ST 501-001 Final Project Report

Sandeep Girada, Tanmay Sah, Alex Tunnell

December 1, 2018

I have neither given nor received unauthorized aid on this assignment. Signature and last four digits of our student ID:
Sandeep Girada (7995), Tanmay Sah (5725), Alex Tunnell (6731).

Introduction

For this project, Adjusted Closing Prices for Alphabet Inc. (GOOGL) and for Amazon.com, Inc. (AMZN) was required for the time period between Nov 11, 2017 and Nov 11, 2018. This data was extracted from Yahoo Finance. We used the R programming language to perform statistical analysis on the data. We also used many packages in R to not only understand our data, but to also explore possibilities outside the instructions given for this project. The data (251 trading days) was used to obtain the log-returns for the stocks. The 0.05 sample quantiles were calculated using the default Quantile function in R. We went through an exhaustive list of distributions and fitting methods in search of a good fit for the given data. We want to show a comparison between two distributions which have shown a significant fit for both GOOGL and AMZN. The two distributions we will be using for comparison in our report are Logistic and Johnson Unbounded (S_U).
We have observed that the leptokurtic nature of the Johnson distribution makes it a more suitable distribution for the Expected Shortfall calculations compared to the logistic. We have also verified all of our findings by writing our own distribution functions for the families of distributions. For this report we are only including inbuilt R package methods for all calculations.

Problem 1 (GOOGL)

Let P_t = Adjusted closing price of t-th day for t = 0, 1, 2, . . . , 250 denotes trading days, such that t = 0 represents Nov 13th, 2017 and t = 250 denotes Nov 9th, 2018. Compute the log-return Rt = $\log(P_t/P_{t-1})$ for t = 1, 2, . . . , 250.

Part (a) Obtain the 0.05 sample quantile $Q_{0.05}$ of $R_t's$ (using the quantile function in R) and also obtain the sample estimate of $E[R|R < Q_{0.05}]$, mean the of those R_t values for which $R_t < Q_{0.05}$.

Answer to (a):

$Q_{0.05}$ = -0.026922 and $E[R|R < Q_{0.05}]$ = -4.2212%

Part (b) Plot the histogram of log-return values using the plotdist function from fitdistrplus package. Next use the descdist function to make a good guess of the probability distribution of $R_t's$ and specify the name of the chosen family of distributions.

Answer to(b):

