Baixando e analisando dados de alta frequência

Lucca Simeoni Pavan *João Carlos de Carvalho*

8 de dezembro de 2016

Sumário


```
knitr::opts_chunk$set(echo = TRUE, cache = FALSE, fig.height = 4, warning = FALSE,
    message = FALSE, error = FALSE, tidy = TRUE, tidy.opts = list(width.cutoff = 70))
```

1 Ranking de negociações

```
library(GetHFData)
tickers_equity <- ghfd_get_available_tickers_from_ftp(my.date = "2016-10-30",
    type.market = "equity", max.dl.tries = 10)
```

```
##
## Reading ftp contents for equity (attempt = 1|10) - Error in reading ftp contents. Trying again..
## Reading ftp contents for equity (attempt = 2|10) Attempt 1 - File exists, skipping dl
```

```
head(tickers_equity, n = 10)
```

```
##    tickers n.trades                      f.name
## 1    PETR4    52393 ftp files/NEG_20161117.zip
## 2    JBSS3    45174 ftp files/NEG_20161117.zip
## 3    ITSA4    39200 ftp files/NEG_20161117.zip
## 4    ITUB4    30529 ftp files/NEG_20161117.zip
## 5    VALE5    30423 ftp files/NEG_20161117.zip
## 6    BVMF3    29099 ftp files/NEG_20161117.zip
## 7    BBDC4    26923 ftp files/NEG_20161117.zip
## 8    ABEV3    26786 ftp files/NEG_20161117.zip
## 9    BBAS3    26672 ftp files/NEG_20161117.zip
## 10   RUMO3    26274 ftp files/NEG_20161117.zip
```

Criando um vetor com as 6 ações mais negociadas em 30/10/2016.

```
top_6 <- c(as.character(head(tickers_equity$tickers)))
print(top_6)
```

```
## [1] "PETR4" "JBSS3" "ITSA4" "ITUB4" "VALE5" "BVMF3"
```

Baixando os dados

```
dados_top6 <- ghfd_get_HF_data(top_6, type.market = "equity", first.date = as.Date("2014-11-03"),
    last.date = as.Date("2016-10-30"), first.time = "9:00:00", last.time = "18:00:00",
    type.output = "agg", agg.diff = "1 hour", dl.dir = "ftp files", max.dl.tries = 10,
    clean.files = FALSE)
```

```r
save(dados_top6, file = "dados_top6.Rda")
head(dados_top6, n = 6)
```

```r
load("dados_top6.Rda")
dim(dados_top6)
```

```
## [1] 22667    13
```

```r
str(dados_top6)
```

```
## 'data.frame':    22667 obs. of  13 variables:
##  $ InstrumentSymbol: chr  "ABEV3" "ABEV3" "ABEV3" "ABEV3" ...
##  $ SessionDate     : Date, format: "2014-11-03" "2014-11-03" ...
##  $ TradeDateTime   : POSIXct, format: "2014-11-03 10:00:00" "2014-11-03 11:00:00" ...
##  $ n.trades        : int  1607 2055 3417 3686 3978 4707 5168 250 1602 1203 ...
##  $ last.price      : num  16.1 16.1 16.2 16.1 16.1 ...
##  $ weighted.price  : num  16.1 16.1 16.2 16.2 16.1 ...
##  $ period.ret      : num  -0.00864 0.00124 0.0056 -0.00124 -0.00372 ...
##  $ period.ret.volat: num  0.000325 0.000324 0.000278 0.000235 0.000263 ...
##  $ sum.qtd         : num  824900 926700 1408500 1034900 1141100 ...
##  $ sum.vol         : num  13291157 14907444 22757436 16729199 18362060 ...
##  $ n.buys          : int  579 1113 1888 2265 1972 1878 2309 23 659 526 ...
##  $ n.sells         : int  1028 942 1529 1421 2006 2829 2859 227 943 677 ...
##  $ Tradetime       : chr  "10:00:00" "11:00:00" "12:00:00" "13:00:00" ...
```

Agora irei criar um banco de dados para cada ação e depois obter os log retornos.

```r
library(dplyr)
dados_ITSA4 <- filter(dados_top6, InstrumentSymbol == "ITSA4") %>%
    select(SessionDate, weighted.price) %>% mutate(log_retorno = log(weighted.price) -
    lag(log(weighted.price)))
dados_PETR4 <- filter(dados_top6, InstrumentSymbol == "PETR4") %>%
    select(SessionDate, weighted.price) %>% mutate(log_retorno = log(weighted.price) -
    lag(log(weighted.price)))
dados_ITUB4 <- filter(dados_top6, InstrumentSymbol == "ITUB4") %>%
    select(SessionDate, weighted.price) %>% mutate(log_retorno = log(weighted.price) -
    lag(log(weighted.price)))
dados_BBDC4 <- filter(dados_top6, InstrumentSymbol == "BBDC4") %>%
    select(SessionDate, weighted.price) %>% mutate(log_retorno = log(weighted.price) -
    lag(log(weighted.price)))
dados_ABEV3 <- filter(dados_top6, InstrumentSymbol == "ABEV3") %>%
    select(SessionDate, weighted.price) %>% mutate(log_retorno = log(weighted.price) -
    lag(log(weighted.price))) %>% mutate(log_retorno = log(weighted.price) -
    lag(log(weighted.price)))
dados_BBSE3 <- filter(dados_top6, InstrumentSymbol == "BBSE3") %>%
    select(SessionDate, weighted.price) %>% mutate(log_retorno = log(weighted.price) -
    lag(log(weighted.price)))
```

