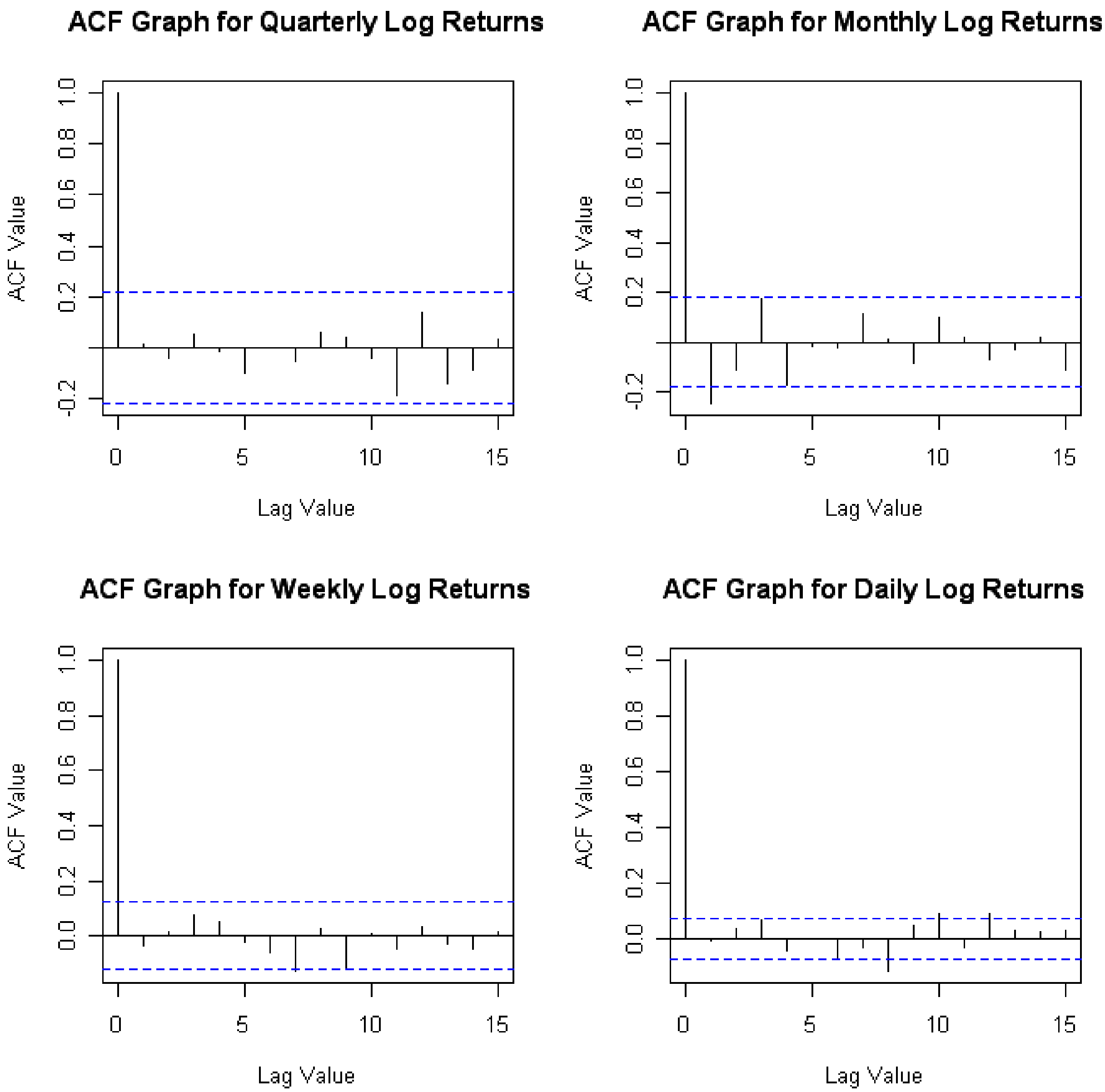
[redacted]

MA 641 – Assignment 2 Report

06-07-08

Problem I: Please plot the first 4 ACF plots on one page (each with its own title and axis labels) and hand that picture in.



Above are all the ACF's on one page as requested.

2. For the CAPM to be valid there should be no serial correlation between successive returns. Does this assumption appear to be valid based on your dataset? Compare the validity of this assumption using the various time frequencies.

