

Measuring Stock Market Risk with Correlation

Liuboyan Bryan

Jiutian education, Beijing, China
liub2109@gmail.com

ABSTRACT

In this report, we investigate the systemic risk via capturing the correlation between the individual stocks and the market. In particular, we consider eight stocks (Microsoft, Exxon Mobil, Caterpillar, Johnson & Johnson, McDonald's, Sandisk, Qualcomm, and Procter & Gamble) and we employ the S&P 500 as the market return. We first evaluate the risk of each individual stock using different risk metrics, e.g., variance (StD), minimax range, and Sharpe Ratio. Furthermore, we employ the regression approach (or beta method) to evaluate the correlation between each individual stock return and the market return, and also compute the R² that measures the explainability level of the market information to each individual stock return. We conclude that Caterpillar has the largest sharpe value, and McDonald's has the largest R square value.

KEYWORDS

Stock Market; Return; Risk measuring

1. INTRODUCTION

Risk measuring is nowadays of great significance for the vast amount of stock corporations, achieving informed investment decisions, controlling the variables to reduce the loss of companies, and maintaining market stability. The purpose of this report is to analyse and measure the risks in the stock market. There is a collection of methods to gauge the risk and volatility of an individual stock, such as by using variance, standard deviation, and value-at-risk etc. However, these means fall short in capturing the systemic risk in terms of the possible correlation with the market. For instance, standard deviation is the one of the way showing the volatility of the stock, but it does not consider to what extent does one specific stock fluctuates along with a standard market index (S&P 500)—a stock market index that tracks the stock performance of the 500 largest companies listed on U.S. stock exchanges. Traditional risk measures such as standard deviation assume that returns are symmetrically distributed (normal). However, real-world returns often exhibit skewness. Skewness provides additional information about the nature of risk, highlighting whether the risk is skewed toward extreme losses or gains. Therefore, we use the beta-method to measure the risk of stock, and betas for any individual stock are determined by simple linear regression.

The remainder of the report contains 3 sections: Section 2: Linear Regression, Section 3: Measuring Stock-Market Risk using Linear Regression, and Section 4: Conclusion. In Section 2, the method of linear regression is introduced, which illustrates the definition of the linear regression model and the use of it in different areas, including the variables in the regression equation. In the next section, it demonstrates the main results and analyses the results with some insightful implications. Using various statistics methods and terminology to investigate the risk from each individual stock. Variance, standard deviation, mean value, expected value, and r square etc. will all be included in this part. Conclusion will be the last part in this report.

2. LINEAR REGRESSION

Linear regression model is a statistical method used to model the relationship between dependent and independent variables and or more than one variable. It aims to predict the dependent variables by using the independent variables. It is used to investigate the trend for different data and also in order to measure the risk in particular areas like stock markets. Statisticians and analysts calculate the gradient and y-intercept for data to construct and plot a regression line and equation, investigating the relations.

Linear regression is widely used in various fields such as economics, finance, biology, engineering, and social sciences for tasks such as predicting sales, estimating the impact of advertising, and assessing disease risk factors.

The simple regression model is " It explains the relations between y and x. It is the parameter of the model or is called intercept and it is important, as this value may help to plot the linear function. is the gradient of the model, representing to what extent does x reflect on y and it can be clearly seen on the graph by witnessing the skewness of the line. Y is the explained variables or dependent variable and X is the explanatory variable or independent variable in the model. is also a crucial term in the equation, representing error term, a random variable that explains the factor of y that cannot be interpreted by the relationship between x and y.

And it must obey the 4 principles:

- (1) $E[\epsilon]=0$;
- (2) The $Var[\epsilon] = \sigma^2$ for different x value is same;
- (3) The values of ϵ is independent;
- (4) The error term ϵ is a normally distributed random variable. In this investigation the simple regression model for risk measuring is defined as $\text{marke } t = \beta_0 + \beta_1 \text{ Total return for the stock} + \epsilon$.