Figure 1: Plotdist of GOOGL data

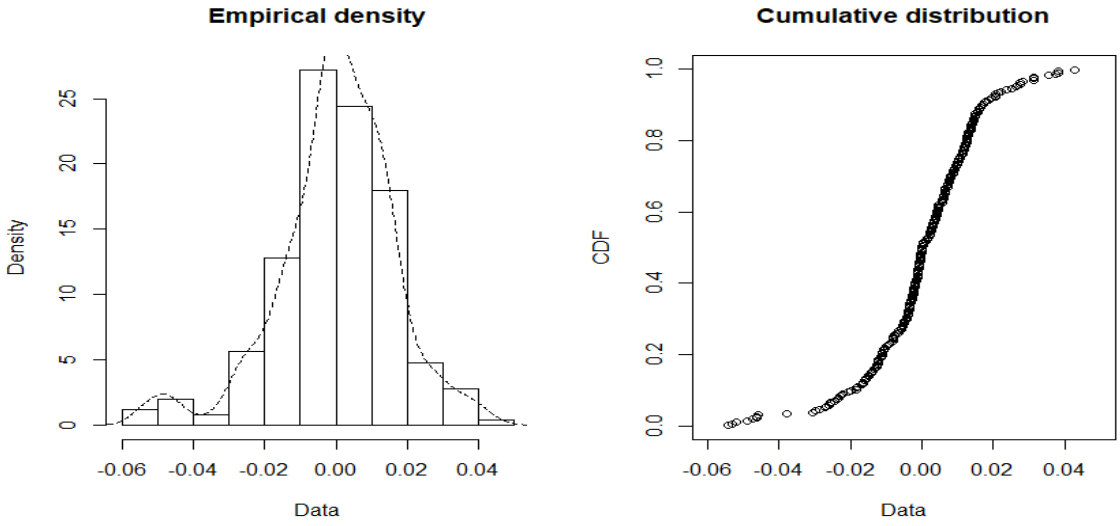
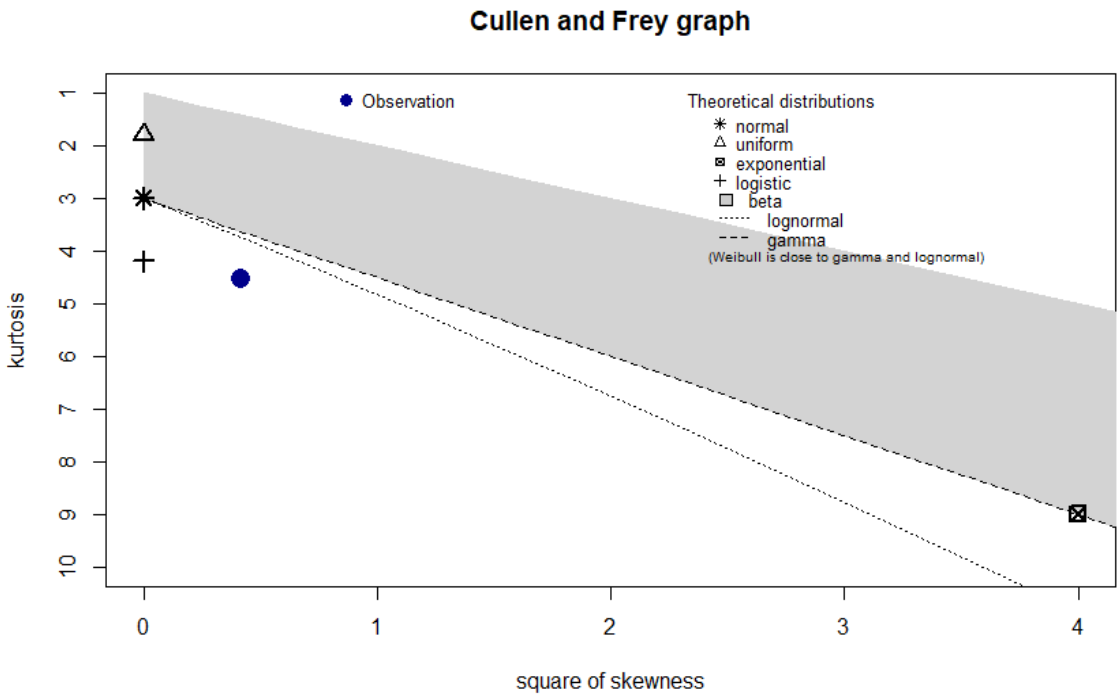


Table 1: Summary Statistics (GOOGL)

Maximum	4.302E-02
Minimum	-5.425E-02
Median	9.329E-05
Mean	1.353E-04
Estimated SD	1.637E-02
Estimated Skewness	-6.473E-01
Estimated Kurtosis	4.533E+00

Figure 2: Cullen and Frey for GOOGL using descdist



From the graph, it is evident that the distribution has more in common with logistic than the rest. Given that the data can take positive and negative values we have ruled out all distributions with a positive support or those with bounded distributions. The bootstrap method did not help in giving more insight about our data than what we already know.

We have chosen the Logistic distribution and Johnson Unbounded distribution to do a comparative analysis.

Part (c) Use the function fitdist to obtain the parameters of the chosen distribution and then use ad.test function from goftest package to determine the suitability of your chosen distribution. A p-value above 0.1 is acceptable (the higher the p-value the more points a team will receive!)

Answer to (c):

We have slightly adjusted the estimations for location and scale with the standard error terms and checked continuously for an increase in pvalue. For GOOGL our location estimates was the same as computed but the scale was reduced by half of the standard error term to increase fitting. We achieved a 0.02 increase in pvalue and 0.1% increase in Expected Shortfall as a result of adjustment.