To test if there CAPM model is valid based on the various time frequencies, we need to use the Portmanteau Ljung&Box test on each of the frequencies.

The results are listed below.

Daily Returns
Box-Ljung test

data: logdailyreturns
X-squared = 37.2172, df = 15, p-value = 0.001176

Weekly Returns
Box-Ljung test

data: logweeklyreturns
X-squared = 13.5627, df = 15, p-value = 0.5589

Monthly Returns
Box-Ljung test

data: logmonthlyreturns
X-squared = 22.6876, df = 15, p-value = 0.091

Quarterly Returns
Box-Ljung test

data: logquarterreturns
X-squared = 10.2227, df = 15, p-value = 0.8055

In each of the frequencies, for the various time frequencies to be able to say that each obeys a CAPM model, we must have the p-values be less than 5%, the typical given alpha level of significance.

Only the daily return data has this property, which makes sense. The daily log return data has a p-value of 0.001176, which is quite significant. This p-value indicates that there is serial correlation between successive returns, which means that past returns play an important role in predicting future returns. This is evident due to the fact that the returns are of very high frequency. The higher frequency data would be expected to show significant correlation. If one thinks realistically, in the stock market, the past day's closing price would highly influence the next day's immediate closing price.   However, the shorter frequency data, ie: the more days, or months spread apart from return data, the less likely the previous return value will play an influence on the successive return value. Thus, one would expect realistically that there be no successive correlation, and that the CAPM model would not be plausible. This is the case with weekly, monthly, and quarterly return data.

Note: the degrees of freedom, df is the lag which I set to be 15 which I thought to be appropriate.

3. What is the difference in the autocorrelation function when looking at data sampled at various frequencies?

To answer this question, we have to look at equations 2.2, and 2.3

$$Equation\ 2.2 = \quad p\hat{}_l = \frac{\sum_{t=l+1}^{T}(r_t - \mu_r)(r_{t-l} - \mu_r)}{\sum_{i=1}^{T}(r_t - \mu_r)^2}$$

$$Equation\ 2.3 = \quad Q(m) = T(T+2)\sum_{l=1}^{m}\frac{p\hat{}_l}{T-l}$$

For each frequency, the T is what changes.

Quarterly lag-l sample ACF	Monthly lag-l sample ACF	Weekly lag-l sample ACF	Daily lag-l sample ACF
[,1]	[,1]	[,1]	[,1]
[1,] 1.000000000	[1,] 1.00000000	[1,] 1.000000000	[1,] 1.000000000
[2,] 0.011910306	[2,] -0.24578333	[2,] -0.036194513	[2,] -0.007408024
[3,] -0.041798615	[3,] -0.10887302	[3,] 0.015571340	[3,] 0.033760269
[4,] 0.050742146	[4,] 0.17138504	[4,] 0.072719970	[4,] 0.064972448
[5,] -0.013456762	[5,] -0.17368347	[5,] 0.047926982	[5,] -0.041204436
[6,] -0.102864366	[6,] -0.01659613	[6,] -0.021530495	[6,] 0.001000025
[7,] -0.002037081	[7,] -0.02364262	[7,] -0.062618661	[7,] -0.074060158
[8,] -0.055821265	[8,] 0.11170125	[8,] -0.127268000	[8,] -0.029683964
[9,] 0.059362958	[9,] 0.01408017	[9,] 0.024885782	[9,] -0.115663722
[10,] 0.042255162	[10,] -0.08013735	[10,] -0.114625574	[10,] 0.048946757
[11,] -0.038134234	[11,] 0.09811048	[11,] 0.004641831	[11,] 0.088013318
[12,] -0.185755485	[12,] 0.01479431	[12,] -0.049772952	[12,] -0.032566488
[13,] 0.137584016	[13,] -0.06847515	[13,] 0.029909249	[13,] 0.089736703
[14,] -0.141385956	[14,] -0.02739230	[14,] -0.031472005	[14,] 0.032574346
[15,] -0.090118939	[15,] 0.01817994	[15,] -0.045728054	[15,] 0.024593277
[16,] 0.035856260	[16,] -0.10752107	[16,] 0.011978258	[16,] 0.028876080

The lower the frequency, the lower the value T is, the greater magnitude values there are overall in the lag-l sample ACF's. So, thus the frequency is the one factor that the lag-l sample ACF's depend upon. The lag-l sample ACF's depend upon the mean of the returns, which is influenced by the frequency. The lower the frequency, the higher the mean return magnitude wise, which will also come into play, thus one would expect the lowest frequency data to have sample ACF's of high extremes which does happen.