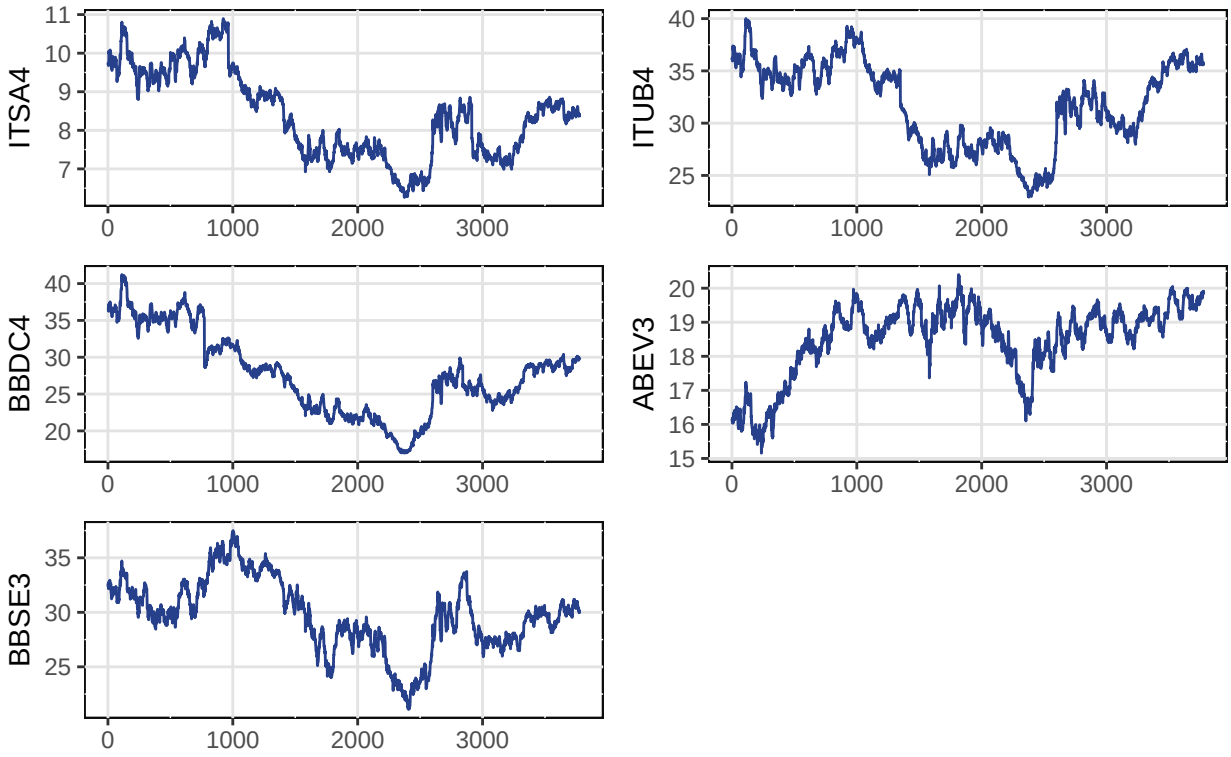
Removendo NAs.

```r
dados_BBSE3 <- dados_BBSE3[2:3778, ]
dados_ABEV3 <- dados_ABEV3[2:3778, ]
dados_BBDC4 <- dados_BBDC4[2:3778, ]
dados_ITUB4 <- dados_ITUB4[2:3778, ]
dados_PETR4 <- dados_PETR4[2:3777, ]
dados_ITSA4 <- dados_ITSA4[2:3778, ]
```

2 Descrição dos Dados

Representação dos preços.

```
matriz_preco <- data.frame(ITSA4 = (dados_ITSA4$weighted.price), ITUB4 = (dados_ITUB4$weighted.price),
    BBDC4 = (dados_BBDC4$weighted.price), ABEV3 = (dados_ABEV3$weighted.price),
    BBSE3 = (dados_BBSE3$weighted.price))
library(BMR)
gtsplot(matriz_preco)
```




```
summary(matriz_preco)
```

```
##      ITSA4            ITUB4           BBDC4           ABEV3
##  Min.   : 6.268   Min.   :22.93   Min.   :16.98   Min.   :15.16
##  1st Qu.: 7.429   1st Qu.:28.40   1st Qu.:23.33   1st Qu.:18.15
##  Median : 8.353   Median :32.66   Median :27.61   Median :18.79
##  Mean   : 8.390   Mean   :31.81   Mean   :27.62   Mean   :18.52
##  3rd Qu.: 9.339   3rd Qu.:35.25   3rd Qu.:31.15   3rd Qu.:19.26
##  Max.   :10.897   Max.   :40.00   Max.   :41.22   Max.   :20.40
##      BBSE3
##  Min.   :21.12
##  1st Qu.:27.69
##  Median :29.80
##  Mean   :29.71
##  3rd Qu.:31.99
##  Max.   :37.50
```

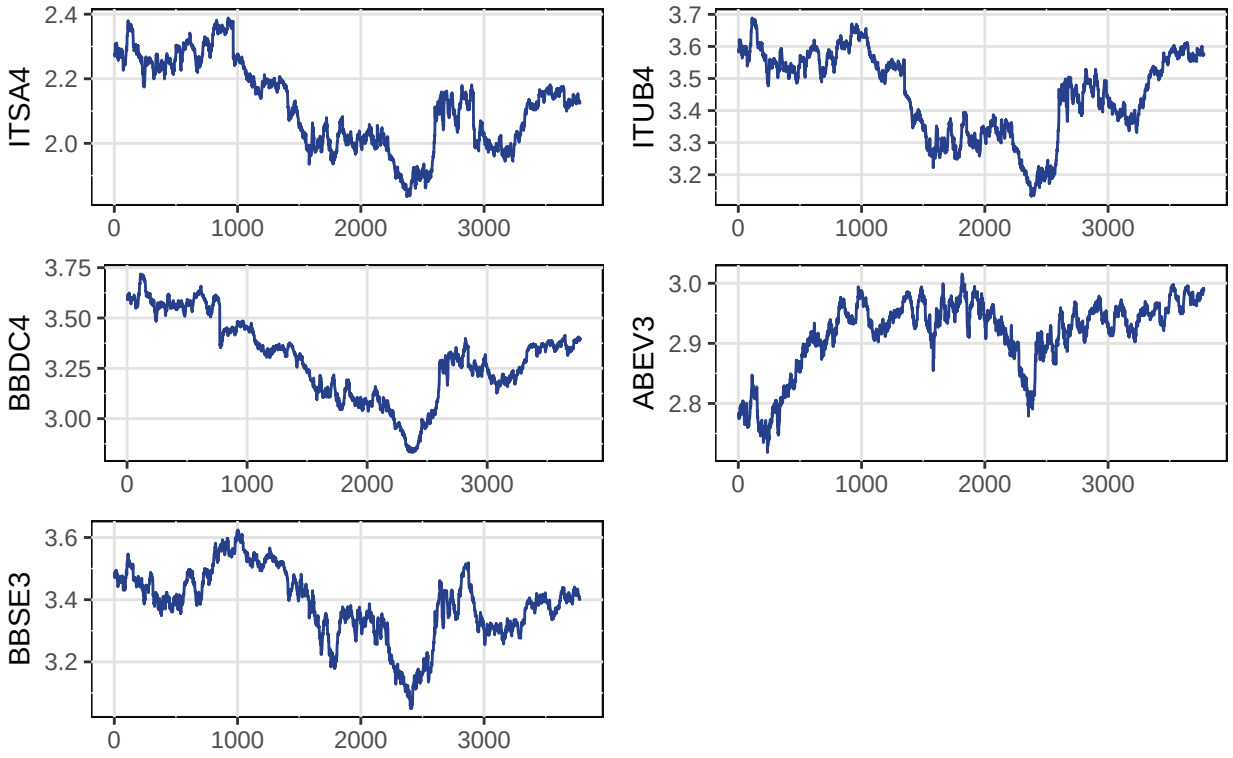
```
cov(matriz_preco)
```

```
##             ITSA4      ITUB4     BBDC4       ABEV3      BBSE3
## ITSA4  1.2634470  4.1766452  5.701124 -0.2390199  3.0699712
```

```
## ITUB4  4.1766452 16.6835114 19.773773 -0.1993035 10.9002367
## BBDC4  5.7011238 19.7737733 30.416024 -2.0025240 12.7153579
## ABEV3 -0.2390199 -0.1993035 -2.002524  1.1330387  0.3676802
## BBSE3  3.0699712 10.9002367 12.715358  0.3676802 10.7679356
```

