**Task 1.**

Data for the task were taken from the table of companies *fhn, fosl, gci.*

In the first task, 3 companys were selected and calculated:

1.Logarithmic rate of the sample with a lag of 1 – variable lr( index) in each file

2.The logarithm of the ratio mx\mn is the variable mr( index) in each file

3.Logarithm of the opn\clo ratio - variable or( index) in each file

Example code from one of the files:

setwd("C:\\Users\\oboga\\Documents\\hw2\\files for ex 1") (directory choice)

x<-read.csv("table\_fhn.csv",header=F) (file reading)

colnames(x) <- c("dat","z","opn","mx","mn","clo","vol") (changing names of the column in the table)

edit(x) (editing the table)

lr <-log(x$clo[-length(x$clo)]/x$clo[-1]) (calculating variables)

mr <- log(x$mx/x$mn)

or<-log(x$opn/x$clo)

Received three vectors for each of the companys, the length of each of them is equal to the quantity of lines in each file.

**Tasks 2 і 3.**

*Data for the 2 task were taken from the company table ftr.*

To build a box with a mustache, we need to know the median, the first and third quartiles, as well as the maximum and minimum sample values. It is necessary to determine what values we will interpret as emissions. To do this, calculate the interquartile scope and set this value by multiplying it by 3\2. Then everything that will get and this interval we will not consider emissions.

medlr<- median(lr) (median calculation)

quantile(lr) (formula for calculating quantiles and assigning variable values)

fqlr <- -0.009938708

tqlt <- 0.009828089

IQRlr <- IQR(lr)\*3/2 (calculating the scope)

There is no large semantic difference between the hand-made figure ( Appendix A) and the figure made with the program, but for a clearer reflection to the hundredths, without number simplifications and mechanical errors, the use of the program is necessary. In addition, in the figure made by hand, emissions are marked only in the places of their largest cluster, while in the figure, the executed in the program can clearly see the minimum and maximum value, « density» emissions and so on.

*Data for the 3 task were taken from the table of the company ftr, gd, fisv.*

To build a box with a mustache, the code from task 1 was used to calculate the variables. The code to build the box:

boxplot(lr,mr,or) ( function for building a box with a mustache)

the resulting boxes can be seen in Appendix B, Appendix C and Appendix D. Comparing the figures in these applications, we assume that they have mpilic features: boxes for lr and or are symmetrical, and for mr – asymmetric.

**Task 4.**

Variables lr mr or were calculated for all companies of the option. As a result, 6 figures were obtained - 2 for each variable: with and without emissions.

Example code from a file for a picture of the variable lr:

boxplot(lr1, lr2, lr3, lr4, lr5, lr6, lr7, lr8, lr9, lr10, lr11, lr12, lr13, lr14, lr15, lr16, lr17, lr18, lr19, lr20) (code for box with emissions)

boxplot(lr1, lr2, lr3, lr4, lr5, lr6, lr7, lr8, lr9, lr10, lr11, lr12, lr13, lr14, lr15, lr16, lr17, lr18, lr19, lr20, outline = F, names = c("lr1", "lr2", "lr3", "lr4", "lr5", "lr6", "lr7", "lr8", "lr9", "lr10", "lr11", "lr12","lr13", "lr14", "lr15", "lr16", "lr17", "lr18","lr19", "lr20")) (code for box without emissions)

The common feature for mustache boxes in Appendix E is their median, it is 0 everywhere, the distribution is symmetrical everywhere. Compared to a box for company 11, the size of others does not fluctuate much. The most plausible reason for this is that in the table for company 11 the values were twice less than for all others, so the built box does not reflect reality. The same can be noted for mustache boxes in Appendix E. For Appendix F, the distribution for the mustache box is asymmetric, the median is closer down, and the tail upwards is stretched more than down. In addition, looking at Figure 2 of Appendix F, you can see that there are no lower emissions.

Analyzing each of the boxes, you can get enough valuable information about the scatter of values to determine the choice of company whose shares are worth buying.

**Task 5.**

*Data for the performance of 5 tasks were taken from the table of the firm fslr, fitb, gis .*

According to the results of task 4, after comparing the variables and their boxes for 20 firms, three firms and variables were selected, namely: the variable lr of the firm fslr (Appendix H), variable or firm fitb (Appendix K) and variable mr of firm gis (Appendix L). Based on the data obtained, histograms of variables were constructed.

Example code:

hist(lr11) (function for constructing histogram)

Examining the values in Appendix L can conclude that the distribution is unimodal (asymmetric ) with a heavy right tail, ie asymmetry is additional.

The distribution in Appendix K and Appendix H is quite symmetrical, but as we shed data from different firms, the frequency of data and the distribution of variables are different.

For example, for Appendix H, the most common logarithmic rate of return was between -0.05 and 0.05, at each of the intervals the value is almost 700.

**Завдання 6.**

*Data for the 6 task were taken from the company table fslr.*

As a result of task 4, the company fslr was selected. A histogram was built for it, on which the values of the three variables are compared: logarithmic rate of the sample with lag 1, logarithm of the ratio mx\mn, logarithm of the ratio opn\clo. (Appendix M)

brks<-seq(min(lr,mr, or)-0.01,max(lr,mr, or)+0.01,length.out=50) ( breakdown of lilage>

hist(lr,breaks=brks,probability=T,

main = paste("Historgam based on lr, mr, or in fslr "),

col=rgb(0,0,1,1/4),ylim=c(0,30)) ( histogram creation)

hist(mr,breaks=brks,probability=T,

col=rgb(1,0,0,1/4), add=T) (adding variables)

hist(or,breaks=brks,probability=T,

col=rgb(1/4,0,1/4,1/4), add=T)

legend(x=0.2,y=30,c("lr "," mr "," "or"),

fill=c(rgb(0,0,1,1/4),rgb(1,0,0,1/4),

rgb(1/4,0,1/4,1/4))) ( legend)

Analyzing Appendix M, one can make a conclusion about the treatment of shares of the selected company, for example: the histogram shows that the most common profit was when buying shares at the opening of the exchange and selling at the average price was more, than when buying a stock at opening and selling during the day. In addition, although the distribution of histograms is relatively symmetrical, the figure shows that for the nominal lr of the two most common percentages, the higher frequency has the one that is negative, and for the variable or – vice versa. In addition, the histogram mr shows that the profit when buying at the minimum and selling at the maximum price is very rarely zero and there is a high probability of earning a minimum percentage.

**Task 7.**

*Data for the 5 task were taken from the company table* *fhn, fls та glw.*

Three companies were selected to compare histograms of the variable mr and a histogram ( Appendix N) was built. Just as we compare the maximum profit as a percentage for 1 day, comparing the minimum and maximum prices of selected companies, we can draw some conclusions. For example, if you choose the stock on which there is the greatest probability of earning, it would be the company glw, because there are less days when there was zero profit, in addition, the trait that distinguishes these shares from others is that the histogram falls more smoothly with an increase in interest, ie there is a greater probability of earning a minimum percentage. If you use this criterion, then the next is fhn, and the last – fls.

