Financial Econometrics

Assignment Regression Analysis: Computing Beta Coefficients

Jens Matthé

Master Business Engineering

S0151880

Part 1: Beta coefficients and frequencies

1. First view to the data

The data includes the daily return indices of 50 Stocks together with the return and price index. The first trading day is January 2, 2001 and the last one April 4, 2020. There are missing values for Amadeus IT group, Deutsche Boerse, Engie, Inditex and Linde. In total the file contains 5027 recorded trading days.

2. The daily returns from 2015

The data is loaded as irregular time series with a daily frequency. The daily returns are computed with the lags. The function computes the first relative difference i.e. the percentage that indicates how the next value differs from the previous value for example a first relative difference of 0.97 indicates a decrease in value of 3% and a first relative value of 1.05 indicates an increase of 5%. The daily returns were computed as the first relative differences minus 1. This calculation is explained by the formula below.

$$Return_{t+1} = \frac{Index_{t+1}}{Index_t} - 1$$

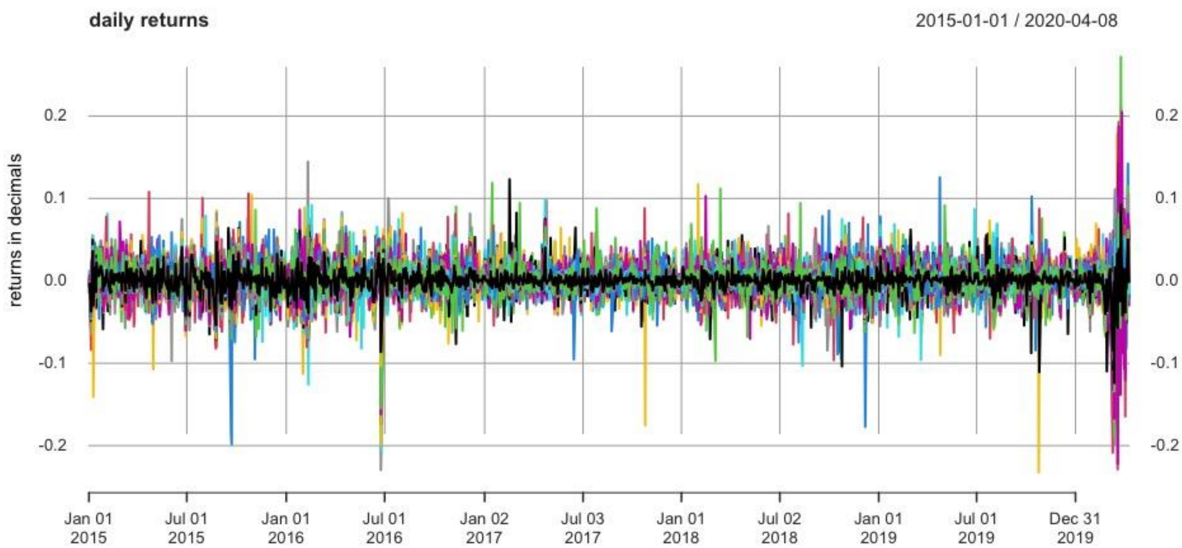

figure 1: daily returns in decimals

In figure 1 the daily returns are shown. The difference between the 50 stocks is not visible but it is clear that most of the returns are between 0% and 10% except from some peaks, mostly at the end of the dataset. It is also clear that the returns are oscillating a lot.

3. Single Index Model for daily returns

The Single Index Model is computed for every stock. For each stock a linear regression is done with the daily returns of the stocks as dependent variable and the daily return of the return index as independent variable. For the first order linear regression a β_0 is found as intercept and a β_1 as slope. The descriptive statistics for the betas are computed below.