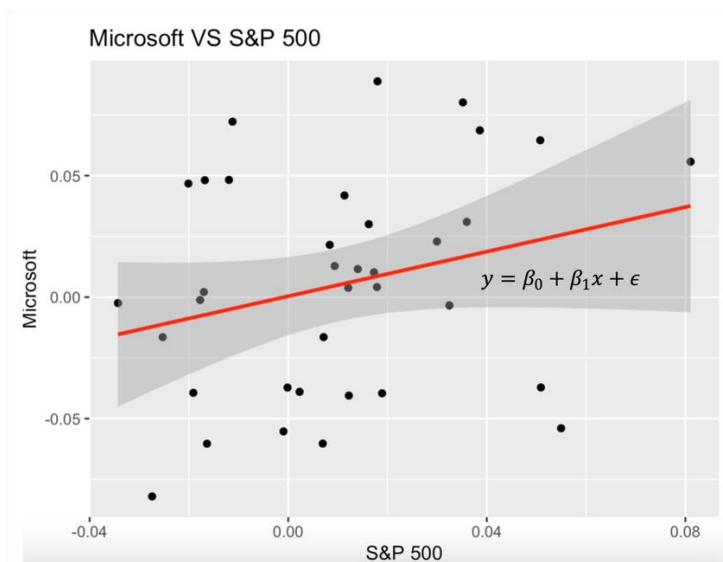


Figure 1. The regression model

Regression models assume a linear relationship between the independent and dependent variables, allowing a direct explanation of how changes in the independent variables affect the dependent variable. It helps to predict the linear relations between variables. By plotting the regression line, R-squared, Adjusted R-squared, and residuals are significant. R^2 is defined as ESS/TSS and the sum of squares (ESS) is the portion of the total variation explained by the regression model and total Sum of Squares (TSS) is the total variation explained in the dependent variable.

3. MEASURING STOCK-MARKET RISK USING LINEAR REGRESSION

3.1. Descriptive Statistics of Individual Stock

In this subsection, we present the descriptive statistics of each individual stock, including mean value, standard deviation, maximum, minimum, and sharpe ratio for each individual stock. In particular, Figure 1 plots the histogram of total return for each individual stock, and the statistics of mean, standard deviation, max-min range, and Sharpe ratio for measuring their risk are presented in Table 1. Here the Sharpe Ratio: is calculated as $(\text{Mean}-R_0)/\text{StD}$ where we let $R_0=0$.

The histograms in Table 1 illustrate the total return for each individual stock. According to the histograms below, it is generally illustrated the spread of different individual stocks. Sandisk is the most volatile stock among all, since it is the most dispersed data on histogram, with the the range between -0.2 and 0.6, holding the largest standard deviation (0.1953974) — measuring the spread between data and data. And therefore, it is the most volatile stock. However, McDonald’s is the least volatile stock, which means the spread of each data is little, with standard deviation (0.006809637).

The histograms in Table 1 illustrate the total return for each individual stocks. According to the histograms below, it is generally illustrated the spread of different individual stocks. Sandisk is the most volatile stock among all, since it is the most dispersed data on histogram, with the the range between -0.2 and 0.6, holding the largest standard deviation (0.1953974) — measuring the spread between data and data. And therefore, it is the most volatile stock. However, McDonald’s is the least volatile stock, which means the spread of each data is little, with standard deviation (0.006809637).