Conclusion: Histograms and mustache boxes are built using the same information and carry a similar semantic load, but each of these data images is convenient to use for different purposes. When working with boxes, you can get more information about data distribution. When working with histograms, it is easier to make assumptions and look for a connection between the data.

Appendix

AppendixА

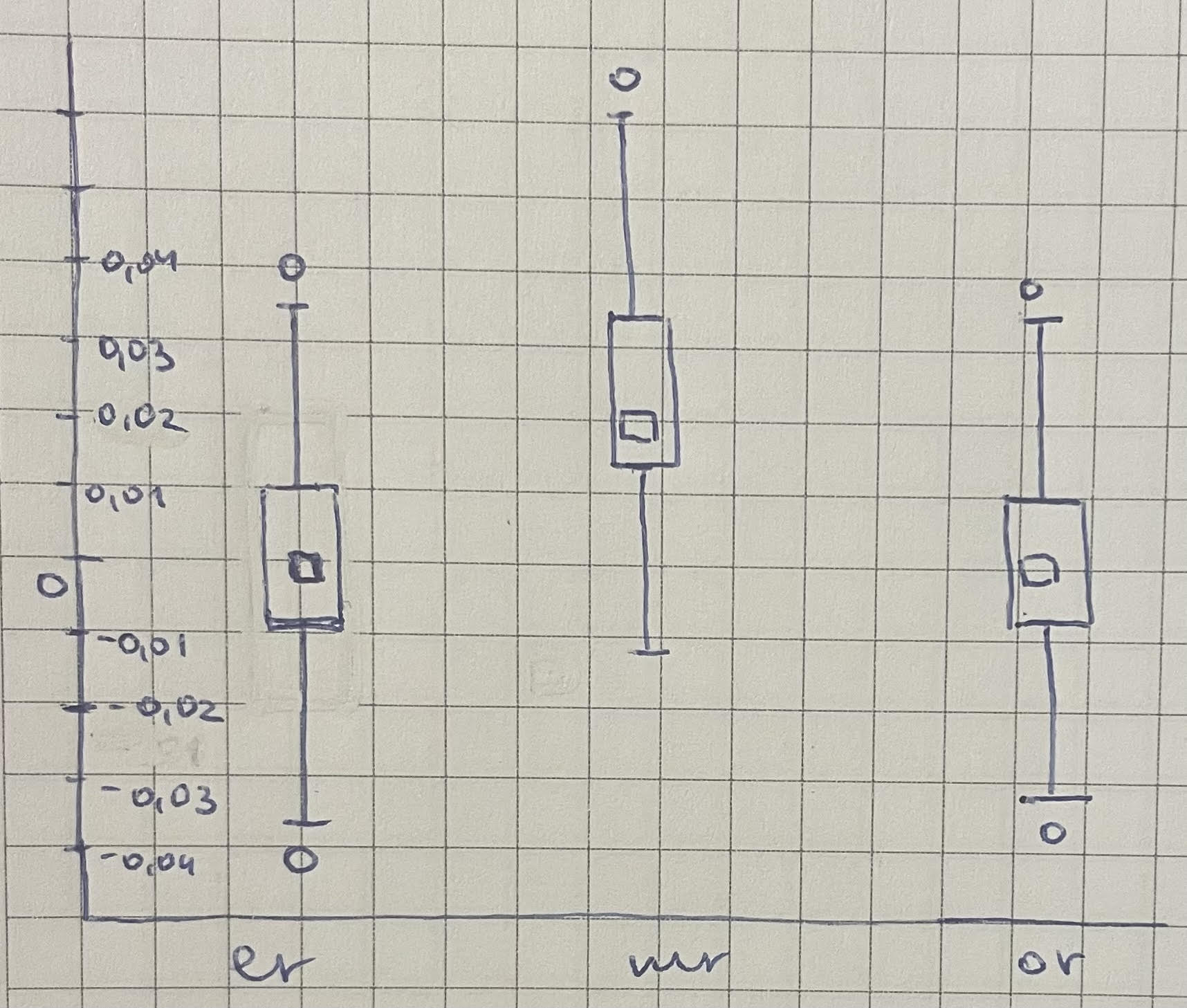


Figure1. Box with mustache for ftr, built by hand

Appendix B

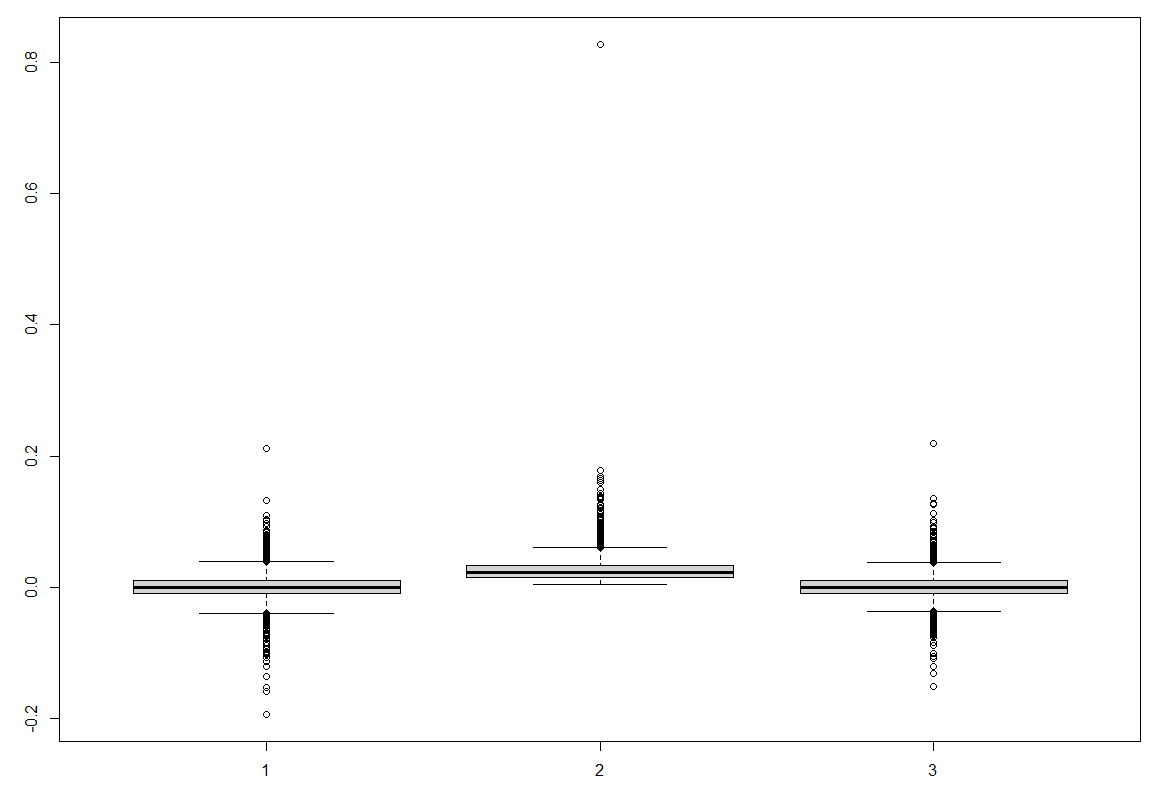


Figure 1 . box with mustache for the company ftr, built with R

Appendix C

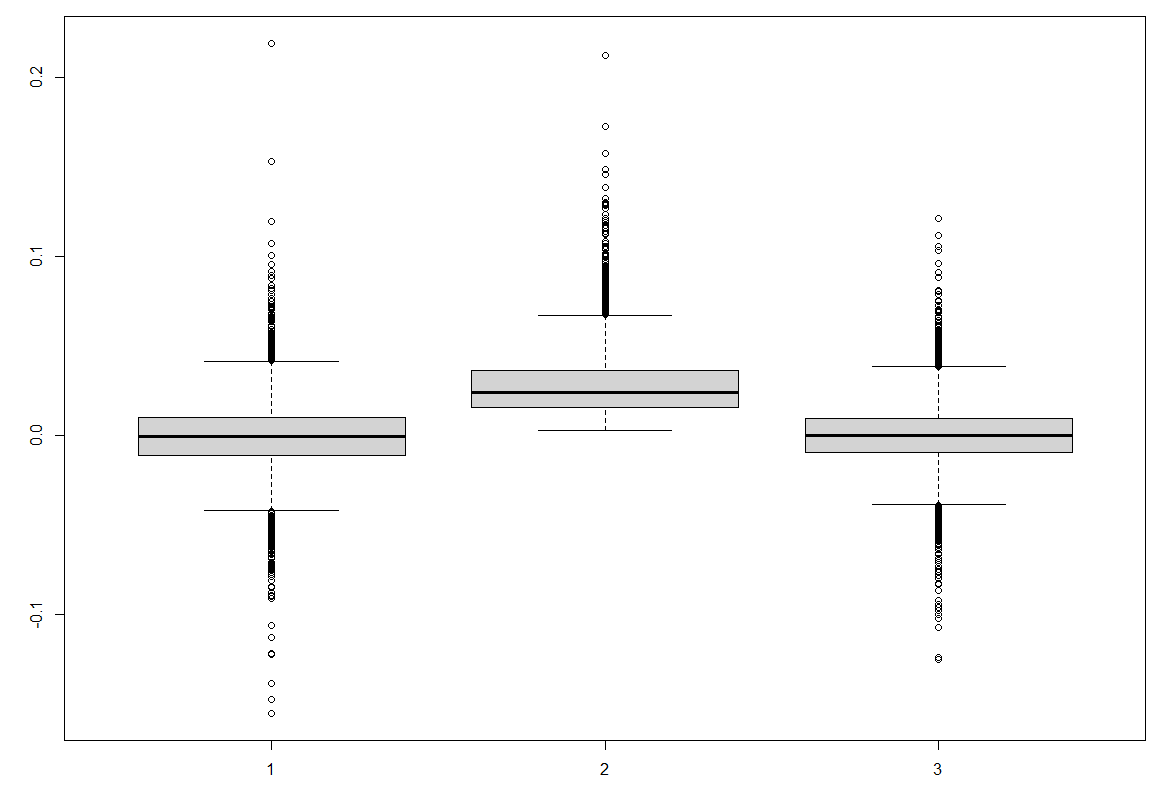


Figure1. Box with mustache for fisv

Appendix D

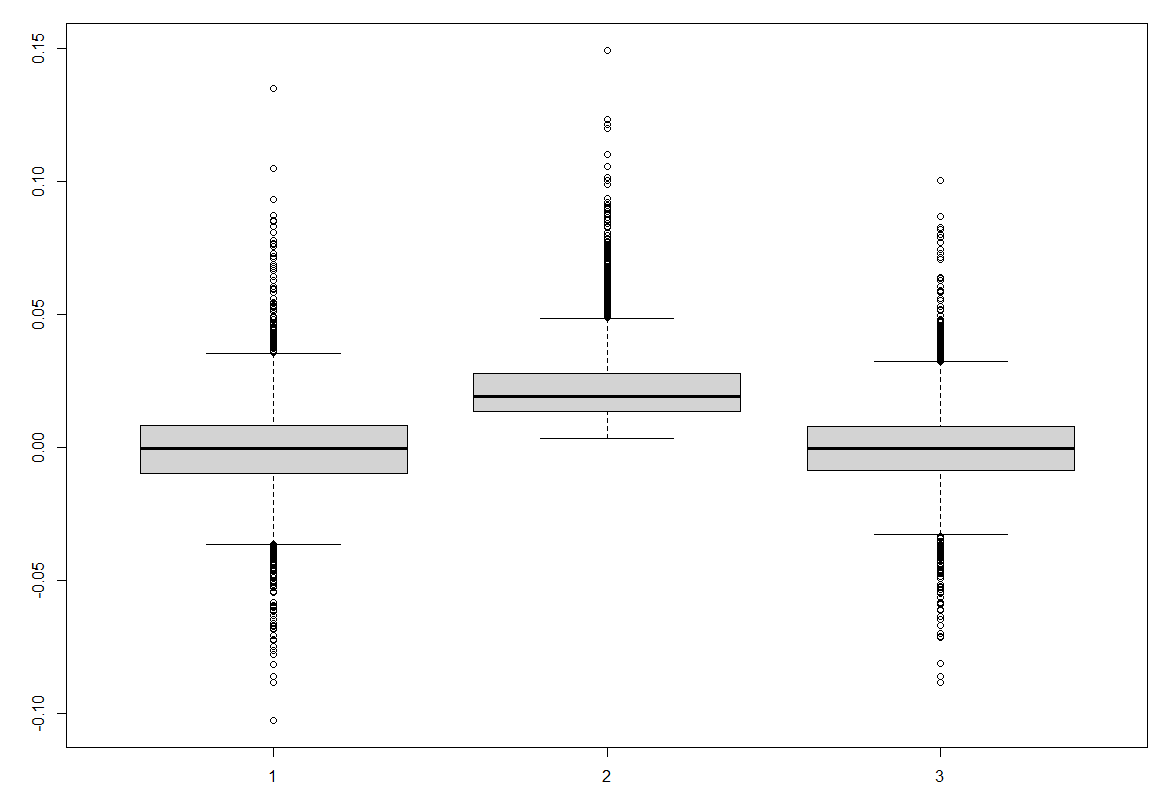


