FTEC5580 Project 1

(1) Find a parametric distribution that fits well to the net return data of your stock. Here, let's simply ignore that a net return has to be greater than−1 and you can use distributions living on the real line such as normal to model net returns. In this part, you must show sufficient analysis to justify your choice.

First, descriptive statistics on the NVDA net return. As shown in Figure 1, NVDA's net return mostly fluctuates within the range of -10% and 10%.

Table 1. NVDA Net Return Basic statistics

Mean	Stand Deviation	Skewness	Kurtosis
0.1489641	2.7453	-0.5030	9.7634

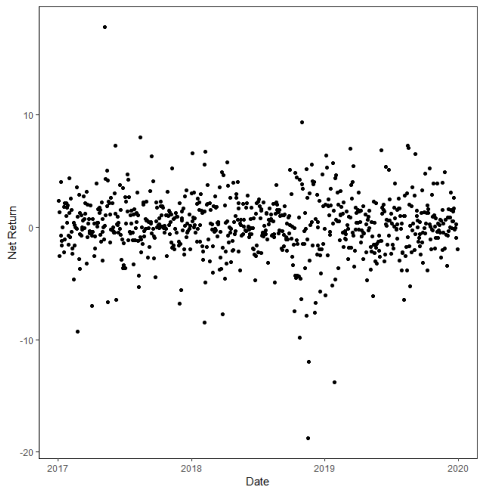

Figure 1. Net Return Scatter Plot

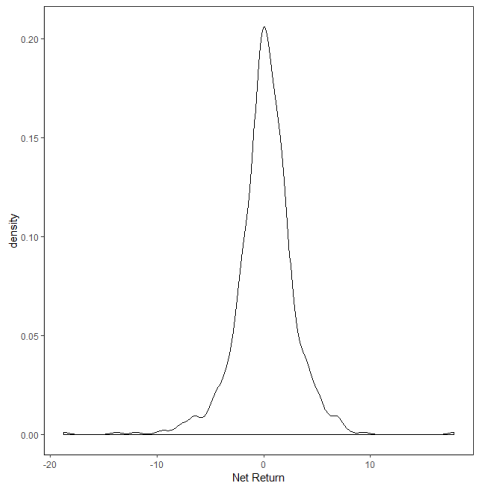

Figure 2. Net Return Probability Density Function

Next, fit the data to a parametric distribution. The R function with "mle" method is used to fit the net return into a normal distribution, a logistic distribution, and a t-distribution. By comparing the model diagnostic plot and loglikelihood statistics, the t distributions fit significantly better in the PP and QQ plots than others and have the smallest log-likelihood statistics. Then the goodness-of-fit test was performed on the three fitted distributions, and the K-S statistic for the t distribution was 0.0188. Look up the critical value in the K-S table value, I have 753

observations in my sample. At an alpha level of 0.05, the K-S table value is 0.0496. Since 0.0188 is less than 0.0496, the null hypothesis (that the distributions are the same) is accepted. In summary, the **t-distribution** is considered to be the most suitable parametric distribution that fits well to the net return data of NVDA.

