factors and optimize the values to yield the absolute possible lowest MAPE.

Using our most accurate short term forecast, we can predict an AAPL price of $152.52 ±1% and a TSLA price of $1,093.89 ±1.9%.

**Long Term Forecasting**

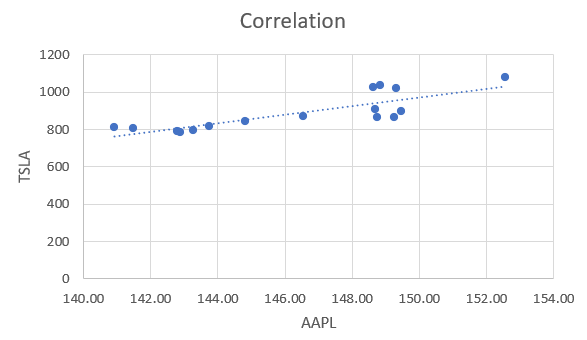
Even though short term forecasting is very useful, it is hard to make financial decisions using only a forecast of the next day’s prices. To enhance our portfolio evaluation, we created long term linear forecasts for AAPL and TSLA by forecasting out to 11/4/2021 (period 160). To ensure our forecast is accurate, we started our forecast back at 8/16/2021 (period 101) despite having actual data for those dates. Using an alpha of .75, we created these linear long term forecasts which project AAPL to be $143.30 ±1.4% and TSLA to be $921.47 ±2.5% on 11/4/2021.



Our Mean Absolute Percent Errors for each forecast are 1.4% for AAPL and 2.5% for TSLA. Since TSLA’s price is larger than AAPL’s, we naturally expected a larger error. Compared to our short term forecast errors (1.0% and 1.9%), we have larger errors for the long term forecast. Again, this was expected because as time increases, uncertainty increases. Even though our linear forecast seems to have the actual data bounce around it, we miss the day-to-day accuracy in favor of the longer term trend.

**Regression**

Diving deeper into our portfolio evaluation, we used regression techniques in order to see if we were able to predict TSLA (y) prices given certain AAPL (x) prices as well as identify any correlations between these stocks. In particular, they seemed to be very similar in October. So, we calculated the correlation between them and plotted their prices for periods 138 – 153.

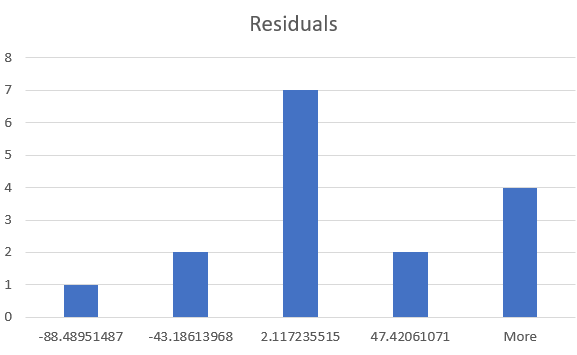


Based on the scatter plot, they do seem to behave similarly. With a correlation coefficient of .82, we can confirm that these assets were strongly and positively correlated in October of 2021. As AAPL increased, TSLA increased. With an R2 value of .67, 67% of the variation in TSLA stock can be explained by AAPL stock.

There could be many potential reasons why these two stocks are correlated, however we think the primary cause is simply due to their size and scale. Both companies are considered “tech” companies and are both in the top 10 in the world in terms of market capitalization. Any experienced investor would feel comfortable trading these stocks because each of these company’s risk of failure is low, and each contains high upside due to each of their advances in technology. An investor who likes tech stocks but doesn’t want to invest all their capital in one company, would likely invest in other top tech stocks with these two having similar value propositions. September is widely known as a poor month to invest in the stock market with October reaping the benefits as the market pushes back up. Many large hedge funds have fiscal years ending in Q3 so they look to sell off their losing positions to save on capital gains taxes and lock in their yearly performance ratings. In times of uncertainty, investors managing their own and other people’s money look to minimize risk. Since these stocks are top performing stocks of the last 5 or more years, they are generally regarded as safe bets.

Next in our analysis, we fit a linear model to our paired data, treating AAPL as our x variable and TSLA as our y variable. Since we are creating a linear model with 1 independent and 1 dependent variable, we can use the formula y = mx+b. We need to calculate the intercept and slope of our model. Once we do that, we can plug in AAPL stock prices and predict TSLA stock prices. We used linear algebra, multiplied matrices together, and calculated our intercept to be -2479.87 and our slope to be 23.01. Our model: TSLA stock = 23.01\*AAPL stock -2479.87. If we use our model to calculate our predicted TSLA values given each AAPL value for periods 138 – 153, we can calculate our residuals and see how accurate this model is and to see if a linear model is appropriate to use by fitting the data.

**Residual Analysis**

We calculated our 16 residuals by using the difference between the actual TSLA price for each of the 16 periods and our predicted TSLA price from our model. The sum of all these residuals was 0 so we can confirm that our model perfectly averages the positive differences with the negative differences. Our model has an acceptable amount of error with an SSE of 46274 and an MSE of 3005. In order to ensure that a linear model is appropriate to use for our data, we need to conduct residual analysis by plotting a normal probability plot of the residuals and performing a chi-square test for normality. 

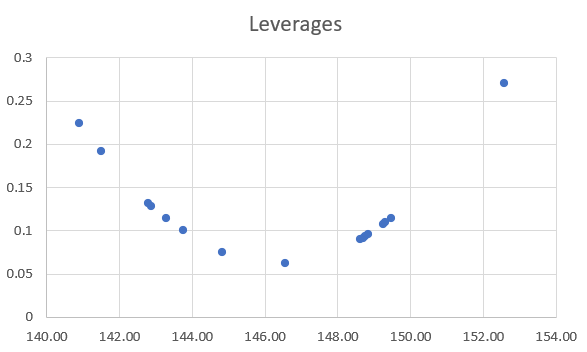
Using features in Excel, we created this plot of residuals that shows they generally are normally distributed. To mathematically confirm, however, we conducted the chi-square test for normality with an alpha of .05 and the following null and alternative hypotheses.