Figure 1 . Box with mustache for gd

Appendix E

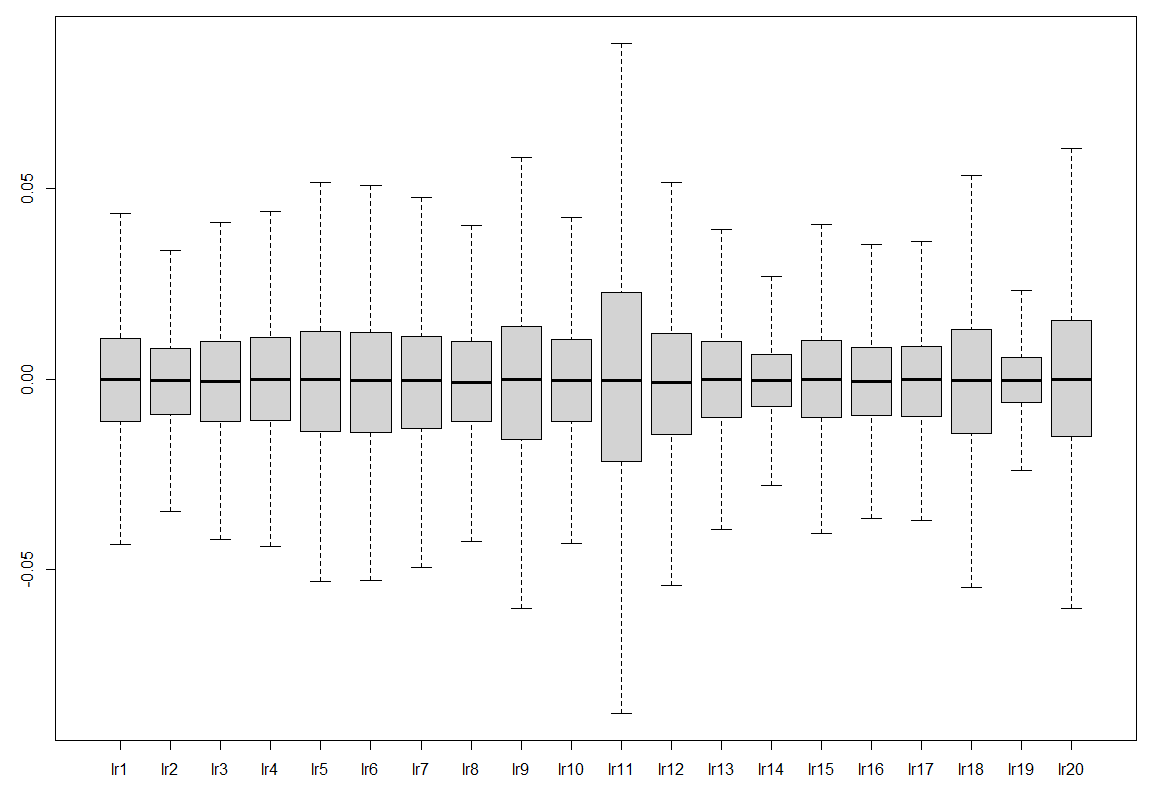
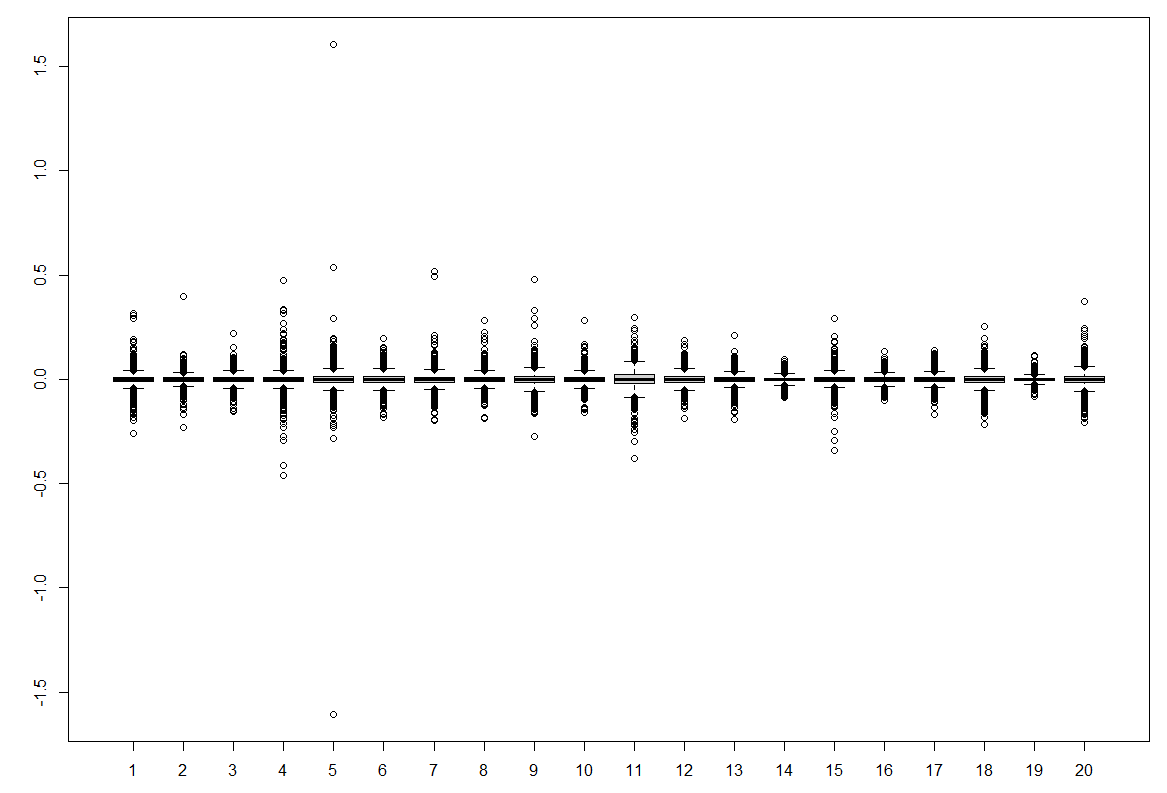
 

Figure1. Box with mustache of logarithmic norm of a stop with a lag of 1 for all 20 companies without emissions

Figure2. Box with mustache of logarithmic norm of a sample with a lag of 1 for all 20 companies with emissions