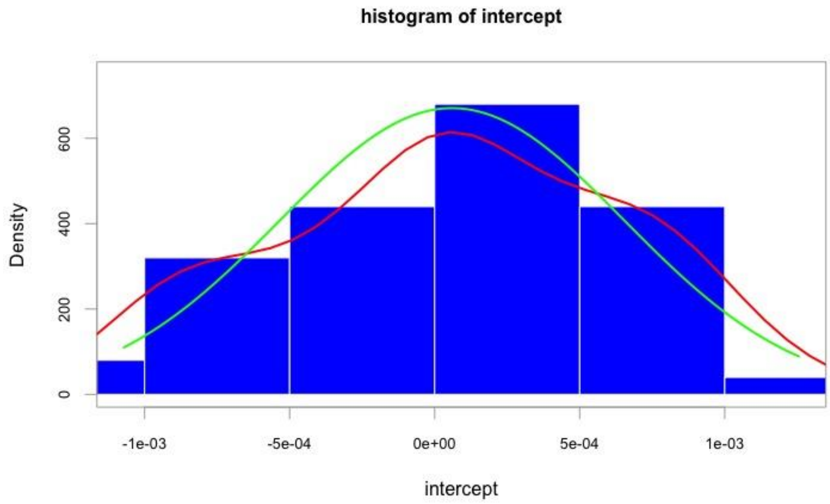

figure 2: histogram intercepts

The histogram of the 50 computed intercepts is together with the computed and normal density function shown in figure 2. Both the median as the mean is equal to 0, the maximum and minimum value is respectively 0.0013 and -0.0011 with a variance and standard deviation of 0. The skewness factor is equal to -0.1317 so the data is a little bit negative skew. The kurtosis factor is equal to -0.8271 so the tails of the distribution are heavier, which can be seen in the difference between the green and red tails in the figure.

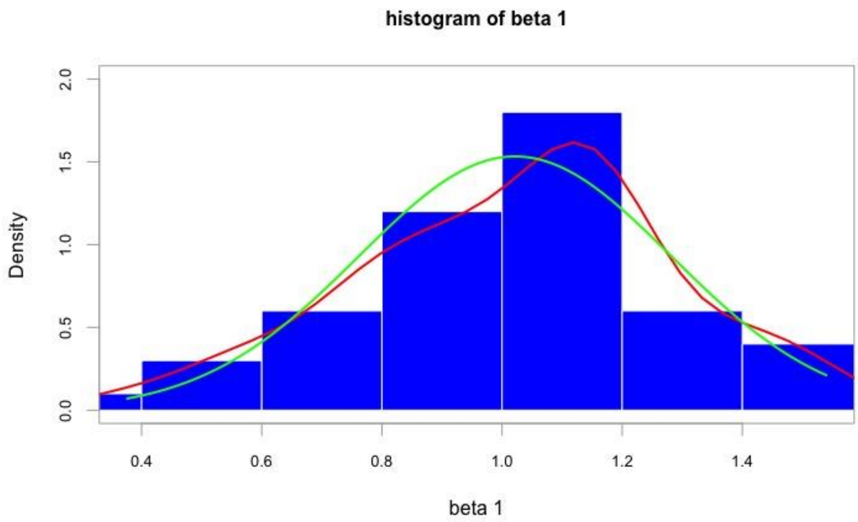

figure 3: histogram of β_1

The histogram of the 50 computed β_1 coefficients is together with the computed and normal density function shown in figure 3. The median is equal to 1.0761, the mean is equal to 1.0218, the maximum and minimum value is respectively 1.5397 and 0.3761 with a variance of 0.0678 and standard deviation of 0.2603. The skewness factor is equal to -0.2647 so the data is a little bit negative skew which is shown in the figure by the longer left tail. The kurtosis factor is equal to -0.2311 so the tails of the distribution are a little heavier, which can be seen in the difference between the green and red tails in the figure.

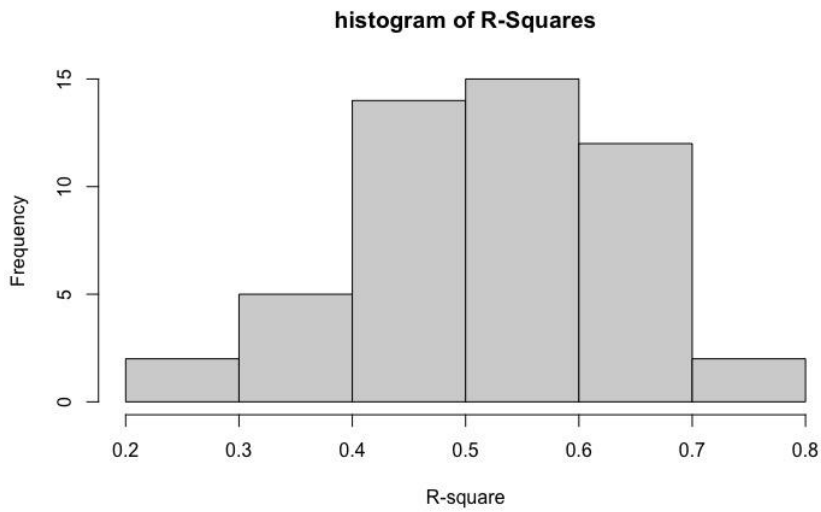

figure 4: histogram of the R-Squares

The histogram of the 50 computed R^2coefficients is shown in figure 4. The median is equal to 0.5327, the mean is equal to 0.5276, the maximum and minimum value is respectively 0.7655 and 0.2112 with a variance of 0.0141 and standard deviation of 0.1186.

