

# Q.E.D. Stock Analysis for IEOR 4150

Yili Yang, Wanling Wu, Qilin Peng

## Introduction

As the financial analysts of the Q.E.D. investment Inc., we are professional in performing statistical analysis to assist our client's decision making.

American stocks have reached to fresh record, according to the Wall Street Journal (Gold & Driebusch, 2016). Citing this equity appreciation background, our major client, IEOR4150 Corp., has approached us with ten candidate S&P 500 stocks for a comprehensive review on their historical log-returns throughout the past five years. We have developed an interactive decision-supporting App using R Shiny and HTML ([https://yili-yang.shinyapps.io/shiny\\_app\\_win/](https://yili-yang.shinyapps.io/shiny_app_win/)) to address the following questions:

- 1) How is the log-return of each candidate stock distributed? (Histograms)
- 2) Does the log-returns follow normal distribution? (Quantile-Quantile Plot)
- 3) Is the log-returns follows a random walk? (Test of Randomness)
- 4) What is the time-series movement of each stock? (Regression on Time)
- 5) What are the confidence intervals for the means and variance given 95% confidence level? (Confidence Intervals)
- 6) What is the historical beta of the past five years of each stock? (CAPM)
- 7) Do two stocks have equal log-return means? (T-Test)
- 8) Whether two stocks have a linear correlation? (Regression)
- 9) Are these ten stocks correlated? (Correlation Coefficient Matrix)

The client can get familiarize with our extraordinary user-friendly application using this manual.

## Data Set

Ten stocks constitute our data set, covering the sectors: technology, finance, Auto manufacturing, pharmaceuticals and retail. The daily closing prices of each stock were extracted from Yahoo Finance and their daily log-returns were calculated from 10/10/2011 to 10/7/ 2016 shown as following:

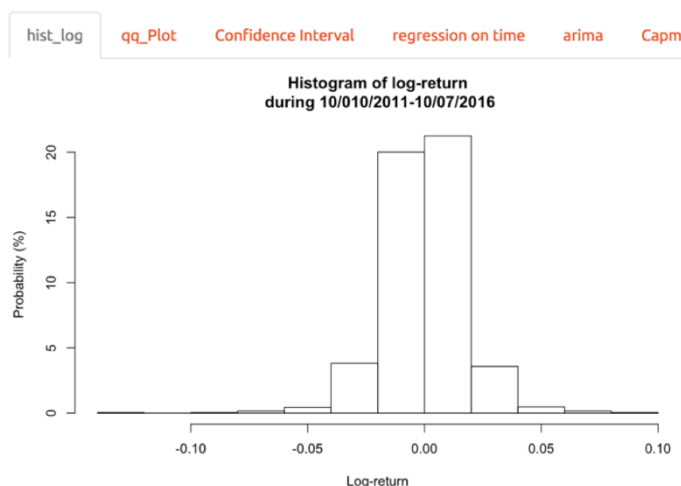
- |   |  |
|---|--|
| 1. <i>Wal-Mart Stores Inc. (WMT)</i>    | 2. <i>Alphabet.Inc. (GOOG)</i>         |
| 3. <i>Apple Inc. (AAPL)</i>             | 4. <i>Tesla.Motors,.Inc.. (TSLA)</i>   |
| 5. <i>Toyota Motor Corporation (TM)</i> | 6. <i>Johnson.&amp;.Johnson. (JNJ)</i> |
| 7. <i>GlaxoSmithKline plc (GSK)</i>     | 8. <i>The Goldman Sachs Group (GS)</i> |
| 9. <i>Morgan Stanley (MS)</i>           | 10. <i>Emerson Electric Co. (EMR)</i>  |

On top of these ten stocks, the 10-year US treasury bond daily prices and the S&P 500 daily index were also collected within the same time frame. They were applied to the CAPM analysis.