Appendix F

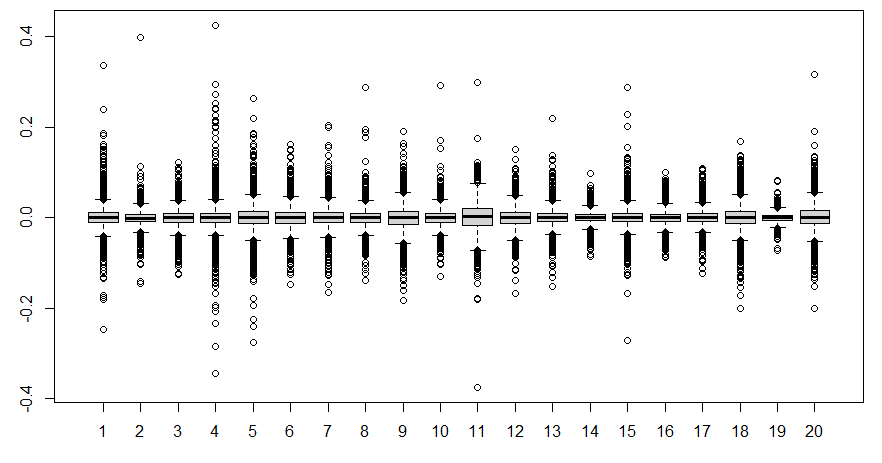
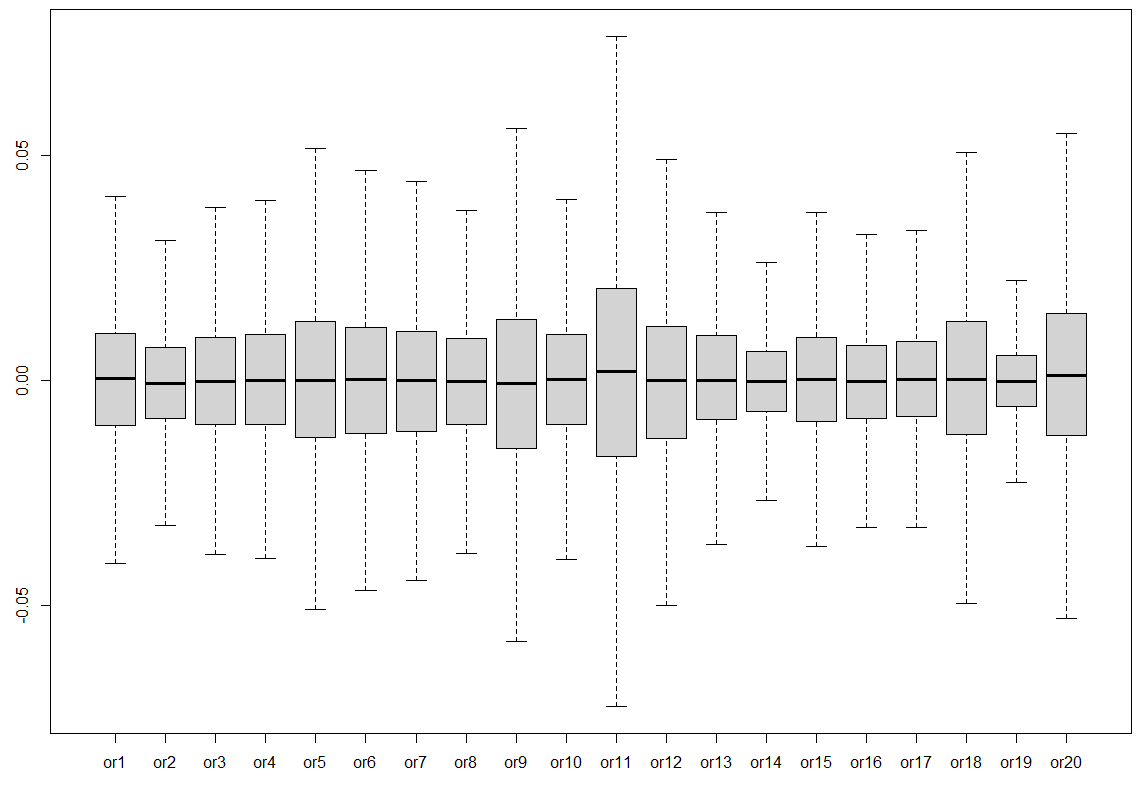


Figure1. Box with mustache logarithm of the opn\clo ratio for all 20 firms without emissions

Figure2. Box with mustache logarithm of the opn\clo ratio for all 20 companies with emissions