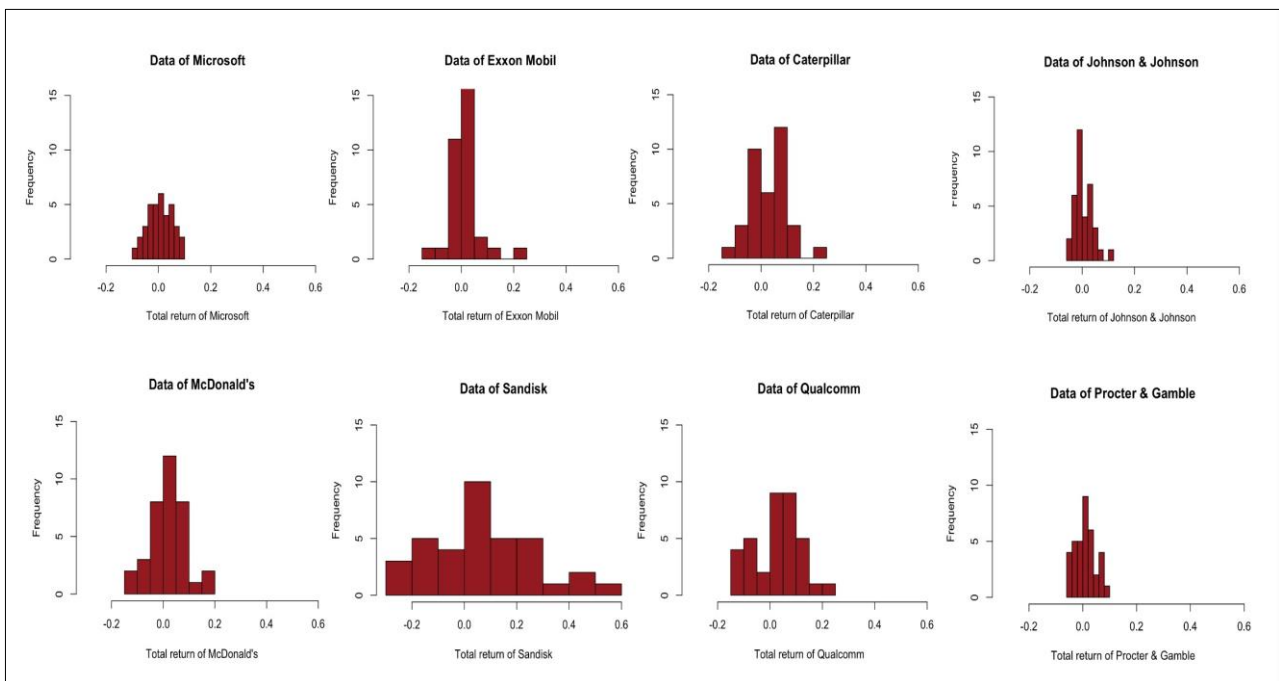


Figure 2. sample distributions of returns of different stocks

The statistics in Figure 2, by comparing the mean value, can reflect that Sandisk is the best stock, as the average for its return is technically the largest and which means it has the most returns of profits. Microsoft; however, lines the last, with the value of and has the least returns of profit on average. Others are similar and not exceed 0.06. Maximum figure tells us the best performance of a stock in their tendency, which illustrates the peak value for a stock to help investors to predict the trend of a stock, and while minimum value gives a limit for a stock to help investors to compare the minimum Return. The difference between maximum and minimum value is also crucial, giving the range of difference between individual stocks. By examining the range, it can give us a general spread of data and can help to locate the value by comparing with the upper quartile, lower quartile, and median. If ,

it shows a positive skew and vice versa. Skewness can influence investment decisions and portfolio management strategies. Investors may prefer positively skewed assets because they have the potential to deliver occasional huge gains, while avoiding negatively skewed assets that can result in unexpectedly large losses. This is particularly significant for investors, who want to minimise the possibility of extreme negative returns.

The Sharpe ratio is an essential tool to help measure the return of an investment compared to its risk. If two portfolios have the same return, the one with the higher Sharpe ratio is considered better because it achieves the same return with less risk. Sharpe Ratio: $(\text{Mean}-R_0)/\text{StD}$ where we let $R_0=0$. It shows how much excess return you will get from the extra volatility of holding a riskier asset. The higher the sharpe value a higher return per unit of risk for an investment, which is a good thing. Caterpillar has the largest sharpe value (0.438970148), which means that with one unit of volatility or risk increases, caterpillar can receive the largest amount of returns among all stocks., but it does not mean that it has the largest volatility. On the other hand, Procter & Gamble has the smallest sharpe ratio (0.02861256), the worst risk-adjusted return and investors need to avoid this stock.