Now, it is time to answer the difference as in the test-statistics for the chi-squared test done in the Portmanteau Box-Ljung test.

Let's look at the Chi-squared test statistic values and try to determine another trend in the data:

Daily Returns
X-squared = 37.2172

Weekly Returns
X-squared = 13.5627

Monthly Returns
X-squared = 22.6876

Quarterly Returns
X-squared = 10.2227

$$Equation\ 2.3 = \quad Q(m) = T(T+2)\sum_{l=1}^{m} \frac{p^{\wedge}{}_{l}}{T-l}$$

Now, we expect that with higher frequency data, the lag-l sample ACF's are lower, which thus implies that in the summation sign of Equation 2.3, the summation is going to be quite small, however, the quadratic term in the Box-Ljung Test overrides this small value in the summation because of the high quadratic value using high frequency data. Thus, high frequency data would naturally tend to have higher test-statistics and thus have a statistical significance in showing that there is serial correlation at a certain fixed significance level of α. The data supports this, with the daily return data having the highest test statistics and the quarterly returns data having the lowest test statistic at a default level of α=5%.

==

4. Finally, test for normality of the returns. Perform a Jaque-Bera test to check normality for each of the data under consideration. What do you see? Explain.

What we must first do is calculate the t-ratios of test statistics for skewness and kurtosis for each frequency type.

Daily Returns
t-statistic for skewness: 6.090283
t-statistic for kurtosis: 59.04953
Jaque-Berra Value: 3523.938
p-value for Jaque-Berra Test: = 1-pchisq(JB,2) = ≅0

For daily return data, the t-statistic for kurtosis is very high, which accounts for the high Jaque-Berra test statistic which accounts for a very small p-value which is statistically significant at the 5% level. Thus, the null hypothesis is rejected of normality. Thus, daily log return data is not normally distributed. Both t-statistics are greater than $z_{\frac{\alpha}{2}}$. So, thus three tests of normality all fail with their null hypotheses for daily log return data.

We shall do the same analysis for Weekly Returns

Weekly Returns
t-statistic for skewness: -1.405587
t-statistic for kurtosis: 12.01468
Jaque-Berra Value: 146.3282
p-value for Jaque-Berra Test: = 1-pchisq(JB,2) = ≅0

Only the t-statistic for kurtosis and the Jaque-Berra Value show statistical significance and refute the null hypothesis of normality in distribution for weekly log return data.

Let's repeat the same analysis for monthly returns

Monthly Returns
t-statistic for skewness: 0.4141321
t-statistic for kurtosis: 4.627147
Jaque-Berra Value: 21.58200
p-value for Jaque-Berra Test: = 1-pchisq(JB,2) = 2.058395e-05

Again, the same two tests of using the t-statistic for kurtosis and the Jaque-Berra test statistic refute the null hypothesis of normality of data. So with 2/3 tests showing the data fails to be normally distributed, it is safe to say that monthly log return data is not normally distributed.

Let's repeat the same analysis for quarterly returns.

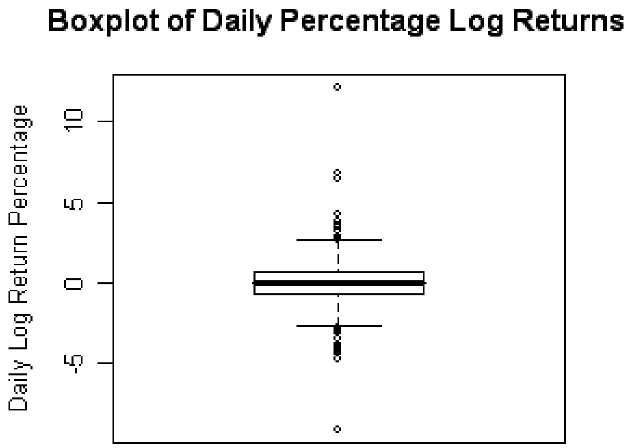
Quarterly Returns
t-statistic for skewness: 1.125770
t-statistic for kurtosis: 4.50318
Jaque-Berra Value: 21.54599
p-value for Jaque-Berra Test: = 1-pchisq(JB,2) = 2.095789e-05