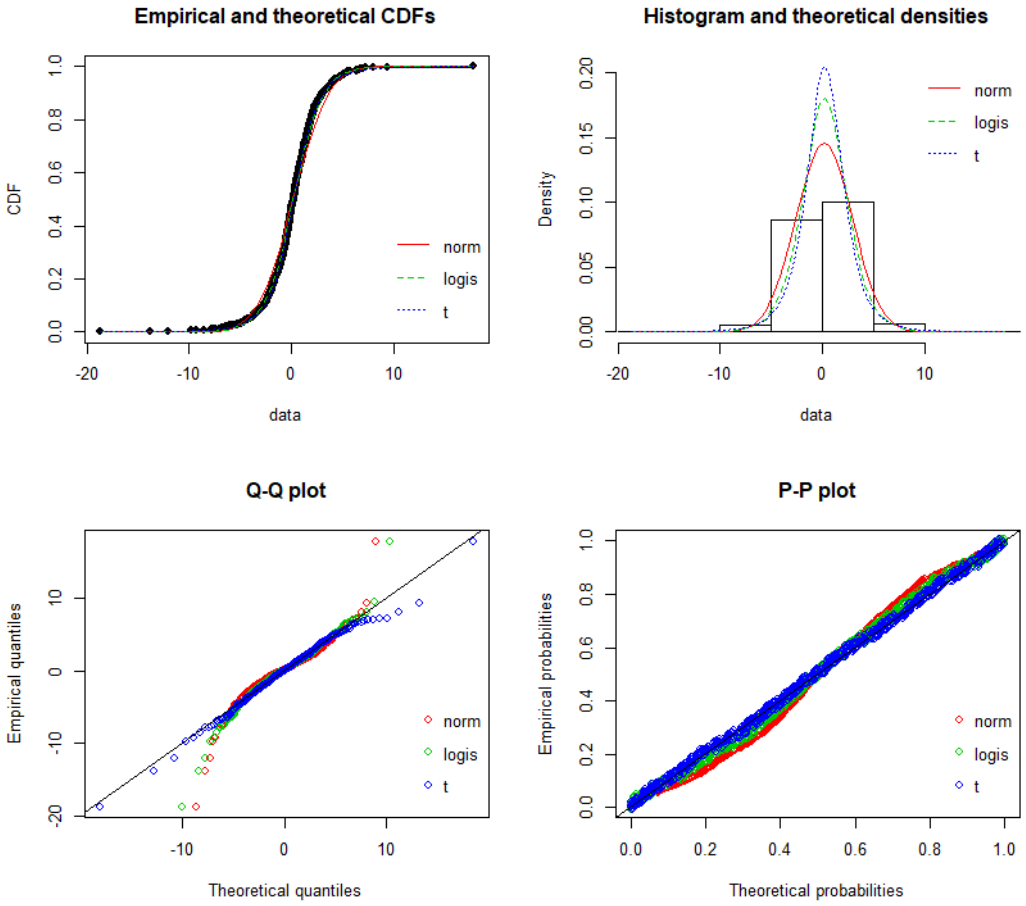

Figure 3. Model Diagnostic plot

Table2. Loglikelihood Statistics

	normal	**logistic**	**t**
Loglikelihood	1825.97	1768.31	1753.08

Table3. Goodness-of-fit Statistics

	normal	logistic	t
Kolmogorov-Smirnov statistic	0.0775	0.0384	**0.0188**
Cramer-von Mises statistic	1.5601	0.2522	0.0344
Anderson-Darling statistic	9.0042	1.6322	0.2772
Akaike's Information Criterion	3655.9430	3540.6120	3512.1520
Bayesian Information Criterion	3665.1880	3549.8580	3526.0200

(2) Train the Fama-French three factor model. Check the model assumptions and address any problems that occur. Is alpha= 0 for your stock? Which factors are statistically significant?

a) Make OLS linear regression model for data. The model results are as follows.The intercept of the model is 0.019, so the alpha is not equal to zero. Within the 99% confidence interval, Mk,t-RF, SMB and HML are all statistically significant. The model passed the F test, indicating that the model as a whole is significant, but the adjusted R-squared is only 40.35%.

Table4. Model Output

	Estimate	Std. Error	t value	Pr (>|t|)
(Intercept)	0.019	0.078	0.250	0.802
Mkt-RF	1.919	0.096	20.073	<0.001
SMB	0.352	0.160	2.201	0.028
HML	-0.728	0.142	-5.144	<0.001

Residual standard error: 2.12 on 748 degrees of freedom
Multiple R-squared: 0.4059, Adjusted R-squared: 0.4035
F-statistic: 170.3 on 3 and 748 DF, p-value: < 2.2e-16

b) Diagnose the model and check the model assumptions.

First, test whether the residuals are normally distributed. As shown in Figure 4, some points fall outside the confidence interval. After the Shapiro-Wilk normality test, the p-value is approximately 0, so the model's residuals do not have a normal

distribution, and **the assumption of residuals are normally distributed does not hold.**

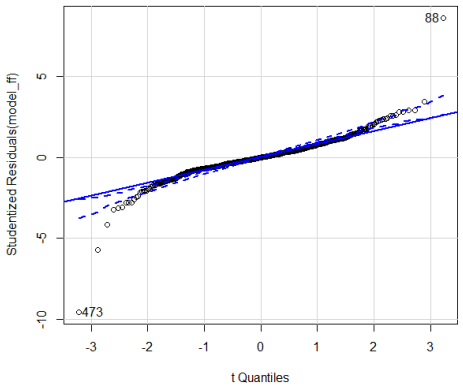

Figure 4. Residual Q-Q Plot

Second, test whether the errors have the same variance. After the Non-constant Variance Score test, the p-value is equal to 0.400, and **the error variance invariance hypothesis holds**.

Third, test whether the errors are uncorrelated. After Durbin-Watson test, the p-value is equal to 0.402, indicating that there is no autocorrelation between the error terms. **The assumption of errors are uncorrelated holds.**

Forth, test whether the predictors affect the response in a linear way. Through the component plus residual plot, we can see that there is a linear relationship in the graph. **The assumption of the predictors affect the response in a linear way holds.**

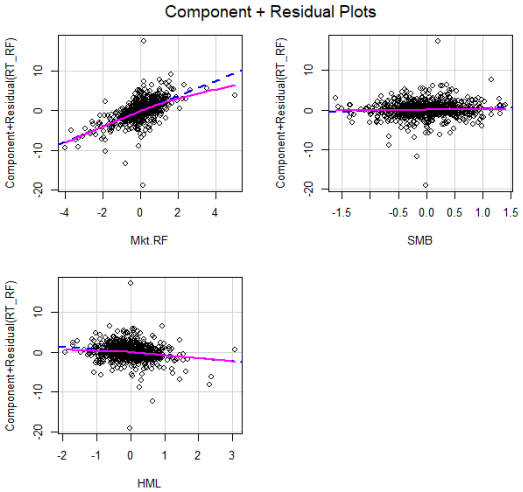

Figure 5. Component Plus Residual Plot

c) Test Multicollinearity. Using Variance Inflation Factor to test for multicollinearity, the results show that there is no multicollinearity problem.

Table5. VIF Test Output

	Mkt-RF	**SMB**	**HML**
VIF	1.056693	1.025203	1.031766

(3) Compare the CAPM model and the Fama-French three factor model in terms of adjustedR2, Cp, BIC, and CV test error (consider 10-fold CV and LOOCV). Which model is better? In this part, we can assume the typical assumptions for linear regression are satisfied.

The model's formulas are as follows:

$$CAPM: R_t - \mu_t = \alpha + \beta_1(R_{M,t} - \mu_t) + \in_t$$

$$Fama - French: R_t - \mu_t = \alpha + \beta_1(R_{M,t} - \mu_t) + \beta_2 SMB_t + \beta_3 HML_t + \in_t$$

From table6, Fama-French three factors model has lager adjusted R^2 and smaller C^p, BIC and test error both in 10-fold CV and LOOCV test error. In summary, the Fama-French three factor model is the better one.

Table 6. The Comparison of Two Models

	Adjusted R^2	**C^p**	**BIC**	**10-fold CV test error**	**LOOCV test error**
CAPM	0.380	31.406	-347.387	4.479	4.688
Fama-French	0.404	4.000	-365.071	4.257	4.523