## Analysis

### Single Stock Analysis

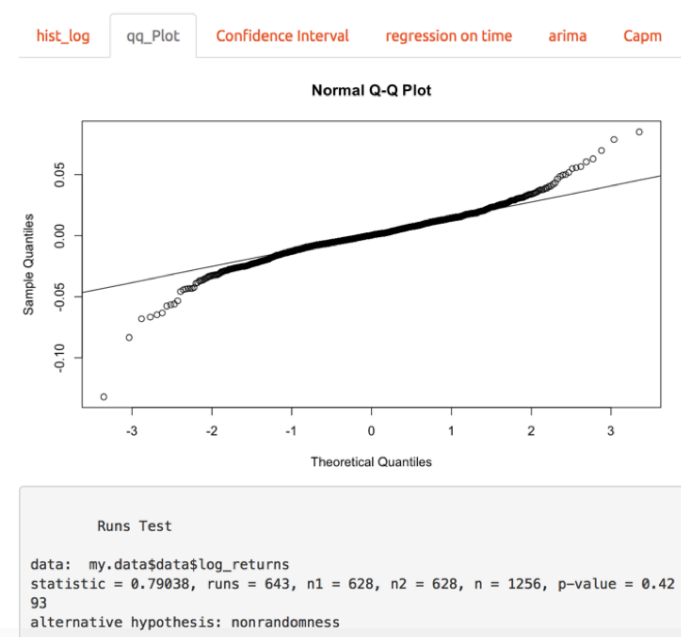
#### (1) Histograms



Take Apple Inc. as an example, the histogram of all the log-returns are approximately bell shape, which gives us an intuition that the log-returns could be approximately normal. The share split issue needs to be address because it created outliers of the log-returns. Two share splits are found in our data set: Apple's share price dropped from 645.57 to 93.70 on 6/9/2014 and the price of Alphabet, also

known as Google, dropped from 1131.97 to 558.46 on 3/27/2014. We fixed this problem by assigning the daily log-return on that date to be 0.

#### (2) Q-Q Plot and Run Test



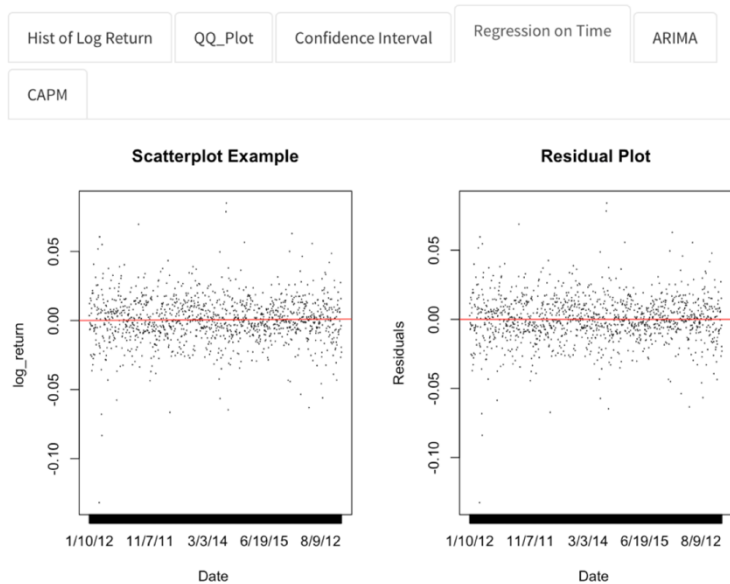
The Normal Q-Q Plots of all the ten stocks are similar as the one above. Though the log-returns have fat tails, we could still regard the log-returns follow approximately normal distribution.

**Table 1: Run Test Results**

Stock	AAPL	GOOG	GS	GSK	JNJ	MS	TSLA	TM	WMT	EMR
<b>P-Value</b>	0.4293	0.5346	0.1421	0.6114	0.0625	0.1016	0.5724	0.0444*	0.0239*	0.8655

To determine whether the log-returns are random or not, we have conducted the run tests on each stock with 5% level of significance. The null hypothesis is that the log-returns are random while the alternative hypothesis assumes non-randomness. From the above run test results summary, we found that only Toyota and Walmart have a p-value lower than 5%, yielding a rejection of the null hypothesis. Hence we know that most of the ten stocks' log-returns are random.

### (3) Regression on Time



The plots of regression on time showed no linear relationship of the regression between pairs of stocks.

**Table 2: Regression on Time Summary**

Stock	Slope	Intercept	Multiple $R^2$	F Test P-Value
<b>AAPL</b>	7.829e-07	6.733e-05	0.0002883	0.5476
<b>GOOG</b>	5.421e-07	5.128e-04	0.0001781	0.6365
<b>GS</b>	1.324e-06	-3.800e-04	0.0009002	0.2878
<b>MS</b>	1.176e-06	-1.448e-04	0.0004083	0.4742
<b>GSK</b>	7.829e-07	6.733e-05	0.0002883	0.5476
<b>JNJ</b>	1.383e-06	-4.457e-04	0.001342	0.1943

<b>TSLA</b>	1.383e-06	-4.977e-04	0.001363	0.1909
<b>TM</b>	1.383e-06	-4.457e-04	0.001342	0.1943
<b>WMT</b>	1.154e-06	-5.459e-04	0.00156	0.1616
<b>EMR</b>	9.480e-07	-4.775e-04	0.0006436	0.3688

The null hypothesis of F test assumes the slope of regression to be zero. With a 95% confidence level, we find that we do not have sufficient evidence to reject the null hypothesis. Therefore, all the regressions on time have a slope of 0.