H0: The residuals are normally distributed

Ha: The residuals are *not* normally distributed

Our p-value is .24. Since it is larger than our alpha of .05, we failed to reject the null hypothesis and accept that the residuals are normally distributed. Our test statistic (2.85) was also less than our cut-off value (5.99) so we can again, accept that our residuals are normally distributed.

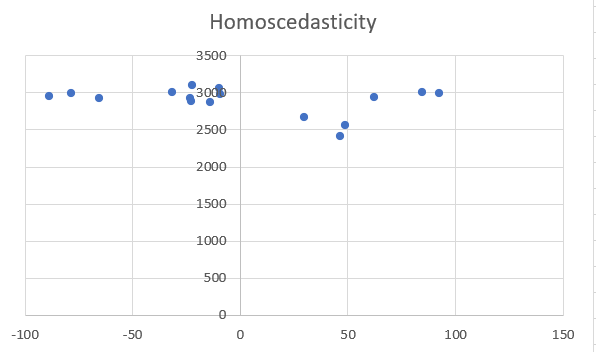
We then calculated the leverages of each X value to see how much influence each value has on our model. All of our leverages seem to influence the model equally, except for our X value of $152.57. That is more than twice the average of our leverages which indicates it is unusually close to our regression line compared to the other values and potentially influencing the model more than the other values. The plot of our residuals also clearly shows our outlier to the top right.



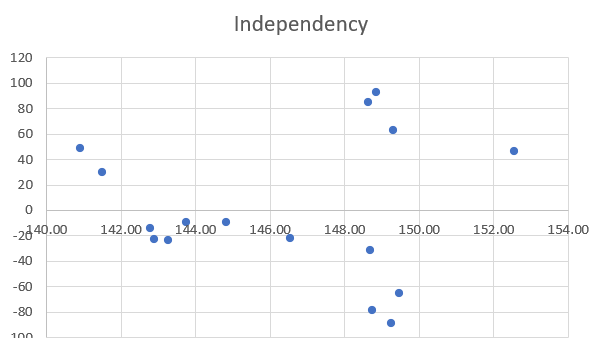
In order for our residuals to be accurate and fully accepted, they must meet the following criteria:

* Be normally distributed (normalcy)
* Have a constant variance (homoscedasticity)
* Be independent of X (independency)

We confirmed they are normally based on our previous analysis. Before checking the homoscedasticity and independency, we created a symmetrical variance-covariance matrix. We then plotted the diagonal entries of our matrix (variance) to the residuals in order to identify the homoscedasticity and ensure our residuals have a constant variance. Since we see a consistent pattern in the chart below, it is fair to assume the variances are consistent enough.



Our last check, independency, can be proven in two ways. First, the nondiagonal entries of our variance-covariance matrix show the covariances of all possible pairs of residuals. The values are really high on the diagonal because the covariance between the 12th residual and the 12th residual, for example, would be high because they are the same variable. The covariances between the other combinations need to be low to show they are not correlated with each other. No value in the matrix comes anywhere near the covariances on the diagonals so it is very likely they are all independent of each other. To confirm, we can plot the residuals against the X values. Since no pattern occurred in the plot below, we can confirm independency.



**Autocorrelation**

The last metric we calculated was our Durbin Watson statistic for autocorrelation. Since our value of .61 is less than 2, we have positive autocorrelation. In this case, if the stock price of TSLA went up yesterday, it is more likely to go up today. If we had negative autocorrelation, then a stock increase yesterday would yield a high likelihood of a stock decrease today. Naturally, we would expect some autocorrelation in stock prices changes. If a company is doing well, then their stock price should increase over many days. The company’s stock price should not be negatively autocorrelated and thus they would likely see a price decrease if they meet or exceed investor expectations. However, company performance changes and stock prices fluctuate day-to-day. Because of this, we should not have significant autocorrelation because it would skew our model towards recent trends and reduce the effects of company performance. Since we created a linear model, we expected some positive autocorrelation. In order to reduce it, we could try creating a non-linear model that can ebb and flow over time. If we wanted to keep a linear model, we could consider adding other variables to the model in order to mitigate the effects of APPL single-handedly affecting the price of TSLA in our model.

Summary

Based on our analysis above we can suggest portfolio recommendations with different levels of risk. However, since these stocks are highly positively correlated with each other, it is very risky to invest in only these two stocks since they seem subject to the same factors. However, we can only give recommendations for these two stocks since we only analyzed these two stocks. People with fewer financial commitments can be riskier than people with many financial commitments (people with kids and mortgages are more cash dependent and thus should minimize their investment risk.) For myself, I hardly have any financial commitments so I can increase my risk in order to increase my return. I would personally choose to keep 100% of my portfolio in TSLA stock and 0% in AAPL stock. Based on our linear long term forecast, TSLA is trending higher at a faster rate than AAPL. However, our TSLA model does have a higher percentage error so it is assumed to be riskier. Since I am young and am not living paycheck to paycheck, I have enough cash reserves and have a stable job to cover any significant losses if they happen. Because of that, I am seeking the maximum possible return on my investment. Because of our previous analysis, we know that these stocks are highly correlated with each other so I see very little value in holding AAPL. If AAPL goes up, TSLA is likely to increase more. However, my financial goals are not everyone’s goals. For people who are more risk averse or want to preserve more of their capital, I would recommend a 50/50 split between AAPL and TSLA.