Criando matriz de log dos preços.

```
matriz_logpreco <- data.frame(ITSA4 = log(dados_ITSA4$weighted.price),
    ITUB4 = log(dados_ITUB4$weighted.price), BBDC4 = log(dados_BBDC4$weighted.price),
    ABEV3 = log(dados_ABEV3$weighted.price), BBSE3 = log(dados_BBSE3$weighted.price))
library(BMR)
gtsplot(matriz_logpreco)
```




```
summary(matriz_logpreco)
```

```
##      ITSA4           ITUB4           BBDC4           ABEV3
##  Min.   :1.835   Min.   :3.132   Min.   :2.832   Min.   :2.718
##  1st Qu.:2.005   1st Qu.:3.346   1st Qu.:3.150   1st Qu.:2.898
##  Median :2.123   Median :3.486   Median :3.318   Median :2.933
##  Mean   :2.118   Mean   :3.451   Mean   :3.298   Mean   :2.917
##  3rd Qu.:2.234   3rd Qu.:3.562   3rd Qu.:3.439   3rd Qu.:2.958
##  Max.   :2.388   Max.   :3.689   Max.   :3.719   Max.   :3.016
##      BBSE3
##  Min.   :3.050
##  1st Qu.:3.321
##  Median :3.395
##  Mean   :3.385
##  3rd Qu.:3.465
##  Max.   :3.624
```

```r
cov(matriz_logpreco)
```

```
##             ITSA4         ITUB4        BBDC4         ABEV3        BBSE3
## ITSA4  0.01787547  0.0162354679  0.025154358 -0.0014471303 0.0128461741
## ITUB4  0.01623547  0.0175130525  0.024189973 -0.0002988126 0.0123718427
## BBDC4  0.02515436  0.0241899729  0.040909538 -0.0032696184 0.0173792039
## ABEV3 -0.00144713 -0.0002988126 -0.003269618  0.0035513228 0.0007159475
## BBSE3  0.01284617  0.0123718427  0.017379204  0.0007159475 0.0127852450
```

Criando matriz com os dados log retorno.

```r
matriz_logrtn <- data.frame(ITSA4 = dados_ITSA4$log_retorno, ITUB4 = dados_ITUB4$log_retorno,
    BBDC4 = dados_BBDC4$log_retorno, ABEV3 = dados_ABEV3$log_retorno, BBSE3 = dados_BBSE3$log_retorno)
summary(matriz_logrtn)
```

```
##      ITSA4                ITUB4                BBDC4
##  Min.   :-1.079e-01   Min.   :-9.911e-02   Min.   :-1.976e-01
##  1st Qu.:-3.412e-03   1st Qu.:-3.276e-03   1st Qu.:-3.698e-03
##  Median :-1.467e-04   Median :-5.178e-05   Median :-3.910e-06
##  Mean   :-4.198e-05   Mean   :-5.640e-06   Mean   :-5.718e-05
##  3rd Qu.: 3.154e-03   3rd Qu.: 3.205e-03   3rd Qu.: 3.599e-03
##  Max.   : 7.451e-02   Max.   : 8.019e-02   Max.   : 7.990e-02
##      ABEV3                BBSE3
##  Min.   :-4.289e-02   Min.   :-6.050e-02
##  1st Qu.:-2.287e-03   1st Qu.:-3.531e-03
##  Median : 7.486e-05   Median : 1.017e-05
##  Mean   : 5.483e-05   Mean   :-2.386e-05
##  3rd Qu.: 2.359e-03   3rd Qu.: 3.362e-03
##  Max.   : 2.241e-02   Max.   : 8.743e-02
```