Table 1. Descriptive statistics of different stocks

Stocks	Mean	Standard deviation	Maximum	Minimum	Sharpe ratio
Microsoft	0.005025556	0.04537158	0.08883	-0.08201	0.110764404
Exxon Mobil	0.01663722	0.05534009	0.23217	-0.11656	0.20063594
Caterpillar	0.03009722	0.06856325	0.21847	-0.1006	0.438970148
Johnson& Johnson	0.005295556	0.034866	0.10334	-0.05917	0.151883095
McDonald’s	0.02447417	0.006809637	0.18257	-0.11443	0.359404914
Sandisk	0.06926194	0.1953974	0.50165	-0.28331	0.35446705
Qualcomm	0.02836028	0.08618695	0.21055	-0.1217	0.329055385
Procter & Gamble	0.01058856	0.03706685	0.087833	-0.05365	0.02861256

3.2. Evaluating the Correlation between Stocks and Market

While in Section 3.1 we have discussed the descriptive risk statistics for each individual stock, the potential correlation between each stock and the market have not yet been evaluated, a topic we delve into in this subsection. The regression lines are plotted in Figure 3 and the associated beta values are presented in Table 3.

In particular, One intuitive approach of exploiting the variation in different stocks and markets is to compare the slope of variables of different companies. The beta value is the gradient in a simple linear regression equation and is a measure of a stock's sensitivity to market fluctuations. A beta value greater than 1 means the stock is more volatile than the market, while a beta value less than 1 means the stock is less volatile.

It is clearly demonstrated in table 3 that Sandisk is the most volatile stock among all, with standard deviation (0.1953974). Standard deviation is a way to measure the spread of the data and in the stock market it helps to measure the volatility. According to beta value, sandisk will perform best in an up market, as the beta value is positive, representing positive slope (a positive relationship) and it is the largest beta value among all which is greater than 1. When beta value is greater than 1, stock is more volatile than the market. In addition, the mean value may also help to indicate the good or bad of a stock. Sandisk has the mean value (0.06926194), which is the largest showing when one unit of change may bring 6% profit. Therefore, Sandisk is comparatively the best stock among all.

Qualcomm holds their value best in a down market, as it maintains a negative proportional relationship with the market, with the beta value (-0.12611). When the market is witnessing a recession or downward tendency, it performs best and is, however, experiencing an ascending trend and receiving more profit. Johnson & Johnson stock is very interesting and the beta value for it is

0.008748, showing nearly a horizontal line on slope, which indicates the trend for it is a proximity to the market. When the market increases, it ascends, vice versa.

Table 2. The beta value and intercept of different stocks

Stocks	β value	γ intercept
Microsoft	0.458374	0.000398
Exxon Mobil	0.730914*	0.009258
Caterpillar	1.49321***	0.01502
Johnson & Johnson	0.008748	0.005207
McDonald's	1.503223***	0.009298
Sandisk	2.60493*	0.04296
Qualcomm	-0.12611	0.07284*
Procter & Gamble	0.506541*	0.005475

Figure 3 demonstrates the regression line of different stocks, showing the relationship and trend of each stock and the market. The gradient is beta value and helps to identify how strong the stock is associated with the market. It also shows some outliers in different graphs. Investors may use the regression line and the spread of the data on the graph to decide which stock is the most profitable and which stock is the most healthy one. Residuals provide insight into how well a model explains the stock returns. Examining the residuals on a graph can help a statistics analyst to measure the error, and if the residuals are large, it indicates that a significant portion of the stock's return is not explained by the market movements, vice versa and this unexplained part is considered unsystematic risk, which is specific to the individual stock.