Table 2: Summary from fitdist (GOOGL Logistic)

	Estimate	Std. Error
Location	0.000941299	0.000945213
Scale	0.008714979	0.000447401
Logliklihood	682.1188	
AIC	-1360.238	
BIC	-1353.195	
Correlation Matrix:		
	Location	Scale
Location	1	-0.02618382
Scale	-0.02618382	1

Table 3: Summary from ad.test (GOOGL Logistic)

Null Hypothesis	Logistic Distribution
location	0.000941299
Scale	0.008491278
An	0.89334
p-value	0.4183

Table 4: Modified Summary from JohnsonFit (GOOGL Johnson SU)

	Length	Value	Mode
gamma	1	0.084167168	numeric
delta	1	1.64712	numeric
xi	1	0.005173417	numeric
lambda	1	0.02247362	numeric
type	1	"SU"	character

We could not adjust the Johnson parameter estimates, as the fitting function in SuppDists package does not allow much manipulation.

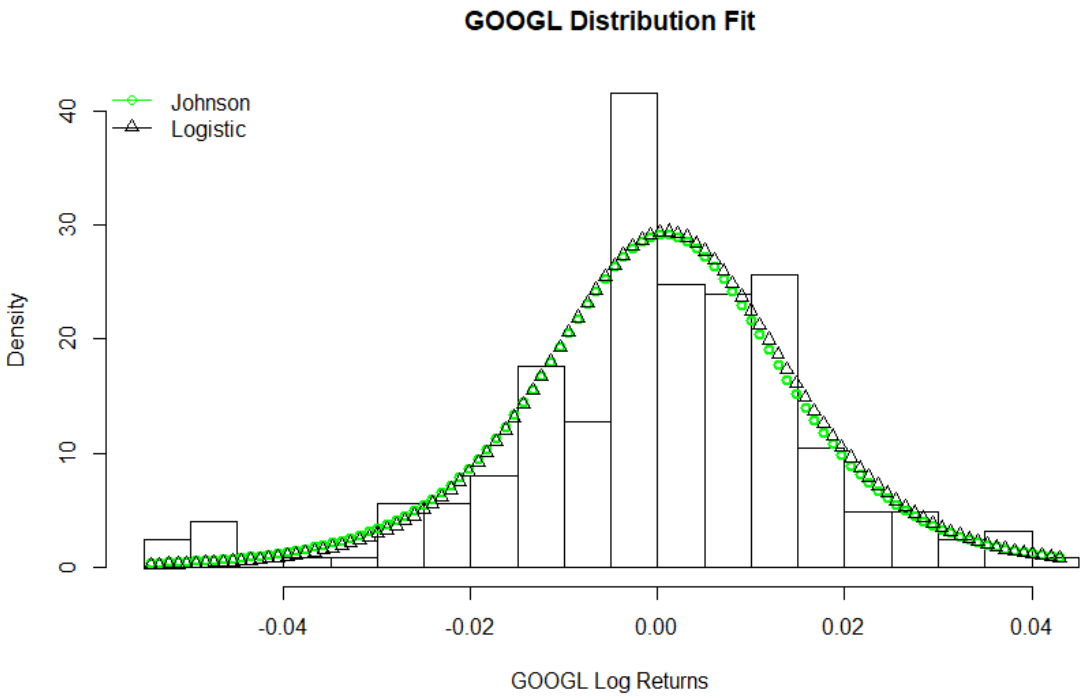
Table 5: Summary from ad.test (GOOGL Johnson SU)

Null Hypothesis	pJohnson
Parameter 1	0.084167168
Parameter 2	1.647119893
Parameter 3	0.001573417
Parameter 4	0.022473622
An	0.90041
p-value	0.4139

Logistic p-value = 0.4183

Johnson Unbounded p-value = 0.4139

Figure 3: Fitting the numerically fit distributions over GOOGL data



Given the close p-values it is very hard to differentiate between the two distributions, but the Expected Shortfall calculations for Johnson are higher given their fat tails.