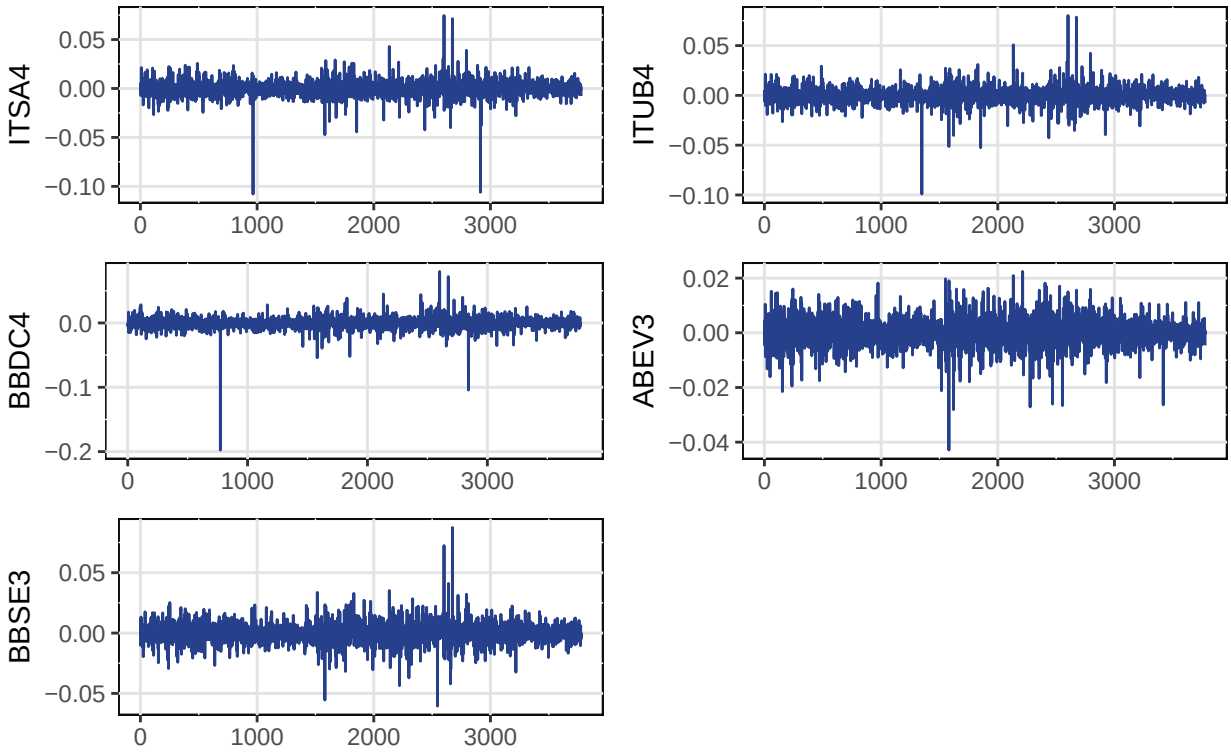
```r
cov(matriz_logrtn)
```

```
##               ITSA4        ITUB4        BBDC4        ABEV3        BBSE3
## ITSA4 5.167608e-05 4.335244e-05 4.281033e-05 1.524743e-05 2.995500e-05
## ITUB4 4.335244e-05 5.104148e-05 4.627070e-05 1.615010e-05 3.064427e-05
## BBDC4 4.281033e-05 4.627070e-05 6.736253e-05 1.655762e-05 3.067263e-05
## ABEV3 1.524743e-05 1.615010e-05 1.655762e-05 2.114459e-05 1.369144e-05
## BBSE3 2.995500e-05 3.064427e-05 3.067263e-05 1.369144e-05 5.436357e-05
```

```r
cor(matriz_logrtn)
```

```
##            ITSA4     ITUB4     BBDC4     ABEV3     BBSE3
## ITSA4 1.0000000 0.8441257 0.7255955 0.4612668 0.5651589
## ITUB4 0.8441257 1.0000000 0.7891058 0.4916024 0.5817463
## BBDC4 0.7255955 0.7891058 1.0000000 0.4387216 0.5068598
## ABEV3 0.4612668 0.4916024 0.4387216 1.0000000 0.4038272
## BBSE3 0.5651589 0.5817463 0.5068598 0.4038272 1.0000000
```

```r
library(BMR)
gtsplot(matriz_logrtn)
```




```
head(matriz_logrtn)
```

```
##           ITSA4         ITUB4        BBDC4         ABEV3        BBSE3
## 1 -0.0079058649 -0.0057653310 -0.004521064 -0.0016150278 -0.010986101
## 2  0.0050112343  0.0045462710 -0.002177734  0.0043904510  0.008955299
## 3 -0.0030588852 -0.0059963761 -0.001079848  0.0004751700 -0.004649955
## 4 -0.0036985681 -0.0022077264 -0.006330370 -0.0045601580 -0.009555026
## 5  0.0019715511  0.0028860314  0.003117587 -0.0008106029  0.013210150
## 6  0.0002158613 -0.0003814315 -0.004567171 -0.0029404438 -0.006192878
```

3 Testes de Estacionariedade

```
library(BMR)
library(knitr)
stat1 <- stationarity(matriz_preco, 4, 8)
```

```
kable(stat1$KPSS, caption = "Teste KPSS (preço)", format = "latex", booktabs = TRUE,
    longtable = TRUE, digits = 2)
```

Tabela 1: Teste KPSS (preço)

	ITSA4	ITUB4	BBDC4	ABEV3	BBSE3	1 Pct	2.5 Pct	5 Pct	10 Pct
Time Trend:	11.18	12.47	15.64	7.35	6.16	0.22	0.18	0.15	0.12
No Trend:	39.72	20.86	37.71	22.77	25.09	0.74	0.57	0.46	0.35

```
kable(stat1$ADF, caption = "Teste ADF (preço)", format = "latex", booktabs = TRUE,
    longtable = TRUE, digits = 2)
```

Tabela 2: Teste ADF (preço)

	ITSA4	ITUB4	BBDC4	ABEV3	BBSE3	1 Pct	2.5 Pct	5 Pct	10 Pct
Time Trend:	-2.09	-1.73	-1.40	-3.10	-2.46	-3.96	-3.66	-3.41	-3.12
Constant:	-1.98	-1.94	-1.85	-2.81	-2.34	-3.43	-3.12	-2.86	-2.57
Neither:	-0.61	-0.30	-0.80	0.57	-0.45	-2.58	-2.23	-1.95	-1.62

```
kable(stat1$ADFLags, caption = "Defagens do teste ADF (preço)", format = "latex",
    booktabs = TRUE, longtable = TRUE, digits = 2)
```

Tabela 3: Defagens do teste ADF (preço)

	Trend Model	Drift Model	None
ITSA4	1	1	1
ITUB4	1	1	1
BBDC4	1	1	1
ABEV3	3	3	3
BBSE3	1	1	1

```
stat2 <- stationarity(matriz_logpreco, 4, 8)
```

```
kable(stat2$KPSS, caption = "Teste KPSS (log preço)", format = "latex",
    booktabs = TRUE, longtable = TRUE, digits = 2)
```

Tabela 4: Teste KPSS (log preço)

	ITSA4	ITUB4	BBDC4	ABEV3	BBSE3	1 Pct	2.5 Pct	5 Pct	10 Pct
Time Trend:	11.09	12.39	14.92	7.46	6.27	0.22	0.18	0.15	0.12
No Trend:	38.14	20.20	34.35	23.00	24.10	0.74	0.57	0.46	0.35

```
kable(stat2$ADF, caption = "Teste ADF (log preço)", format = "latex", booktabs = TRUE,
    longtable = TRUE, digits = 2)
```

Tabela 5: Teste ADF (log preço)