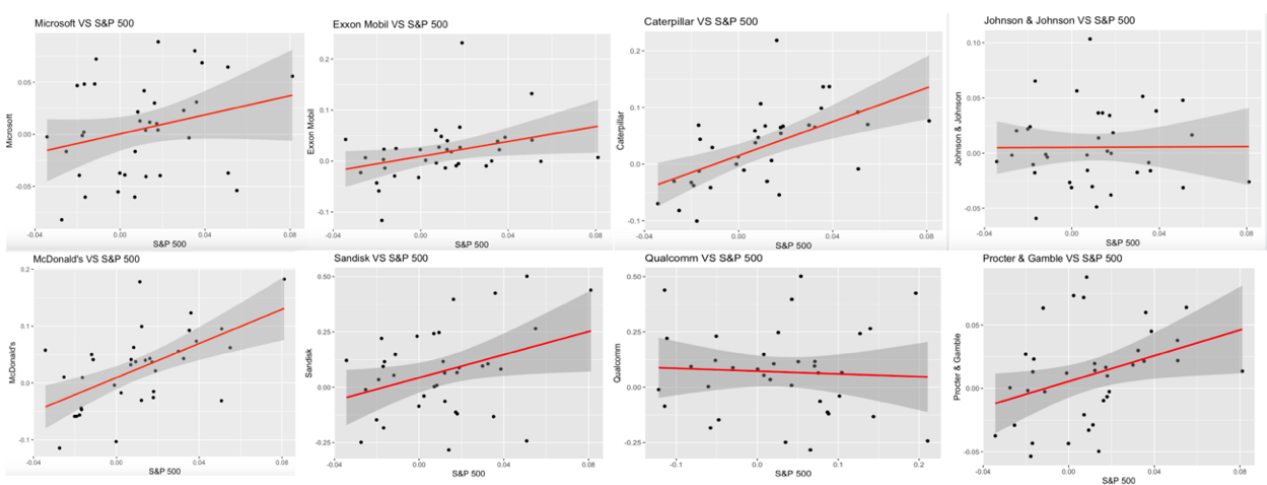


Figure 3. Regression lines for the correlation between each individual stock and the market

3.3. Explaining the Stock-Return via the Market-Return Information

Furthermore, we try to explain the stock returns with the market information using the regression results. In particular, we calculate the total sum of squares (SST) and the explained sum of squares (SSE) for each stock return with respect to the market return.

Table 3. SST, SSE, and R² of different stocks

Stocks	SST	SSE	R ² =SSE/SST
Microsoft	0.07205032	0.005097827	0.070753704
Exxon Mobil	0.1071884	0.01296217	0.120928851
Caterpillar	0.1645322	0.05409904	0.328805182
Johnson& Johnson	0.04254733	1.85659E-06	4.36E-05
McDonald's	0.1622991	0.05482669	0.337812656
Sandisk	1.336306	0.1646408	0.123205912
Qualcomm	0.25998660	0.04850091	0.186551576
Procter & Gamble	0.0480883	0.0006225499	0.129459744

The amount of the individual stock that can be explained by the SP 500 is measured by the R², which is defined by

$$R^2 = \frac{SSE}{SST} \in [0,1],$$

Where

$$SSE = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad SST = \sum_{i=1}^n (y_i - \bar{y})^2.$$

From the results (Table 4), the conclusion can be drawn that McDonald's has the largest R square value, which means that the regression is most reliable among all and the regression line is the best fitted to explain the relationship between individual stocks and S&P 500. However, Johnson& Johnson has the least R square value that is nearly 0, meaning a poor relationship can be explained by the market return. Others are all around in the range between 0 and 1. R²=0: The independent variables do not explain any of the variability in the dependent variable. R²=1: The independent variables explain all the variability in the dependent variable. And the larger the R² value represents a better fit of the model. When R² equals 0.07, it reflects that 7% of the variance in the dependent variable is explained by the independent variable and the remaining 93% of the variance is due to factors not included in the model or random variability.