Part (d) Let $f_R(r)$ denote the chosen probability density function (PDF) of the log-returns and write down its exact expression. Show that for any constant c,

$$E(R|R<c) = \int_{-\infty}^{c} \frac{r f_R(r)}{F_R(c)} dr \;\; where \;\; F_R(c) = \int_{-\infty}^{c} f_R(r) dr$$

Use your chosen PDF and the above formula to compute the following summaries (using Monte Carlo or numerical integration methods if needed):

$Q_{0.05}(R) = F_R^{-1}(0.05)$ and $E[R|R < Q_{0.05}]$

and compare the values with the corresponding empirical values that you obtained in (a).

Answer to (d):

The PDF for the Logistic distribution is given by [1]:

$$f_R(r) = \frac{exp(\frac{r-m}{s})}{s(1+exp(\frac{r-m}{s}))^2} \tag{1}$$

Where "m" is the mean or location and "s" is the scale.

The Johnson SU distribution has a PDF with the form [2]:

$$f_R(r) = \frac{\delta}{\sqrt{2\pi((r-xi)^2+\lambda^2)}} exp(-0.5(\gamma + \delta ln(\frac{r-xi+\sqrt{(r-xi)^2+\lambda^2}}{\lambda}))^2) \tag{2}$$

Where γ and δ are the shape parameters. The location-scale parameters are λ and xi.

The proof for part d is as follows:

From the definition of conditional probability we know that:

$$P(R|B) = \frac{P(R \cap B)}{P(B)} \tag{3}$$

Where B = R<c and for our case "c" is constant. Therefore, applying this approach to our situation we see:

$$f_{R|R<c}(r|r<c) = \frac{f_R(r)}{P(R<c)} \tag{4}$$

Where

$$P(R<c) = F_R(c) = \int_{-\infty}^{c} f_R(r) dr \tag{5}$$

Therefore (3) becomes

$$f_{R|R<c}(r|r<c) = \frac{f_R(r)}{F_R(c)} \quad For\ r < c\ and\ 0\ otherwise \tag{6}$$

Given that

$$E(R|R<c) < \infty \tag{7}$$

We now find the expected value of (6) which is:

$$E(R|R<c) = \int_{-\infty}^{c} r f_{R|R<c}(r|r<c)dr \tag{8}$$

This then becomes:

$$E(R|R<c) = \int_{-\infty}^{c} \frac{r f_R(r)}{F_R(c)}dr \tag{9}$$

Table 6: Comparisons of the Quantiles and Expected Shortfalls

	Johnson SU	Logistic
From Distribution Q0.05	-0.02659533	-0.02406
From Distribution ES	-3.79%	-3.277%
	From Data:	
Part (a) Q0.05	-0.026922	
Part (a) ES	-4.22%	

The Expected Shortfall calculated using Johnson is much closer to the empirical estimate calculated from the actual data. Our adjustment of scale parameter for GOOGL has actually increased the Expected Shortfall from Logistic distribution and brought it closer to the empirical estimate.

Problem 2 (AMZN)

This problem is the same as problem one except we are dealing with AMZN not GOOGL. In addition, we do not need to prove the formula for $E(R|R < c)$ as that was completed in problem one.

Answer to (a):

$Q_{0.05}$ = -0.032286 and $E[R|R < Q_{0.05}]$ = -4.8962%

Answer to (b):

