Stock Market Risk Analyzer

Jason Van Bladel
School of Engineering and Computer Science
University of the Pacific
Stockton, California
j_vanbladel@u.pacific.edu

I. Introduction

The stock market is an extremely large complex system that is influenced by many factors and known for its irrational behavior. This complex system consists of thousands of marketable securities ranging from large blue-chip companies to new high-growth startup companies to complicated mutual funds. That being said, my research consisted of finding a way to evaluate the given market price of a marketable security at any point in time and quantify the amount of risk a security has based on statistics within the company. This is important for investors because with any marketable security the investor does not want to overpay for a security and wants to know the amount of risk of a given security. The beauty of the stock market it is complex system that acts on irrational behavior giving investors the ability to take advantage of mispriced marketable securities at a given time. A great way of identifying mispriced securities is to look at the life and blood of every company: balance sheets, income statements, and cash flow. These items do not entirely dictate the price of a given security; however, they give a realistic picture of where the company is heading at any given point in time. My research consisted of taking the stock market data with the daily stock market data for the last 10 years, company balance sheets, income statements, and cash flow statements to create a predictive model to project the risk of a given marketable security using data science techniques on Python.

II. Related Work

There have been many attempts to create a predictive model for stock market and many of these models can predict the value of a marketable security consistently over time due to the irrational behavior within the stock market. Recent attempts to create predictive models of given marketable security have recently utilized neural networks [1][2]. These attempts try to utilize machine learning to create a predictive output, this is an acceptable way to create a predictive model; however, with my project focusing on risk, machine learning is not appropriate. My research will focus on creating regression functions [3] and using this to do a statistical risk analysis [4].

III. Solution and Scalability

On a high level, the solution to my research that I propose is creating 5 regression lines from different statistics that correlate closely to price of the stock to predict the future price of marketable securities and then normalize these regressions to create multiple values in which a statistical analysis can be performed. From this solution, I can calculate a normal distribution of a stock price at a given time and calculation of probability of loss in capital. This is important because my research focuses on accurately analyzing the risk of a marketable security and this solution allows for a large amount of scalability. The solution has the ability to be applied to every security, can create a comprehensive predictive model, has the ability to show stock price at given point in time, and allows for the calculation of probability of loss in capital. This project in the future can be applied to mutual funds or stock market portfolios making the program very scalable to do the dynamic code.

IV. Implementation Details

The research consisted of 3 specific research phases: data gathering and cleaning, generating predictive regression functions, and statistical risk analysis.

A. Data Gathering and Cleaning

For the data gathering and data cleaning phase of the project I used 2 data sources; the first data source came from the Python module yfinance which gives various statistics of a given stock and the last 10 years of price data and the second data source came from an API called financialmodelingprep.com which queries securities balance sheets, income statements, and cash flow statements. Using these 2 data sources, I implemented a Python class called Stock_Storage which can query, sort, and store all the data to be quickly and easily accessed. This process consisted of clean the data to remove all Null values and creating dictionaries to store given values for easy access. The Stock_Storage class is outfitted with many helper functions and can be imported to other stock related projects very easily.

B. Generating Predictive Regression Functions

To generate predictive regression functions for the data, I utilized various Python libraries including numpy, sklearn, statsmodels.api, and scipy. Using these Python libraries in my project I was able to implement linear, exponential, quadratic, logarithmic, and polynomial regressions for the price of any marketable security in the stock market. Using different regression models for different securities is essential because it shows the different outcomes for the security. From the analysis, the polynomial regression showed the highest correlation (r squared value) to the various securities and this makes sense because a polynomial regression has the highest number of coefficients that can dictate the accuracy of the regression line. After generating the different regression lines to the price of the security, I then implemented the regression functions on the securities revenue, debt, eps, and net income because these 4 values are highly correlated to the price of a marketable security.