Appendix G

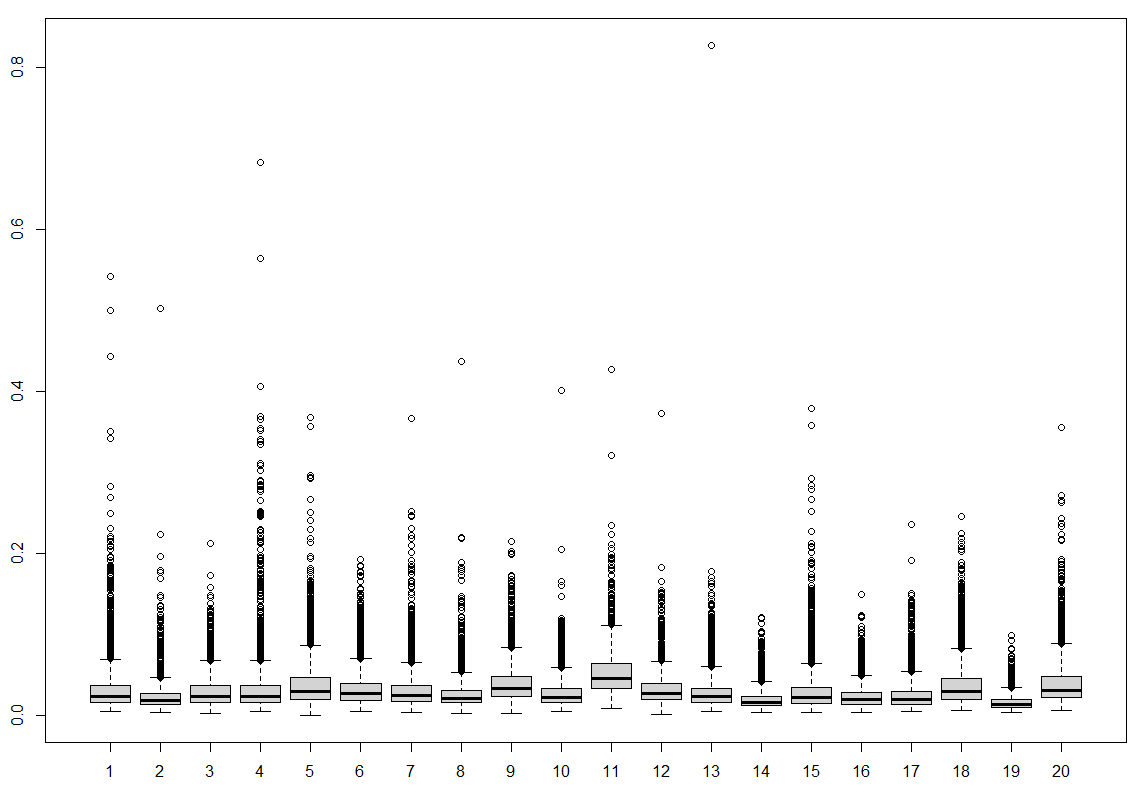
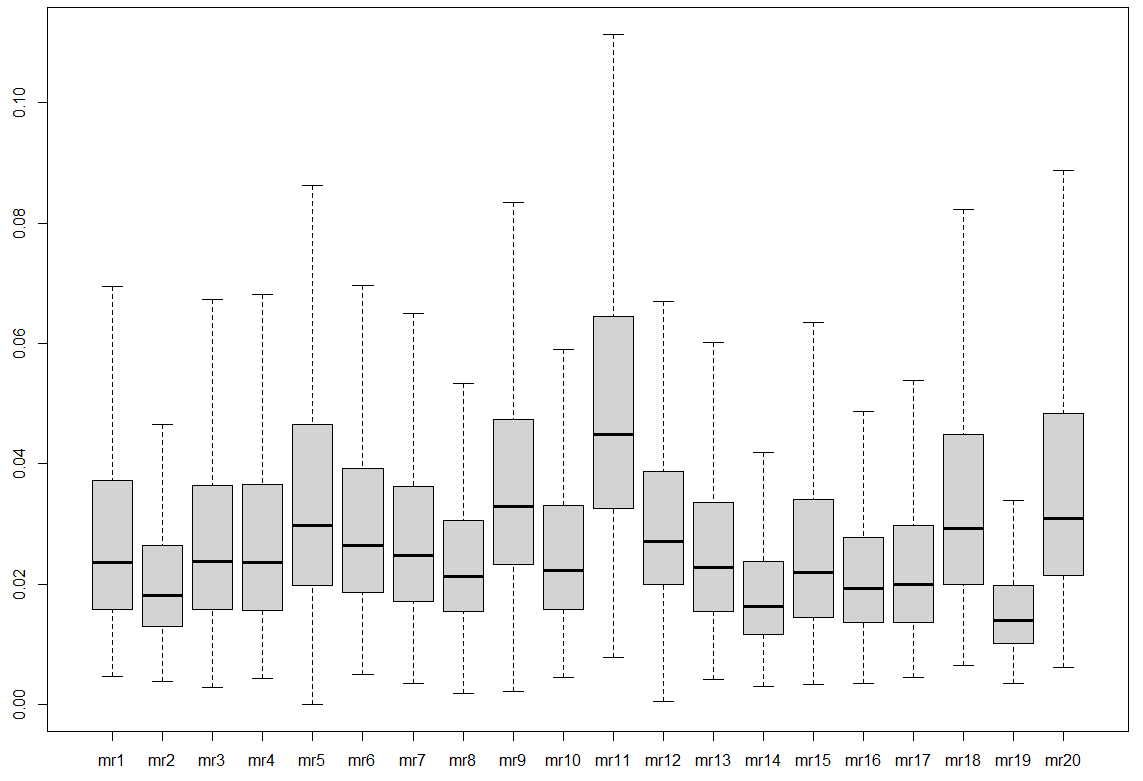


Figure1. Box with mustache logarithm ratio mx\mn for all 20 firms without emissions

Figure2. Box with mustache logarithm of the ratio mx\mn for all 20 companies with emissions

Appendix H

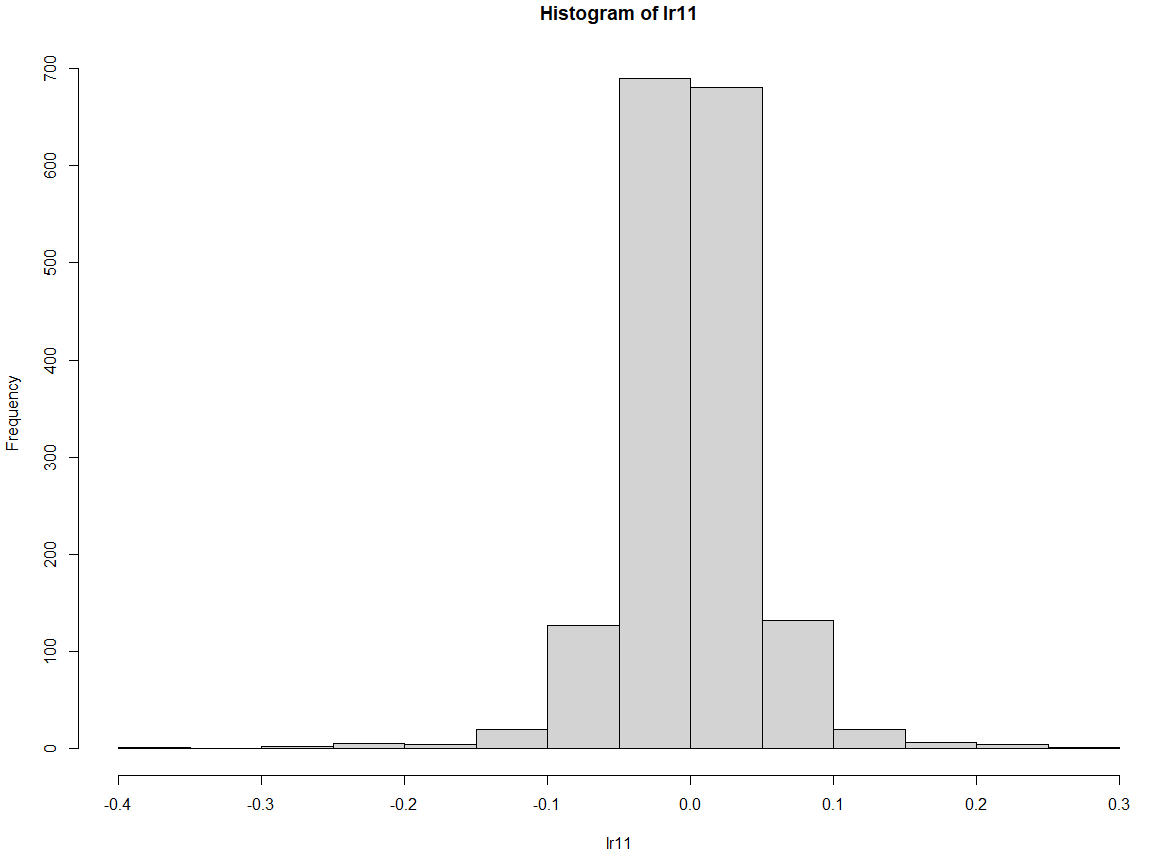


Figure1. Histogram of logarithmic rate of the sample with lag 1 for the firm fslr