Figure 4: Plotdist of AMZN data

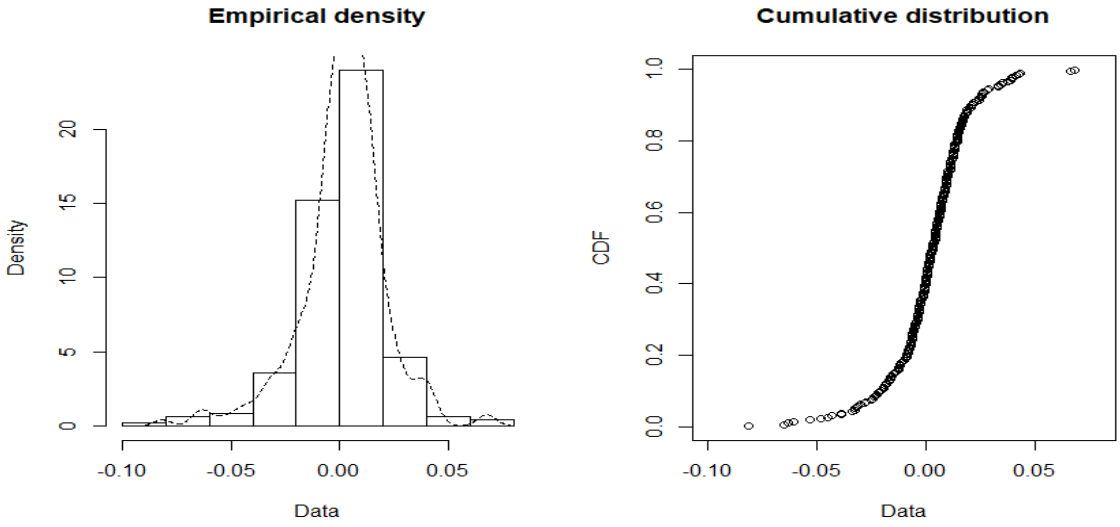
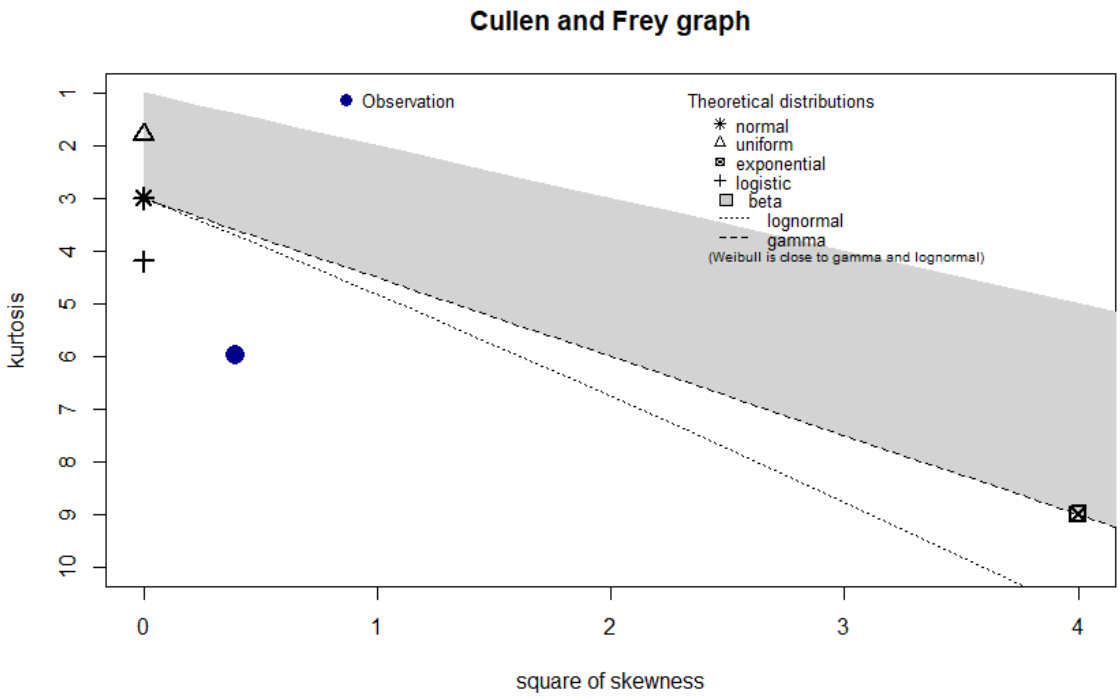


Table 7: Summary Statistics (AMZN)

Maximum	6.849E-02
Minimum	-8.142E-02
Median	3.101E-03
Mean	1.666E-03
Estimated SD	1.942E-02
Estimated Skewness	-6.249E-01
Estimated Kurtosis	5.978E+00

Figure 5: Cullen and Frey for AMZN using descdist



Answer to (c): We have slightly adjusted the estimations for location and scale with the standard error terms and checked continuously for an increase in p-value. For AMZN our location estimates was the same as computed but the scale was reduced by 1.2 times the standard error term to increase fitting. We achieved a 0.07 increase in p-value and 0.25% decrease in Expected Shortfall as a result of adjustment.