C. *Statistical Analysis*

The final phase of the research would be to complete a statistical analysis on the regression functions that were created. For simplicity, only the linear regression models were used within this section of the project. The first step of the statistical analysis was to normalize the linear regression lines for price, revenue, debt, eps, and net income. To do this, each regression lines calculated values at time equal to today 12/7/2020 and 1 year from now 12/7/2021. These values then were normalized to become percent change. The one unique value of debt was inversed because an increase in debt would result in a reduction in value of the security. After all the value of price, revenue, debt, eps, and net income were normalized to become percent change, the average and standard deviation were calculated. Given the average and standard deviation of the percent change in a marketable security, we can convert these values to represent price. This can be used to create a normal distribution and using Z-Score the probability of a security losing capital can be calculated. Finally, from all of these values, the risk of security can be calculated as:

Risk = Probability of Loss x Market Price of Security

This formula allows for the calculation of capital a risk for a given security.

V. RESULTS

The results of implementing the research is that I can create regression functions of price, revenue, debt, earnings per share (EPS), and net income. The price regressions can be seen in figures 1-4, revenue can be seen in figures 5-6, and debt in figures 7-8.

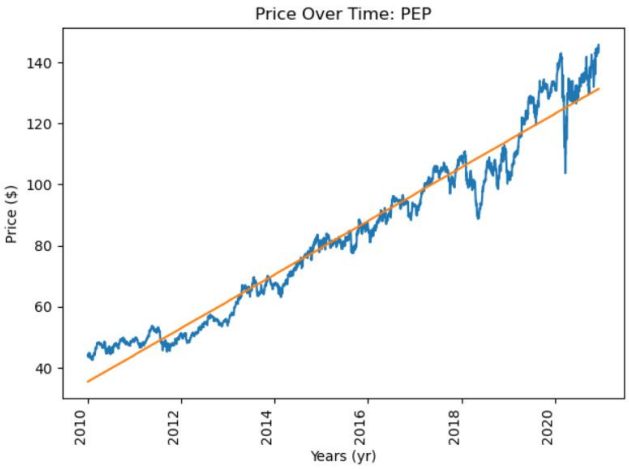