	ITSA4	ITUB4	BBDC4	ABEV3	BBSE3	1 Pct	2.5 Pct	5 Pct	10 Pct
Time Trend:	-2.03	-1.71	-1.27	-3.06	-2.44	-3.96	-3.66	-3.41	-3.12
Constant:	-1.97	-1.91	-1.73	-2.80	-2.34	-3.43	-3.12	-2.86	-2.57
Neither:	-0.45	-0.12	-0.49	0.69	-0.25	-2.58	-2.23	-1.95	-1.62

```
kable(stat2$ADFLags, caption = "Defagens do teste ADF (log preço)", format = "latex",
    booktabs = TRUE, longtable = TRUE, digits = 2)
```

Tabela 6: Defagens do teste ADF (log preço)

	Trend Model	Drift Model	None
ITSA4	1	1	1
ITUB4	1	1	1
BBDC4	1	1	1
ABEV3	3	3	3
BBSE3	1	1	1

```
stat3 <- stationarity(matriz_logrtn, 4, 8)
```

```
kable(stat3$KPSS, caption = "Teste KPSS (log retorno)", format = "latex",
    booktabs = TRUE, longtable = TRUE, digits = 2)
```

Tabela 7: Teste KPSS (log retorno)

	ITSA4	ITUB4	BBDC4	ABEV3	BBSE3	1 Pct	2.5 Pct	5 Pct	10 Pct
Time Trend:	0.04	0.04	0.05	0.03	0.04	0.22	0.18	0.15	0.12
No Trend:	0.09	0.14	0.23	0.05	0.06	0.74	0.57	0.46	0.35

```
kable(stat3$ADF, caption = "Teste ADF (log retorno)", format = "latex",
    booktabs = TRUE, longtable = TRUE, digits = 2)
```

Tabela 8: Teste ADF (log retorno)

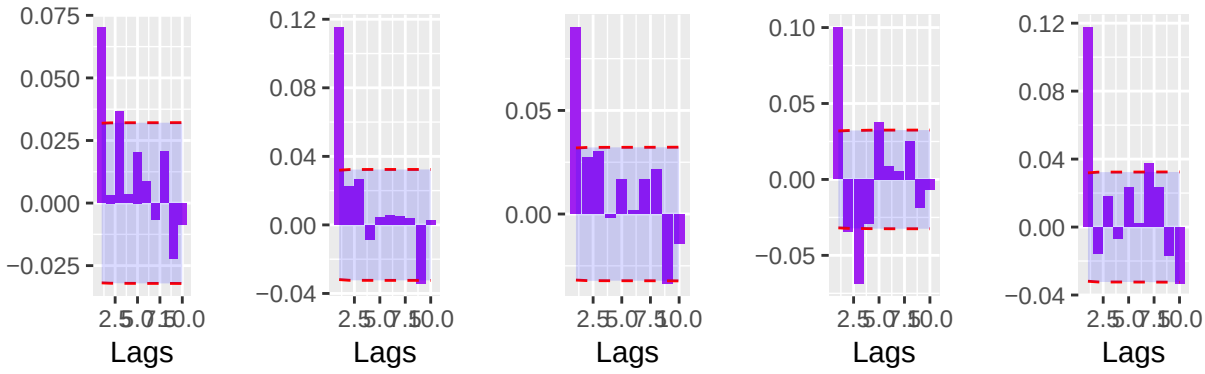
	ITSA4	ITUB4	BBDC4	ABEV3	BBSE3	1 Pct	2.5 Pct	5 Pct	10 Pct
Time Trend:	-41.94	-40.48	-40.65	-37.49	-42.03	-3.96	-3.66	-3.41	-3.12
Constant:	-41.94	-40.47	-40.63	-37.49	-42.03	-3.43	-3.12	-2.86	-2.57
Neither:	-41.94	-40.48	-40.63	-37.49	-42.04	-2.58	-2.23	-1.95	-1.62

```
kable(stat3$ADFLags, caption = "Defagens do teste ADF (log retorno)", format = "latex",
    booktabs = TRUE, longtable = TRUE, digits = 2)
```

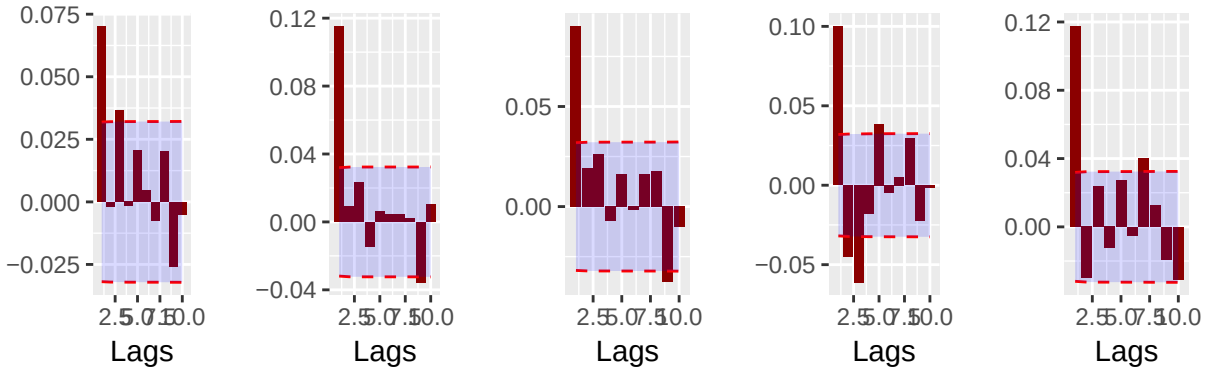
Tabela 9: Defagens do teste ADF (log retorno)

	Trend Model	Drift Model	None
ITSA4	1	1	1
ITUB4	1	1	1
BBDC4	1	1	1
ABEV3	2	2	2
BBSE3	1	1	1

```
library(BMR)
gacf(matriz_logrtn)
```