Table 8: Summary from fitdist (AMZN Logistic)

	Estimate	Std. Error
Location	0.002561909	0.0010628
Scale	0.009902025	0.0005189
Logliklihood	646.534	
AIC	-1289.068	
BIC	-1282.025	
Correlation Matrix:		
	Location	Scale
Location	1	-0.028345
Scale	-0.02834509	1

Table 9: Summary from ad.test (AMZN Logistic)

Null Hypothesis	Logistic Distribution
location	0.002561909
Scale	0.009279349
An	1.1824
p-value	0.2743

Table 10: Summary from JohnsonFit (AMZN Johnson SU)

	Length	Value	Mode
gamma	1	0.150938	numeric
delta	1	0.9153748	numeric
xi	1	0.005074599	numeric
lambda	1	0.01066301	numeric
type	1	"SU"	character

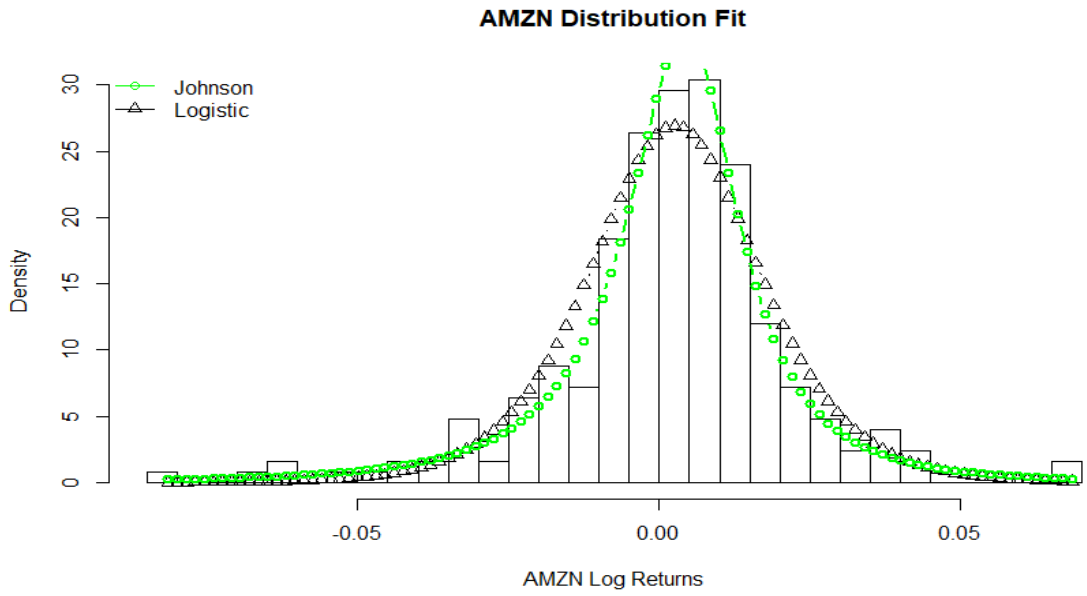
Logistic p-value = 0.2743

Johnson Unbounded p-value = 0.992

Table 11: Summary from ad.test (AMZN Johnson SU)

Null Hypothesis	pJohnson
Parameter 1	0.150938034
Parameter 2	0.915374796
Parameter 3	0.005074599
Parameter 4	0.010663007
An	0.19379
p-value	0.992

Figure 6: Fitting the numerically fit distributions over AMZN data



Given the stark difference in p-value. The distributions are clearly distinguishable for AMZN.