4. Weekly and monthly returns

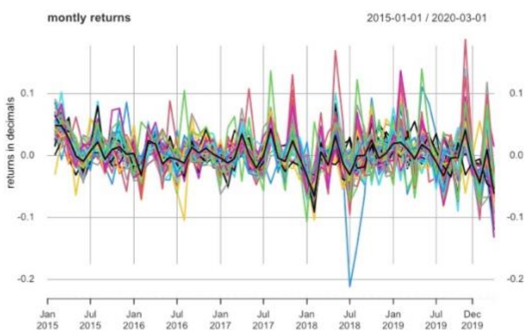

figure 5: monthly returns in decimals

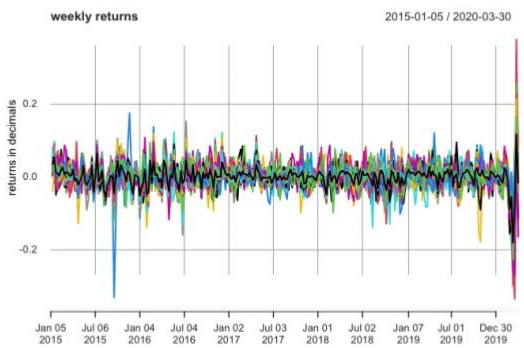

figure 6: weekly returns in decimals

It is clear that the larger the intervals are, the more clear are the peaks in the variance. The differences in variance are more visible in the monthly returns.

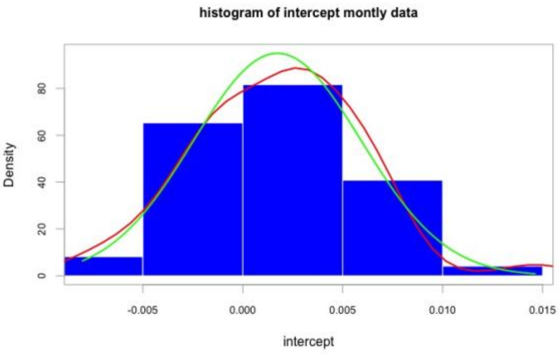

figure 7: histogram of the intercept for monthly data

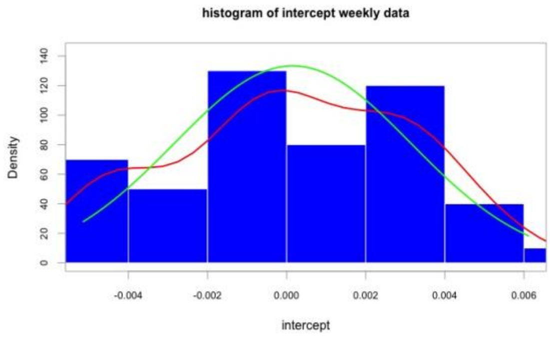

figure 8: histogram of the intercept for weekly data

The mean of the weekly data is equal to 0.0001 and 0.0017 for the monthly data. The variance and standard deviation is 0 and 0.003 for the weekly data and 0 and 0.0042 for the monthly data. The skewness factor is equal to -0.1329 for the weekly data and 0.1811 for the monthly data, so the data is a little bit negative skew for the weekly data and positive skew for the monthly data. The kurtosis factor is equal to -0.8909 for the weekly data and 0.6232 for the monthly data. The larger the intervals are, the larger is the variance and the less heavy are the tails of the distribution.