#### (4) ARIMA

```

Forecast method: BATS(1, {0,0}, 0.951, -)

Model Information:
BATS(1, {0,0}, 0.951, -)

Call: tbats(y = y)

Parameters
  Alpha: 0.006624765
  Beta: -0.0009858228
  Damping Parameter: 0.951082

Seed States:
      [,1]
[1,] -0.006925957
[2,]  0.000467895

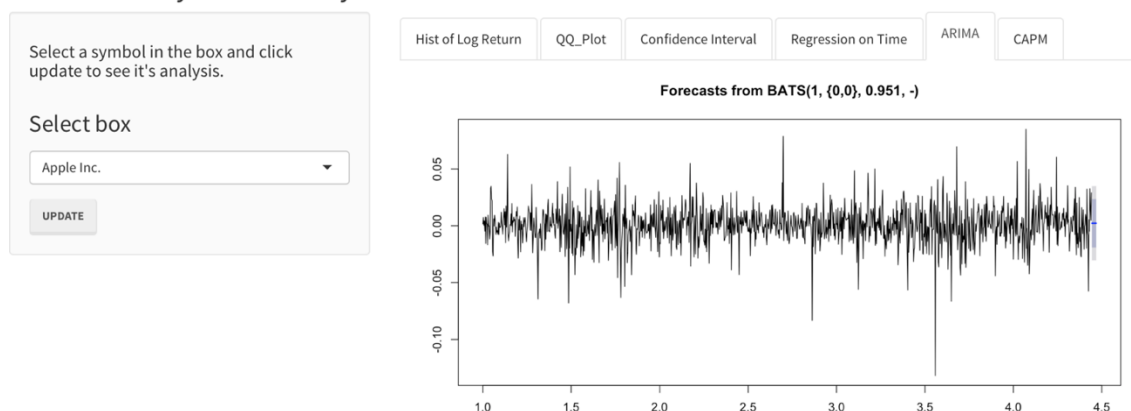
Sigma: 0.01665442
AIC: -1314.471

Error measures:
              ME      RMSE      MAE  MPE  MAPE      MASE      ACF1
Training set -8.015467e-06 0.01665442 0.01191988 -Inf  Inf  0.6731125 0.008060245

Forecasts:
      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
4.443836    0.002219259 -0.01912424  0.02356275 -0.03042280  0.03486132
4.446575    0.002223083 -0.01912076  0.02356692 -0.03041950  0.03486567
4.449315    0.002226719 -0.01911737  0.02357080 -0.03041624  0.03486968
4.452055    0.002230178 -0.01911407  0.02357443 -0.03041304  0.03487339
4.454795    0.002233468 -0.01911089  0.02357782 -0.03040991  0.03487684
4.457534    0.002236597 -0.01910782  0.02358101 -0.03040687  0.03488006
4.460274    0.002239572 -0.01910487  0.02358402 -0.03040394  0.03488308
4.463014    0.002242402 -0.01910205  0.02358686 -0.03040112  0.03488593
4.465753    0.002245094 -0.01909936  0.02358955 -0.03039843  0.03488862
4.468493    0.002247654 -0.01909680  0.02359211 -0.03039588  0.03489118

```

#### Statistics Analysis of One Symbol



To forecast the future of the stocks' performance, we introduced ARIMA and used 5% of significance. The regression result and coefficients can tell us the future of the log-return will be in the gray area with 95% confidence. The exact forecast result can be also seen in the table with 20% significance and 5% significance.

#### (5) Confidence Intervals of Mean and Variance

Since it was unknown that the variances of the normal distributions from which the stock log-returns approximately follow, the t-distribution with  $n - 1$  degree of freedom was applied to find the confidence interval for  $\mu$ . To find the confidence interval for  $\sigma^2$ , we utilized a chi-squared distribution with  $n - 1$  degrees of freedom. The following table is the illustration of the 95% confidence interval of mean and variance of each of the stock.