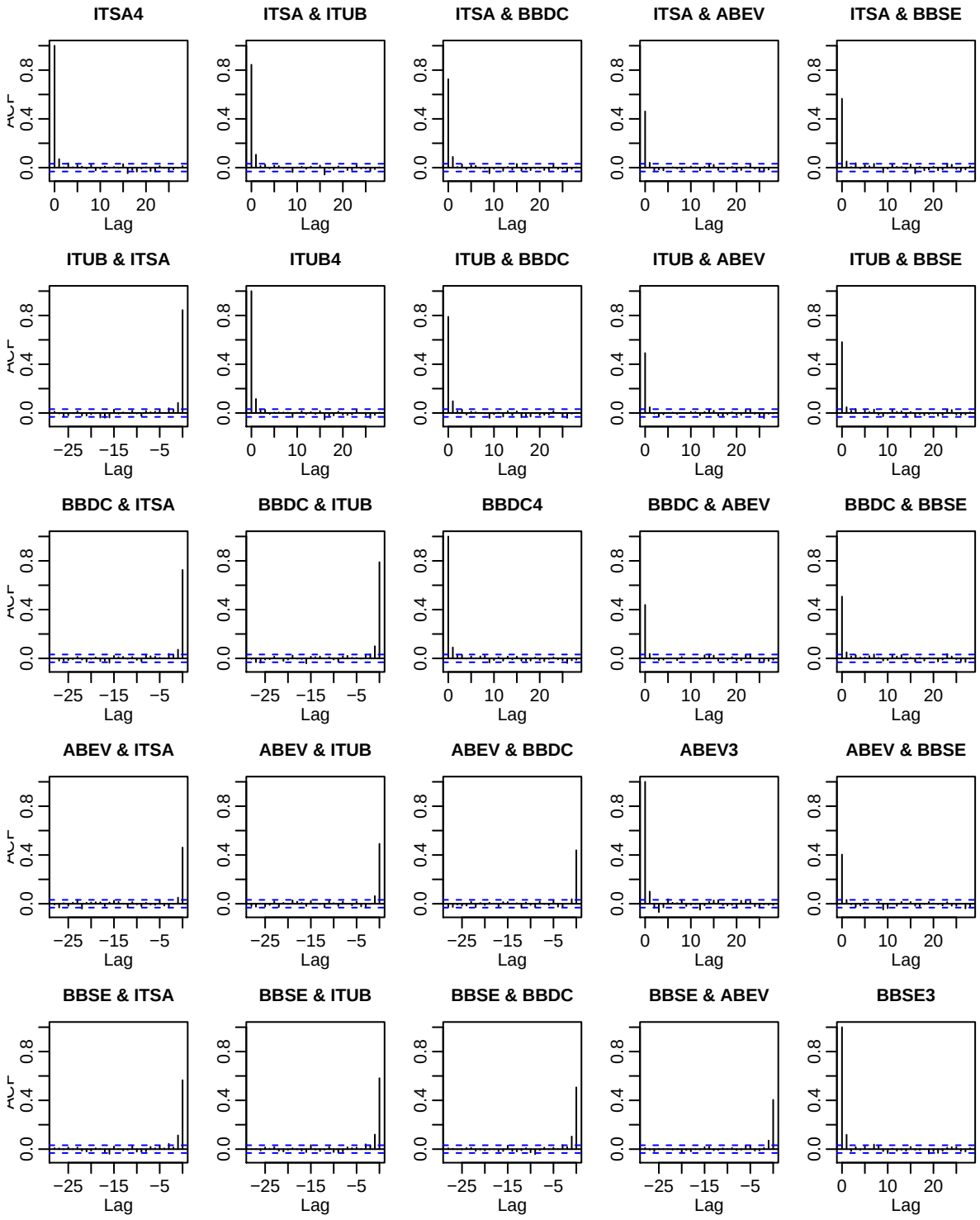



```
gpacf(matriz_logrtn)
```

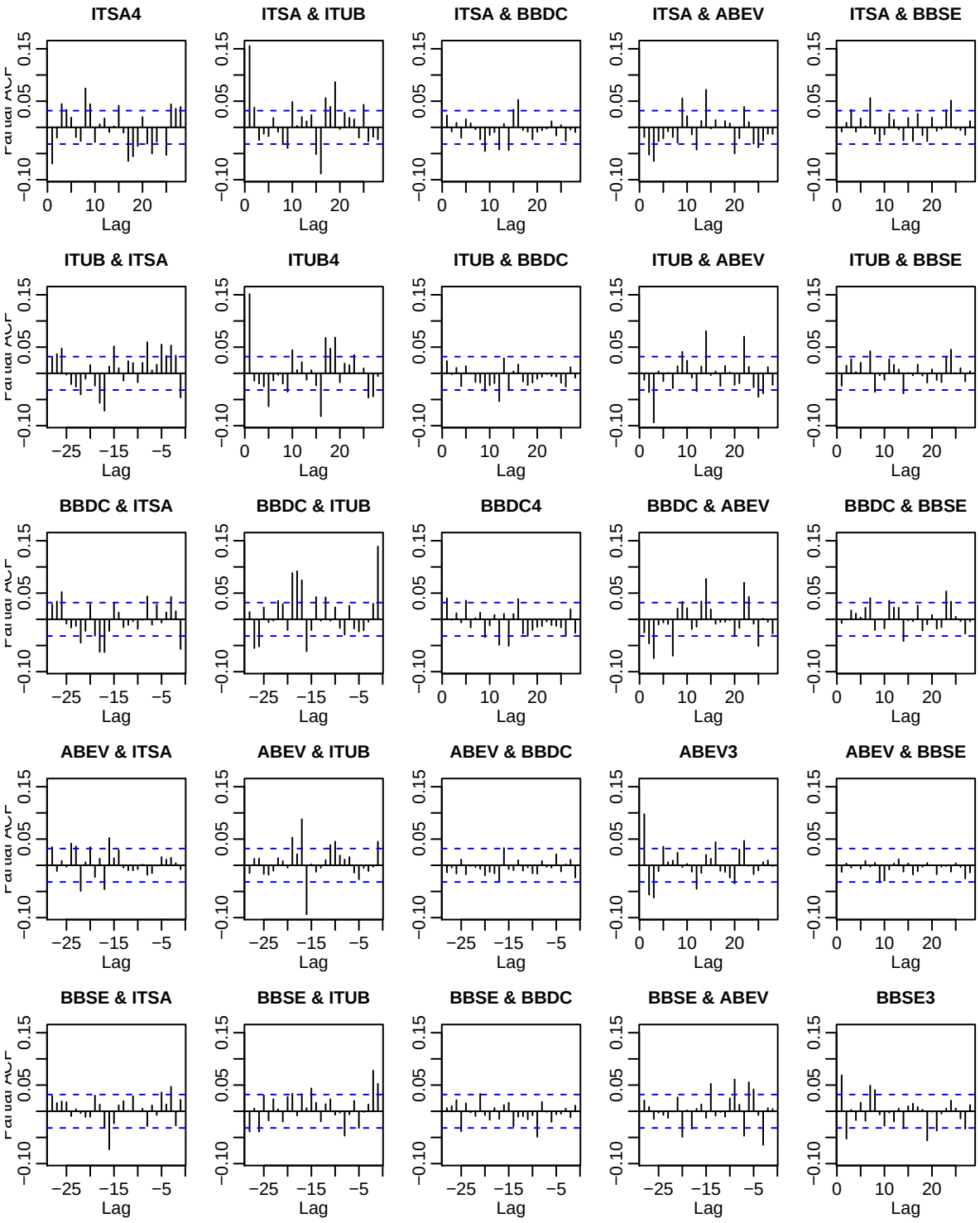



```
save(matriz_preco, file = "matriz_preco.Rda")
save(matriz_logrtn, file = "matriz_logrtn.Rda")
```