Fig. 1 Linear Regression for price for PEP

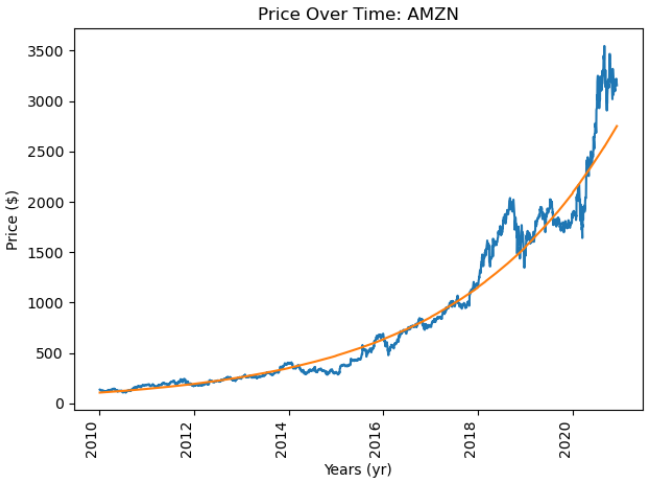

Fig. 2 Exponential Regression for price for AMZN

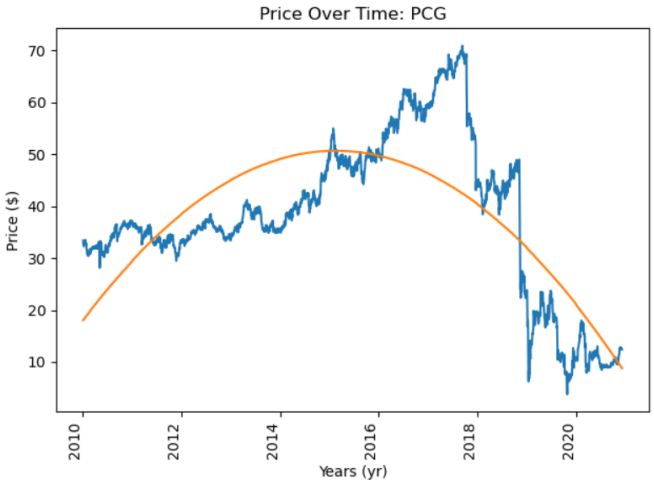

Fig. 3 Quadratic Regression for price of PCG

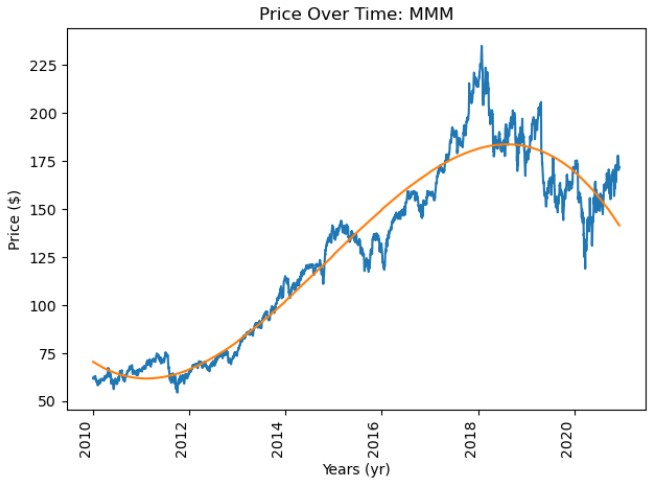

Fig. 4 Polynomial Regression for price of MMM

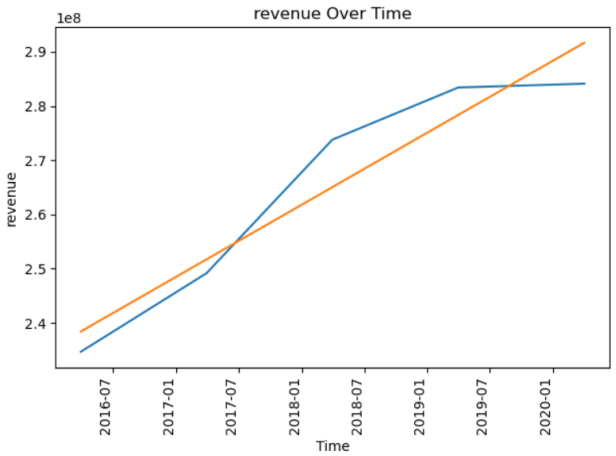

Fig. 5 Linear Regression for Revenue for PETS

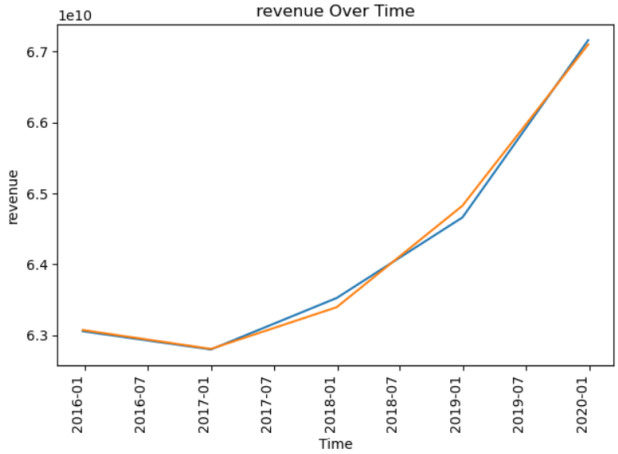

Fig. 6 Quadratic Regression for Revenue for PEP

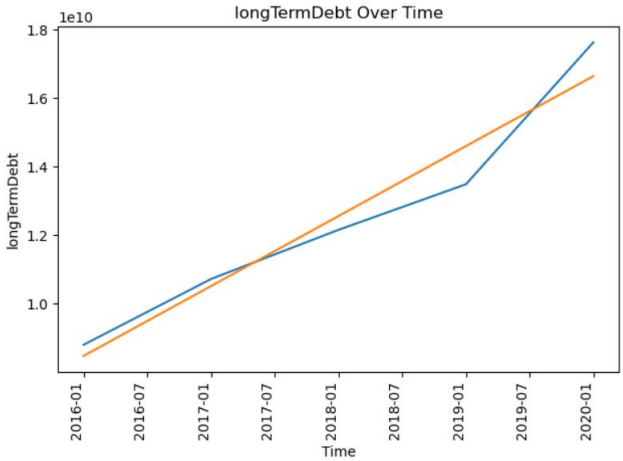

Fig. 7 Linear Regression for Debt for MMM

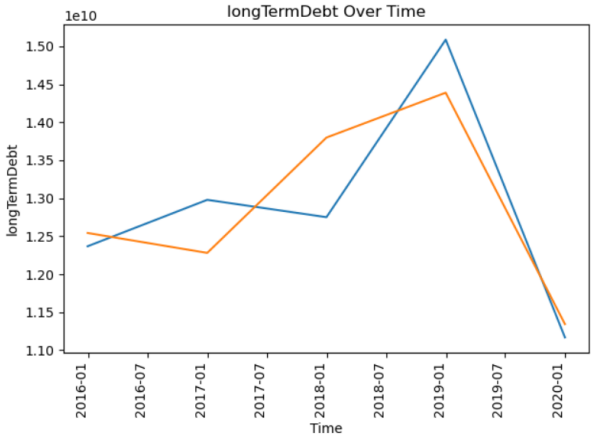

Fig. 7 Polynomial Regression for Debt for MGM

These regression lines were then normalized over the time interval (12-7-2020, 12-7-2021) calculating the avg percent change over time and standard deviation change over time. As seen in Figure 8, we can see these values graphed over the normal distribution.

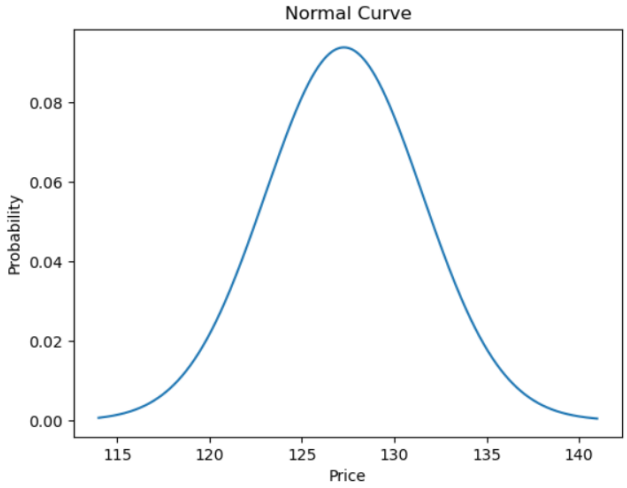

Fig. 8 Normal Distribution for Projected Price 2021-12-7of AAPL

Using the Z-Score we can then calculate the probability in loss in capital and the associated risk for a given security. The calculated values are shown in Table I for the security AAPL.

Table I. Stock AAPL Statistics

Price: 2020-12-7	$122.25
Avg Projected Price 2021-12-7	$127.28
Std Projected Price 2021-12-7	$4.25
Probability of loss	12%
Probability of gain	88%
Associated Risk	$14.51

From these calculated values, we can quantify the amount of risk given within a various stock and I encourage you to download the code to see the how these values change for various securities that are traded on the stock market.