**Table 3: 95% Confidence Interval for  $\mu$  and  $\sigma^2$**

Stock	Confidence Interval for $\mu$	Confidence Interval for $\sigma^2$
AAPL	[0.00053019, 0.00058941]	[0.00027968, 0.00028109]
GOOG	[0.00024780, 0.00033558]	[0.00061438, 0.00061746]
GS	[0.00042433, 0.00048099]	[0.00025601, 0.00025729]
MS	[0.00055782, 0.00063260]	[0.00044585, 0.00044809]
GSK	[-3.41831778e-05, 5.19003332e-06 ]	[0.00012360, 0.00012422]
JNJ	[0.00047432, 0.00050509]	[7.548085704e-05 , 7.58596352e-05 ]
TSLA	[0.00149720,0.00161069]	[0.00102668, 0.00103182]
TM	[0.00039989, 0.00044837]	[0.00018739, 0.00018833]
EMR	[9.47973329e-05 , 0.00014279]	[0.00018364, 0.00018456]
WMT	[0.00016094, 0.00019845]	[0.00011217, 0.00011236]

As shown in the above table, the 95% confidence interval for  $\mu$  of the log-return of all of the stocks were close to zero with the upper and lower bounds not exceeding  $\pm 1$  %. Similarly, the 95% confidence interval for  $\sigma^2$  displayed an interval that was also extremely small, suggesting that the log-returns of all of the stocks possessed little variability given the time span of five years regardless of different industries.

#### (6) Capital Asset Pricing Model (CAPM) Regression

We would like to investigate the systematic (market) risk effect of each stock by capital asset pricing model (CAPM) regression:

$$R_{i,t} - r_f = \alpha_i + \beta_i (R_{m,t} - r_f) + \varepsilon_{i,t}$$

$R_i$  = return on asset  $i$  at time  $t$ .

where  $i=1,\dots,N$  and  $t=1,\dots,T$

$r_f$  = risk free return at time  $t$ .

$R_{m,t}$  = return on the market portfolio at time  $t$ .

$\alpha_i$  and  $\beta_i$  are the coefficients to be estimated.

The log returns of 10-year US treasury bond daily prices and the S&P 500 daily index were use as risk free return and market portfolio return respectively.

**Table 4: CAPM Regression Summary**

Stock	Beta (Slope)	Alpha (Intercept)
AAPL	9.982e-01	9.222e-05
GOOG	1.0299035	-0.0001953
GS	8.983e-01	4.611e-05
MS	0.8318133	0.0002293
GSK	1.0272869	-0.0004999
JNJ	1.042e+00	-4.726e-06
TSLA	0.9417436	0.0011208
TM	9.911e-01	-3.913e-05
EMR	0.9874175	-0.0003422
WMT	1.0543319	-0.0003222

As shown in the above table, the Beta of all the ten companies are very close to one, suggesting that these stocks are closely correlated with the market and tend to move up and down with the market. For GOOG, GSK, JNJ and WMT, their Beta is slightly larger than one, meaning they are more exposed to the market risk and more volatile.

The intercept alpha is referred to as “excess return” or “abnormal rate of return”. The alphas of the stocks are very close to zero, because in CAPM, alpha represents the performance of single stock relative to a benchmark and some nonsystematic risk was not yet diversified.

## Two Stocks Paired Analyses

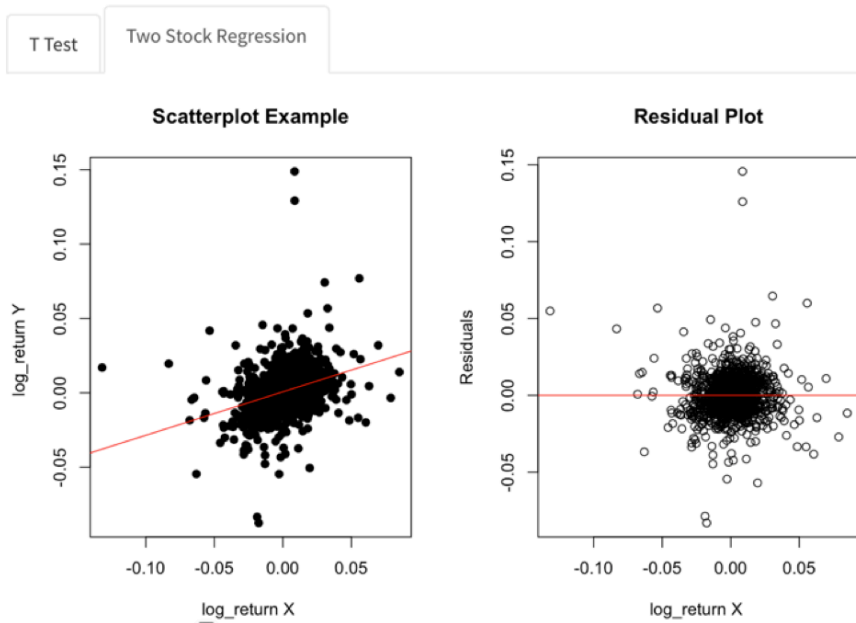
### (7) T-test and Linear Regression

For the analysis of two stocks, we firstly investigate the equality of two population means. With a confidence level of 95%, we used the two-sample t-test with unknown variance (which were assumed to be unequal) to determine whether the means of log returns were equivalent. The null hypothesis is that the true difference in means is equal to 0, and alternative hypothesis is that the true difference in means is not equal to 0. In this report, we illustrate the comparison of stock AAPL with the nine other stocks.