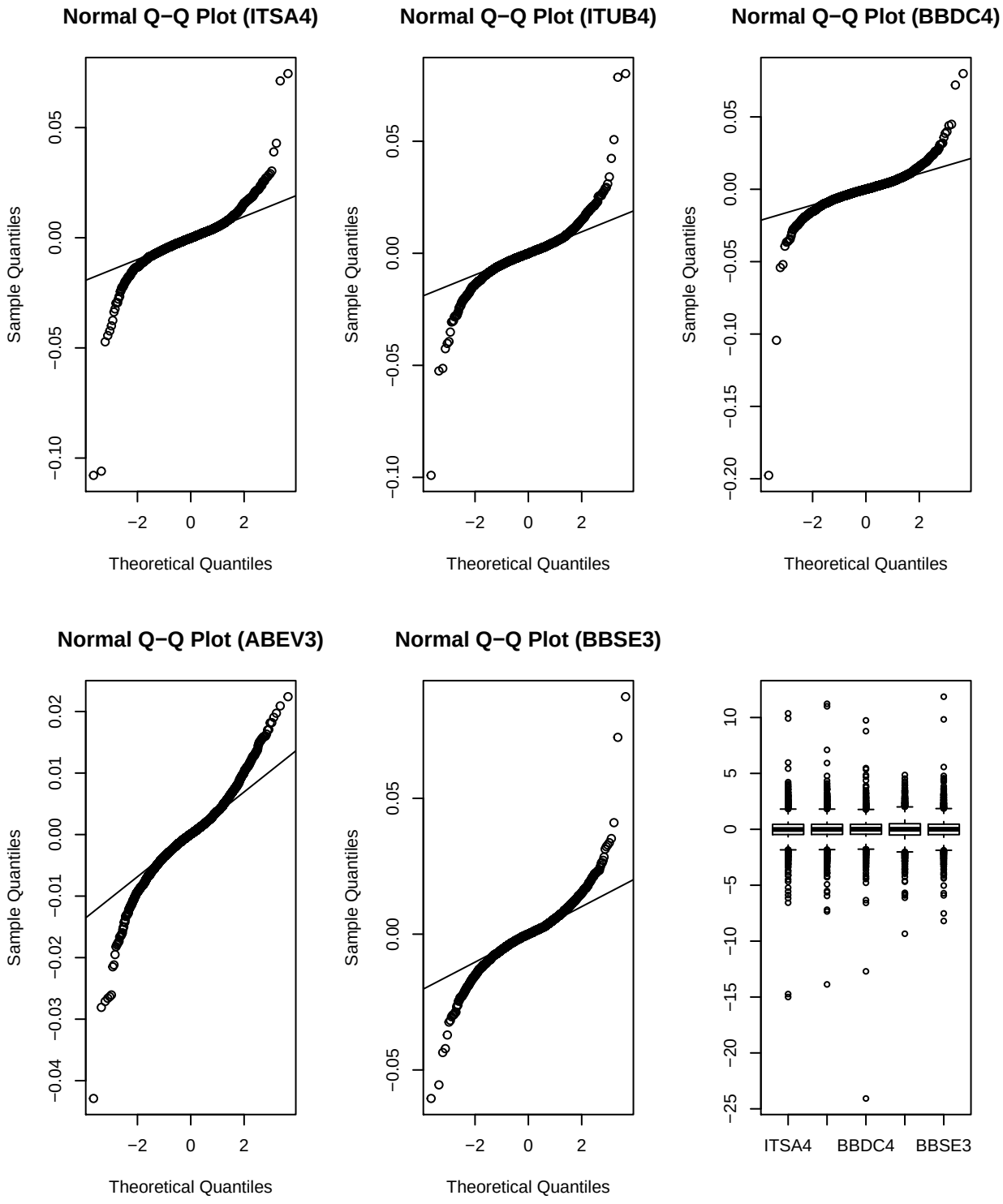
```
acf(matriz_logrtn)
```



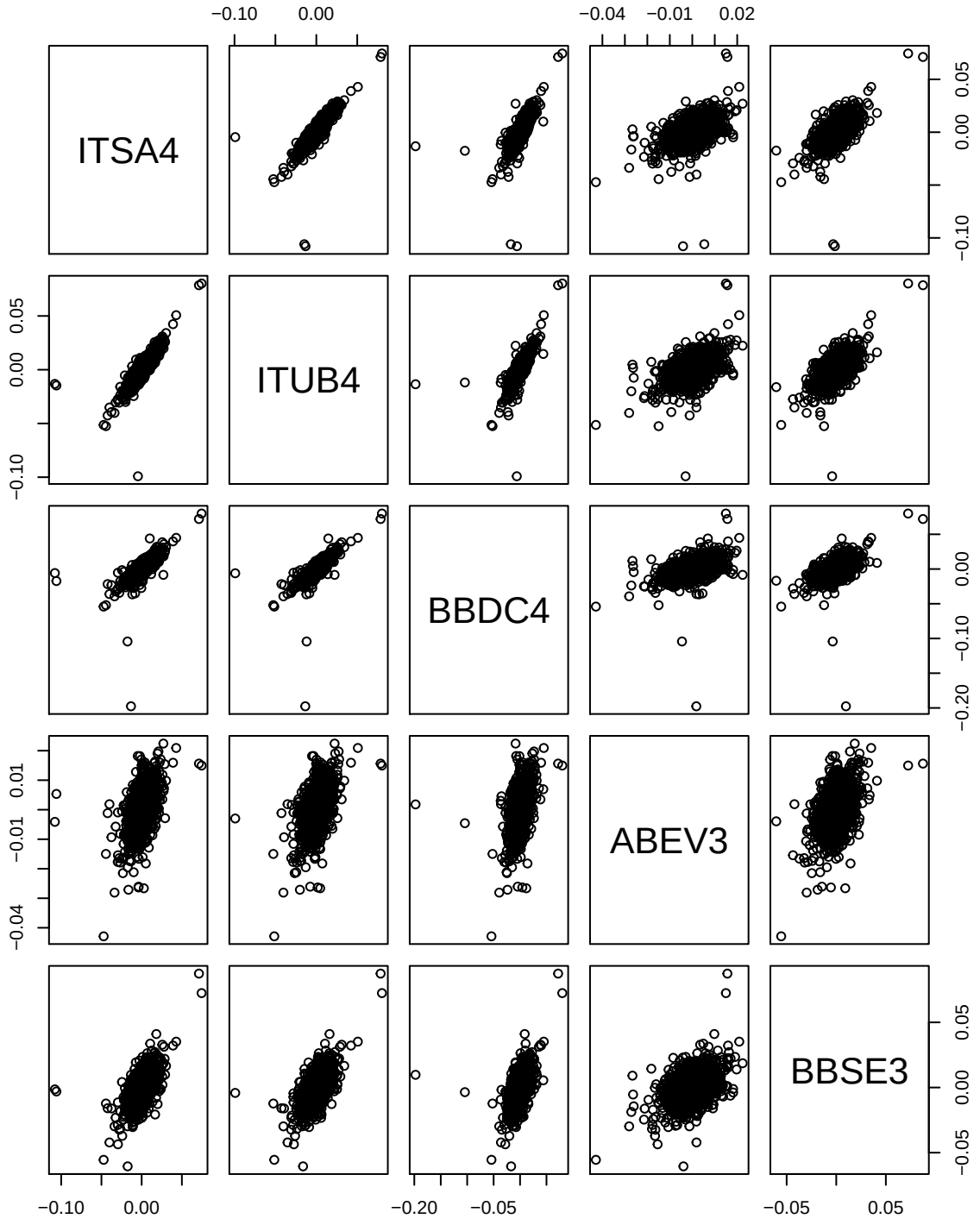

```