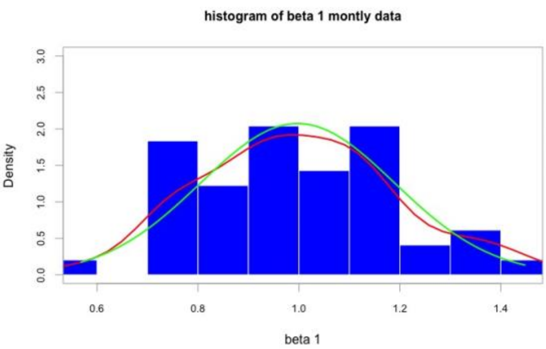

figure 9: histogram of β_1 for monthly data

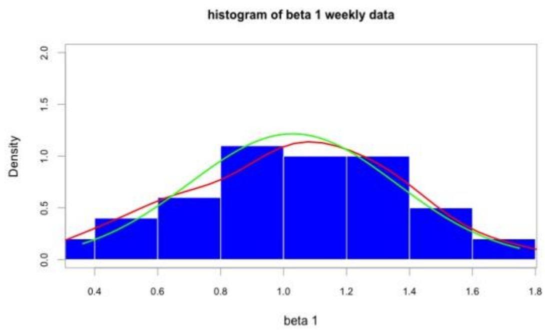

figure 10: histogram of β_1 for monthly data

The mean of the weekly data is equal to 1.0019 and 0.9884 for the monthly data. The variance and standard deviation is 0.1087 and 0.3283 for the weekly data and 0.0370 and 0.1923 for the monthly data. The skewness factor is equal to -0.1244 for the weekly data and 0.2047 for the monthly data, so the data is a little bit negative skew for the weekly data and positive skew for the monthly data. The kurtosis factor is equal to -0.5051 for the weekly data and -0.2894for the monthly data. The larger the intervals are, the larger is the variance and the less heavy are the tails of the distribution.

figure 11: histogram of the R-Squares for monthly data

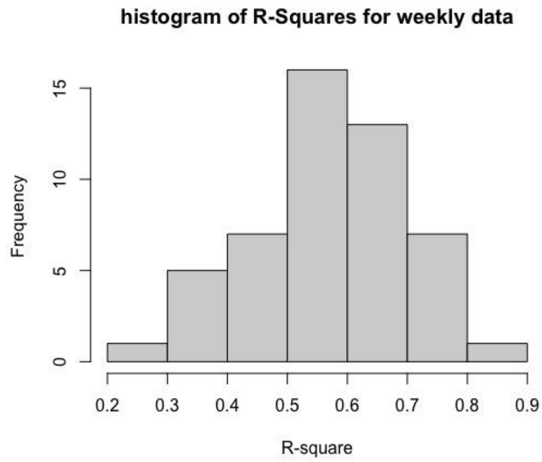

figure 12: histogram of the R-Squares for monthly data

The median is equal to 0.5702 for the weekly data and 0.5489 for the monthly data, the mean is equal to 0.5588 and 0.5135 for the monthly data. The variance is 0.0168 for the weekly data and 0.0230 for the monthly data, and the standard deviation is 0.1294 for the weekly data and 0.1515 for the monthly data. The larger the intervals, the higher the variance and standard deviation of the R-Squared. There is no clear conclusion for the value of the R-Squared, the factor is better for weekly data compared to monthly however the value for daily data is in the middle.

Part 2: Blume and Merrill-Lynch beta coefficients

The first approach included the use of historical betas from period 2016-2017. These betas were copied to use as betas for period 2018-2020. The idea behind this is that the past is the best predictor the future. The second approach included the Blume approach. A set of betas was computed based on the coefficients of period 2014-2015, trained on the data of period 2016-2017. By this method, coefficients were computed for a period and adjusted by the training on independent data of another period. The third method included the Merrill-Lynch approach. The coefficients of period 2014-2015 were used and adjusted by fixed cross-sectional coefficients.

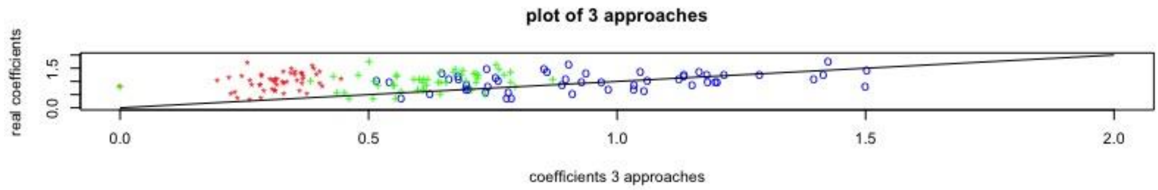

figure 13: plot of the 3 approaches

The blue scatterplot is the first approach, the red scatterplot is the second approach and the green scatterplot is the third approach. The advised method is the first one. The blue scatterplot is the best fit to the 45° line. An explanation for this is that the computed coefficients fit the real values of the computed coefficients for period 2018-2020 very well. These coefficients are the best estimators to make predictions for the returns in period 2018-2020. It is true that the information from the nearest past is the best to make predictions about the future.