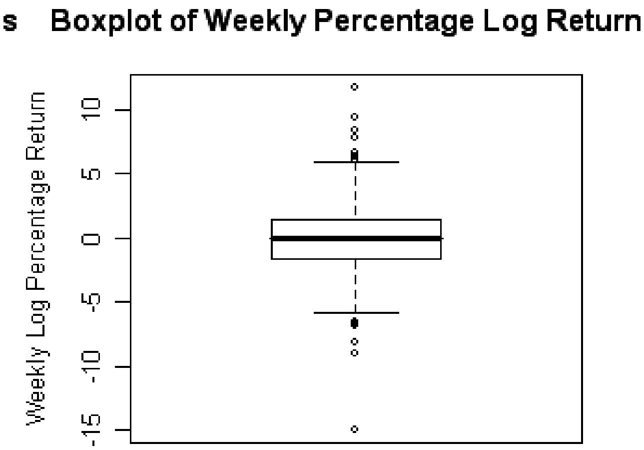
The same results follow again as far as the number and types of tests to prove or disprove normality.

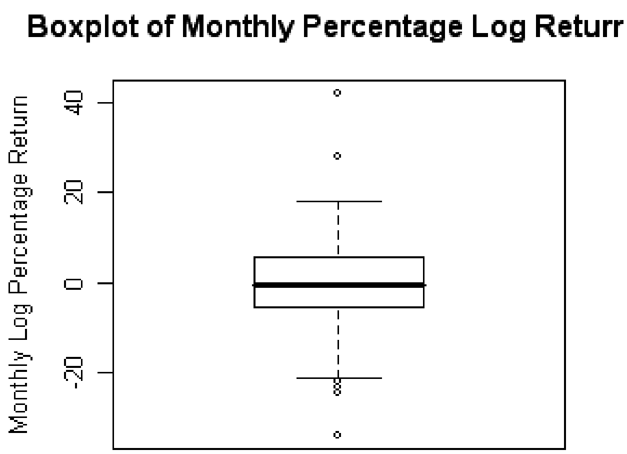
However, I would like to comment further.

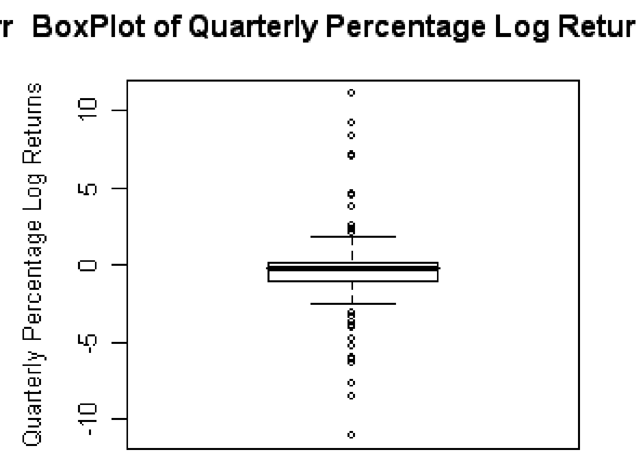
In the high frequency data, the t-statistic for skewness was quite high, and when the frequency was low, the t-statistic for skewness was quite low which meant that as far as skewness tests go, it would seem logical to say that as the frequency decreases, so does the skewness. The same trend follows for the t-statistic for kurtosis. And the same pattern of direct proportions for the Jaque-Berra test statistic follows. The higher the frequency of the data, the higher the test statistic became. The lower the frequency, the lower the JB value became. The lower these values are, the more likely intuitively it is for the data to be more normally distributed.

These trends follow for the most part, as is most evident in the page of box plots for each of the types of frequencies.


Boxplot of Daily Percentage Log Returns
Daily Log Return Percentage
10
5
0
-5


Boxplot of Weekly Percentage Log Return
Weekly Log Percentage Return
10
5
0
-5
-10
-15


Boxplot of Monthly Percentage Log Returr
Monthly Log Percentage Return
40
20
0
-20


BoxPlot of Quarterly Percentage Log Retur
Quarterly Percentage Log Returns
10
5
0
-5
-10