Answer to (d):

Table 12: Comparisons of the Quantiles and Expected Shortfalls

	Johnson SU	Logistic
From Distribution Q0.05	-0.0320944	-0.02476
From Distribution ES	-6.07%	-3.428
	From Data	
Part (a) Q0.05	-0.032286	
Part (a) ES	-4.896%	

The Expected Shortfall calculated using Johnson is much higher than the empirical estimate calculated from the actual data. Our adjustment of scale parameter for AMZN has actually decreased the Expected Shortfall from Logistic distribution and moved it further away from the empirical estimate. Thus we conclude that increasing p-value does not actually result in a better Expected Shortfall estimation.

Problem 3

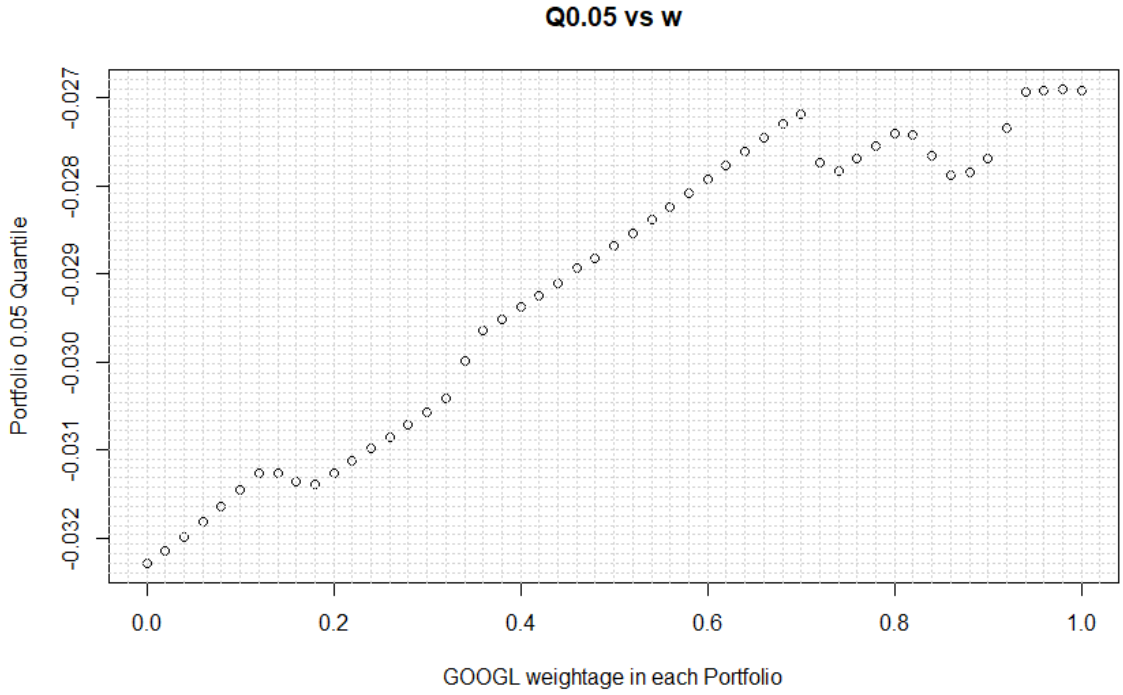
Let P_t^G and P_t^A denote the adjusted closing prices of Google and Amazon, respectively for t = 1, 2, . . . , 250 (as described in problems 1 2 above). Consider a portfolio that combines the two prices as $P_t(\omega) = \omega P_t^G + (1-\omega)P_t^A$ where $\omega \in [0,1]$ is a weight to be determined.

Let $R_t(\omega) = log(P_t(\omega)/P_{t-1}(\omega))$ denote the log return of the portfolio for an arbitrary value of ω.

Part (a) For each $\omega \in [0,1]$ obtain the 0.05 sample quantile $Q_{0.05}(\omega)$ of the $R_t(\omega)$ and plot $Q_{0.05}(\omega)$ as a function of ω on grid of equally spaced 50 values in [0, 1].