Appendix K

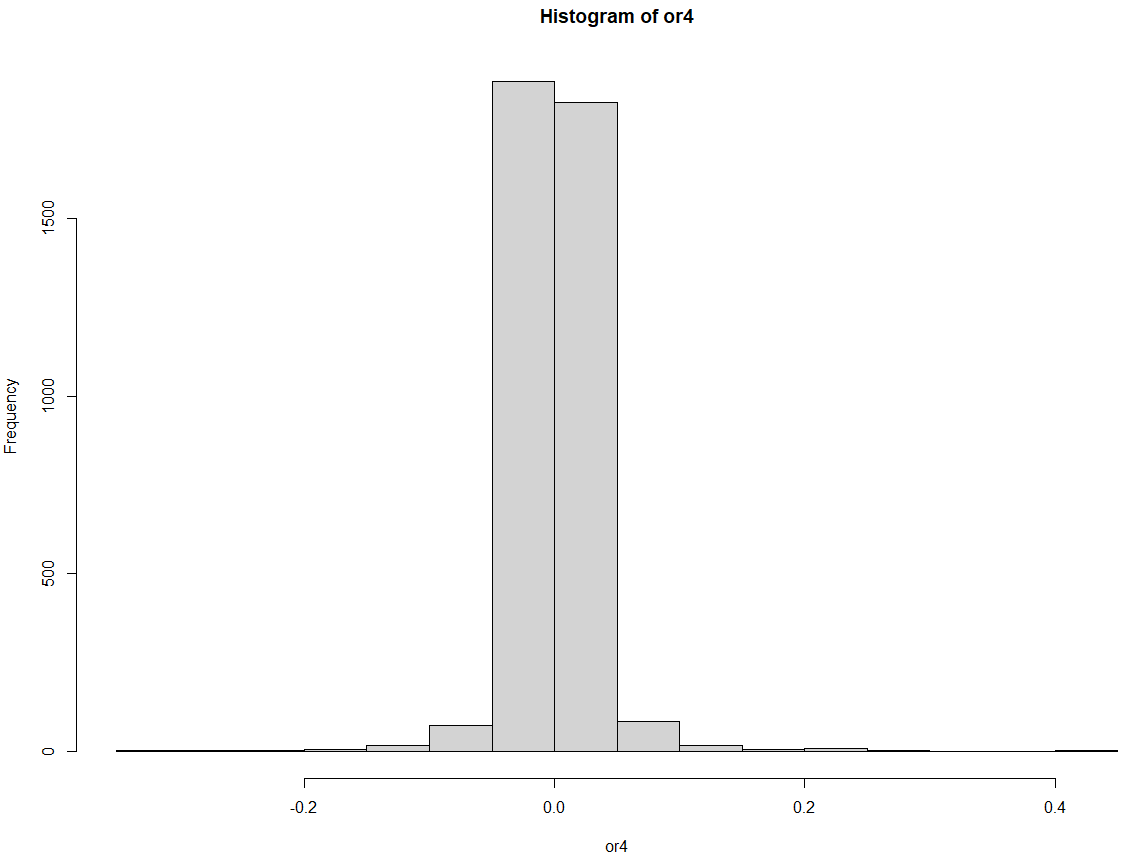


Figure1. histogram of the opn\clo ratio logarithm for fitb

Appendix L

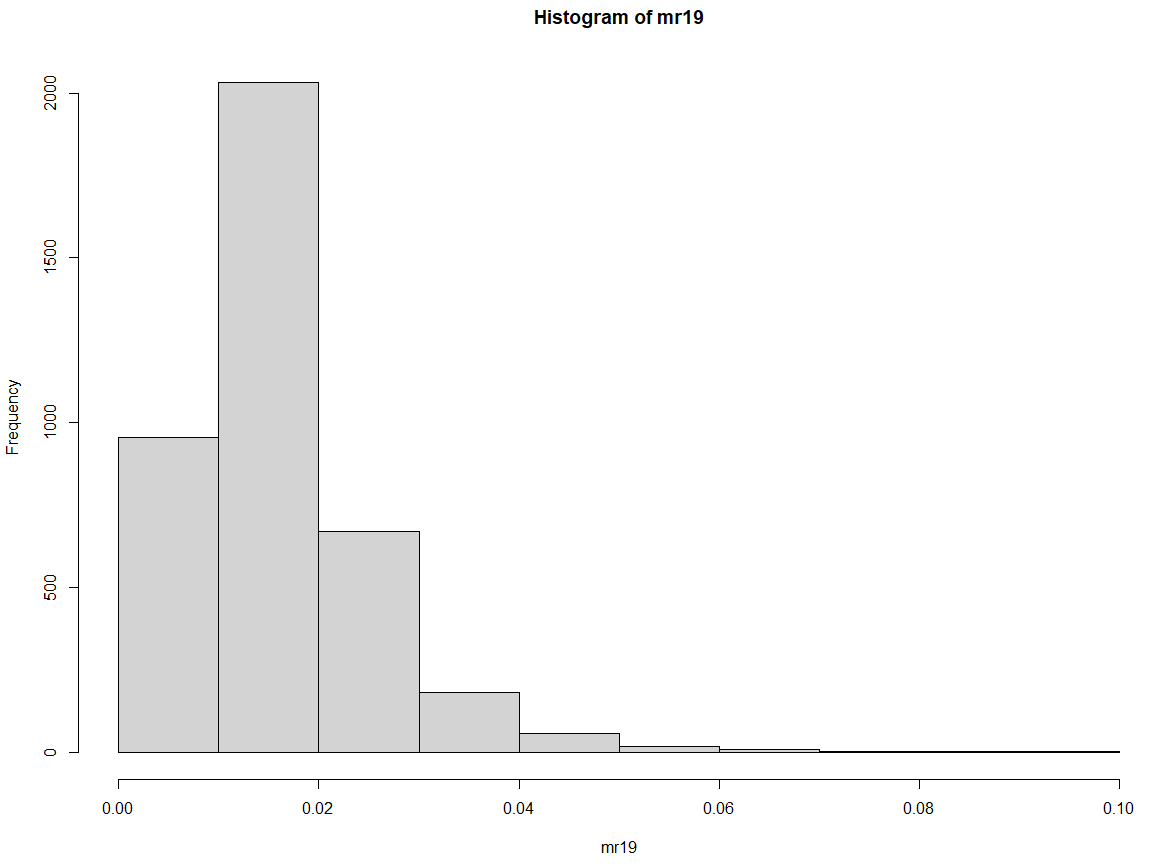


Figure1. histogram logarithm of the mx\mn ratio for the gis firm

Appendix M

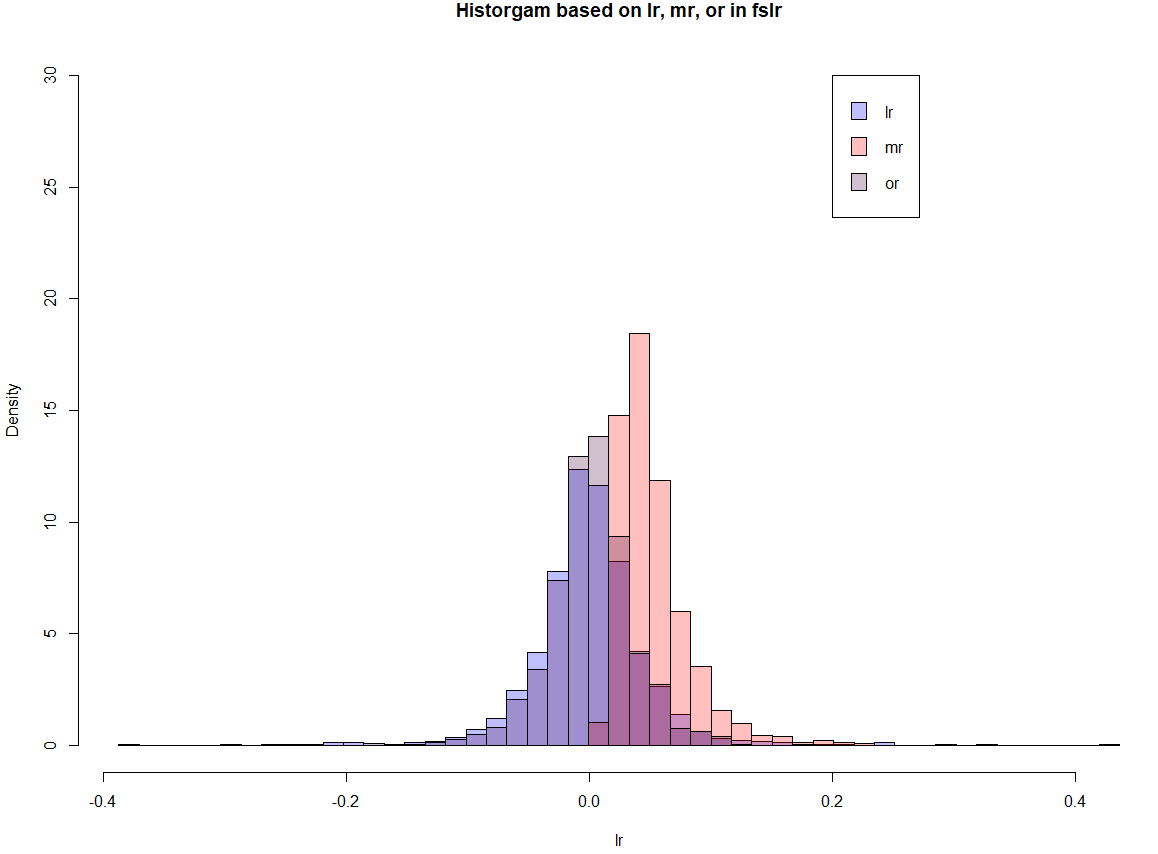