**Table 5: AAPL Two Sample T-test and Linear Regression**

Stock	T test P-Value	Slope	Intercept	Adjusted R <sup>2</sup>
GOOG	0.7508	0.2923594	0.0006901	0.1094
GS	0.8698	0.3218335	0.0002725	0.1124
MS	0.9629	0.1854376	-0.0001183	0.07708
GSK	0.3112	0.1364919	0.0004133	0.06829
JNJ	0.8952	0.3673757	0.0003896	0.08393
TSLA	0.3301	0.3769642	0.0013429	0.03795
TM	0.8241	0.2331104	0.0002936	0.08037
WMT	0.4965	0.1210624	0.0001119	0.03577
EMR	0.4681	2.795e-01	-3.765e-05	0.1182

As shown in the above table, since the P-values are larger than the significance level (0.05), we accept the null hypothesis, indicating that the population means of the two stocks are equal with confidence level of 95% regardless of the industry. This is consistent with the results in confidence interval of mean in part 1, showing that the means of log return are around zero.



We then used a linear regression to investigate any correlation between the log-returns of two stocks. The above two tables shows the linear regression plot and residual plot. It is easy to see a positive correlation among all different pairs of stocks among our data sets. The  $R^2$  indicates the percentage of the response variable variation that is explained by a linear model. Thus we see that any pair of stocks are somehow linearly related.

## (8) Correlation Coefficient Matrix

```
[1] "Correlation Coefficient matrix"
```

	AAPL	GOOG	GS	GSK	JNJ	MS	TSLA	TM	WMT	EMR
AAPL	1.0000	0.3319	0.3364	0.2789	0.2627	0.2910	0.1968	0.2848	0.1912	0.3449
GOOG	0.3319	1.0000	0.4043	0.3187	0.3779	0.3842	0.2664	0.3090	0.2339	0.3807
GS	0.3364	0.4043	1.0000	0.4275	0.4749	0.8632	0.2599	0.4843	0.2707	0.5666
GSK	0.2789	0.3187	0.4275	1.0000	0.4656	0.3770	0.2236	0.3842	0.2467	0.3920
JNJ	0.2627	0.3779	0.4749	0.4656	1.0000	0.4393	0.2021	0.4171	0.3867	0.4704
MS	0.2910	0.3842	0.8632	0.3770	0.4393	1.0000	0.2622	0.4786	0.2555	0.5409
TSLA	0.1968	0.2664	0.2599	0.2236	0.2021	0.2622	1.0000	0.2135	0.1329	0.2704
TM	0.2848	0.3090	0.4843	0.3842	0.4171	0.4786	0.2135	1.0000	0.2484	0.4684
WMT	0.1912	0.2339	0.2707	0.2467	0.3867	0.2555	0.1329	0.2484	1.0000	0.2833
EMR	0.3449	0.3807	0.5666	0.3920	0.4704	0.5409	0.2704	0.4684	0.2833	1.0000

From the correlation coefficient matrix, we can find that all the ten stocks are positively correlated. In order to get rid of the unsystematic risks, we suggest our client to consider several stocks with relatively low correlation to achieve diversification effects. For example, Apple, Tesla and Walmart are less correlated, by holding stocks among different industries, the client would be able to optimize his/her risks.

## Conclusion

By investigating single stocks, we found that the distributions of stocks' log-returns are approximately normal and random. The means of log-return are close to zero and the variance are close to zero as well. The regression on time shows that there is no linear relationship between the log-returns and time. Using CAPM regression, we have calculated the historical beta, i.e., the sensitivity of these ten stocks to market.

For the two stock analysis, the population mean of arbitrage two stocks are equal with confidence level of 95%. And the regression of one log-return on the other indicates that any pair of stocks have positive linear correlation.

Finally, the multiple stocks analysis yields a correlation coefficient matrix that provides the clients a guideline to diversify his/her portfolio by selecting stocks with low correlations to eliminate unsystematic risks.

## References

Gold, R., & Driebusch, C. (2016, 12 9). *U.S. Stocks Close at Records*. Retrieved from The Wall Street Journal: <http://www.wsj.com/articles/european-stocks-inch-up-after-ecb-prolongs-stimulus-148127380>