In addition, it is also easy to see

$$R^2 = \frac{(\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}))^2}{((\sum_{i=1}^n (x_i - \bar{x})^2) \times (\sum_{i=1}^n (y_i - \bar{y})^2))},$$

$$R = \frac{(\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}))}{\sqrt{(\sum_{i=1}^n (x_i - \bar{x})^2) \times (\sum_{i=1}^n (y_i - \bar{y})^2)}, \quad \text{and}$$

$$R = \frac{Cov(X,Y)}{\sigma_X \times \sigma_Y} = \sqrt{\frac{SSE}{SST}}.$$

From the above we can derive that

$$R = \rho_{XY}$$

that is exactly the linear correlation coefficient between the individual stock and the market (SP&500). In other words, if the R² of some individual stock with respect to the market is close to 1, then they are highly correlated, vice versa.

4. CONCLUSION

In conclusion, I have investigated the relations between each individual stock and the market and the risk of each stock and the market, by using different statistics metrics (regression, sharpe ratio, R^2 etc.). By evaluation above, we can precisely get the result that Sandisk has the largest standard deviation and is the most fluctuated stock compared to the market. Sandisk also has the largest range with (0.78496), and larger range indicates greater volatility, which implies higher variation. Therefore sandisk is less predictable and more likely to experience large changes. Caterpillar has the largest sharpe value, whereas, Procter & Gamble has the smallest sharpe ratio (0.02861256), the worst risk-adjusted return and investors need to avoid this stock. In addition, McDonald's is the least volatile stock, with the smallest standard deviation. Lastly, Procter & Gamble has the least range, which means it has a smaller risk among all stocks.

The regression lines plotted and calculated gradient of each stocks clearly show that Sandisk performs best across all stocks, holding the largest beta value (2.60493), which is best fitted with the market; however, it is interesting to see that Qualcomm has the downward sloping (-0.12611), witnessing an reciprocal relation to the market. Others do not exceed 1 and in the range 0 to 1, with a generally flat gradient. In a nutshell, a lower beta means lower risk and potentially lower returns, while a higher beta means higher risk and potentially higher returns. Higher risk can lead to higher returns, it is essential for investors to align their risk tolerance with their investment strategy to mitigate potential adverse impacts.

R^2 is the last part of the results, McDonald's has the largest R^2 , and the regression line is the best fitted to explain the relationship between individual stocks and S&P 500. However, Johnson & Johnson has the least R square value that is nearly 0, meaning a poor relationship can be explained by the market return. Others are all around in the range between 0 and 1.

Granted, there are also some limitations in this regression model, as it is a linear regression and it only contains two variables; however, in reality there are more than 2 factors that affect the stocks and it is a multiple situation. For example, changes in government policies, regulations, or geopolitical events can significantly impact stock prices. In addition, companies with unstable financials, high debt, or inconsistent earnings are riskier investments. And stocks in certain industries (e.g., technology, biotechnology) tend to be more volatile due to rapid innovation and competition. It can also be highly sensitive to the economic cycle of the country. Therefore, there are a majority of influences, instead of the market itself. In the future, it is more reliable to use multiple regression model to evaluate the risk, even though we might not consider all the factors, it still can harness more information than linear regression does.

REFERENCES

- [1] Eckel S, L? Ffler G, Maurer A, et al. Measuring the effects of geographical distance on stock market correlation [J]. Journal of Empirical Finance, 2011, 18(2):237-247. DOI: 10.1016/j.jempfin.2010.12.001.
- [2] Wang X, Cai G. A Combined BP Neural Networks with Correlation Analysis and PCA for Stock Market Prediction [J]. 2009.
- [3] Blackburn D W, Chidambaran N K. Measuring market integration: US stock and REIT markets[C]//2018. DOI:10.1007/978-3-319-99624-0_12.
- [4] Gardiner D F. A re-examination of stock-market risk [D]. 2011.