Figure1. Histogram for variables: logarithmic rate of the sample with a lag of 1 (lr), logarithm of the ratio mx\mn (mr) and logarithm of the ratio opn\clo (or) for the firm flsr

Appendix N

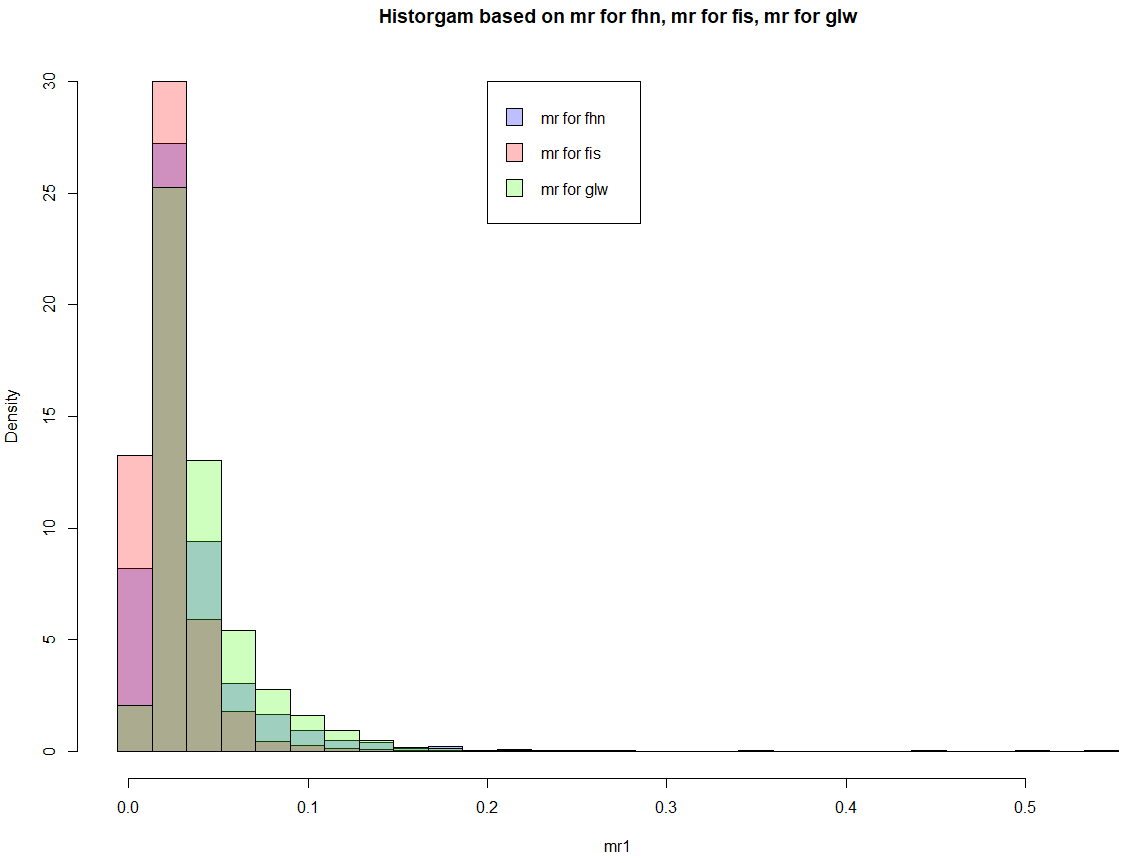


Figure1. Histogram for logarithm of the ratio mx\mn (mr) for fhn, fls and glw