Keep in mind that these calculated values are based on linear regressions.

VI. Conclusion

In conclusion, this project successfully created a predictive model to project the risk of a given marketable security. My research consisted of taking the stock market data with the daily stock market data for the last 10 years, company balance sheets, income statements, and cash flow statements to create a predictive model to project the risk of a given marketable security using data science techniques in Python. This is significant because we can quantify the amount of risk of a given marketable security based on price data, balance sheets, income statements, and cash flow statements giving investors a way to analyze the amount of risk in a given marketable security on the stock exchange.

References

[1] T. Kimoto, K. Asakawa, M. Yoda and M. Takeoka, "Stock market prediction system with modular neural networks," 1990 IJCNN International Joint Conference on Neural Networks, San Diego, CA, USA, 1990, pp. 1-6 vol.1, doi: 10.1109/IJCNN.1990.137535.

[2] Qian, B., Rasheed, K. Stock market prediction with multiple classifiers. Appl Intell 26, 25–33 (2007). https://doi.org/10.1007/s10489-006-0001-7.

[3] Robert F. Whitelaw, Stock Market Risk and Return: An Equilibrium Approach, The Review of Financial Studies, Volume 13, Issue 3, July 2000, Pages 521–547, https://doi.org/10.1093/rfs/13.3.521K. Elissa, “Title of paper if known,” unpublished.

[4] Pindyck, Robert S. "Risk, Inflation, and the Stock Market." The American Economic Review, Vol. 74, No. 3, (June 1984), pp. 335-351.