Answer to (a):

Figure 7: Portfolio 0.05 Quantiles with respective weights

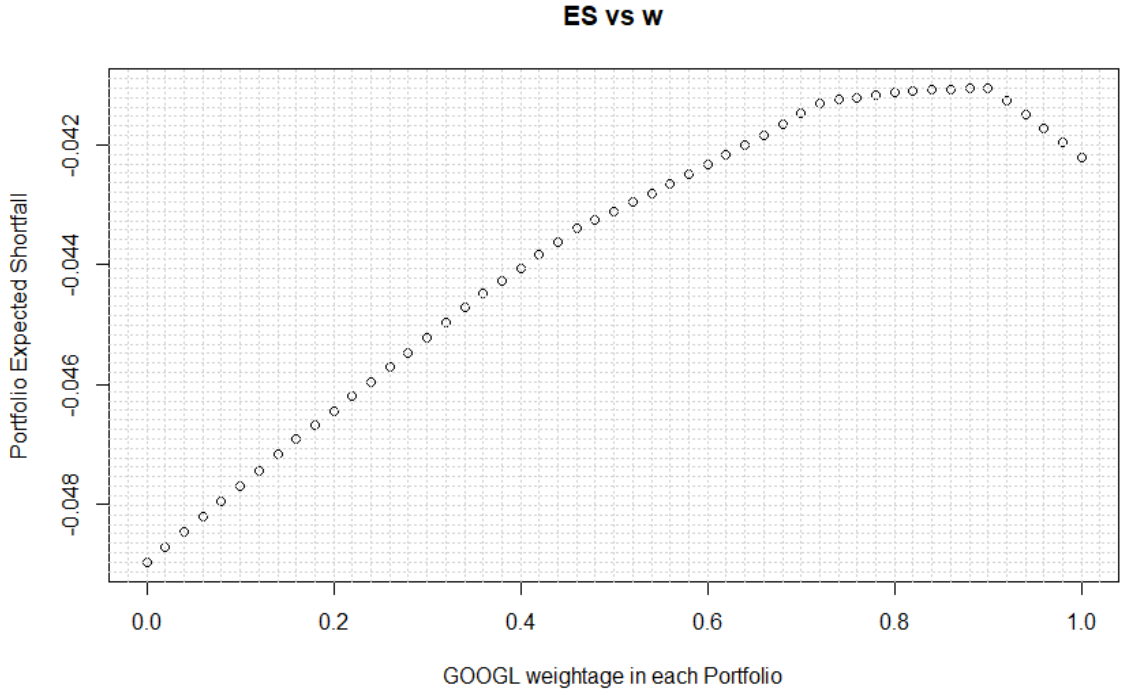


We have generated a data frame of portfolio prices and Log returns to calculate individual quantiles.

Part (b) For each $\omega \in [0,1]$ obtain the sample estimate of $ES(\omega) = E[R(\omega)|R(\omega) < Q_{0.05}(\omega)]$ and plot $ES(\omega)$ as a function of ω on grid of equally spaced 50 values in [0, 1].

Answer to (b):

Figure 8: Portfolio Expected Shortfall with respective weights



Part (c) Find value(s) of ω for which $ES(\omega)$ is minimized and maximized based on a grid search of 50 values.

Answer to (c):

The values of ω for which $ES(\omega)$ is maximized and minimized is:

ω= 0.88 and ω= 0 respectively

Conclusion

We believe that a fat tail distribution like Johnson is a much better fit over the Logistic. It can also be noted that an increasing p-value does not actually result in an increased accuracy of the Expected Shortfall estimate; as the p-value is only an indicator for fit over the entire data while Expected Shortfall is more concerned with the fit of the tail.

We also believe that the empirical estimate of Expected Shortfall may not be a very good estimator, but definitely a useful one given the limited amount of data we have.

References

[1]

PennState, Center for Astrostatistics, The Logistic Distribution, n.d., Viewed 1 December 2018, <http://astrostatistics.psu.edu/su07/R/library/stats/html/Logistic.html>.

[2]

A. R. Godfrey, RDocumentation, JohnsonSU, n.d., Viewed 1 December 2018, <https://www.rdocumentation.org/packages/ExtDist/versions/0.6-3/topics/JohnsonSU>.