acf(matriz_logrtn, type = "partial")
```



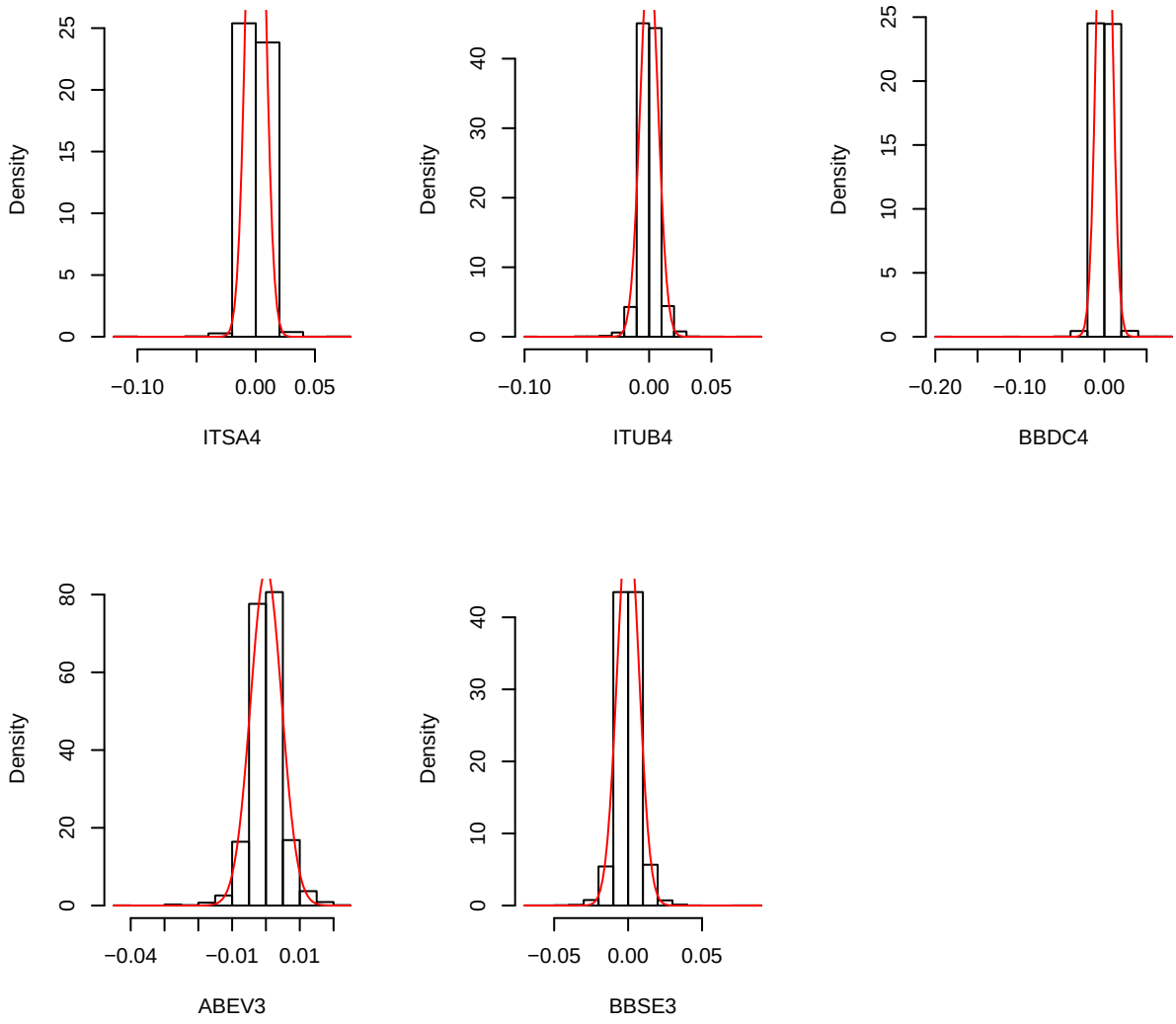

```
library(MVN)
uniPlot(matriz_logrtn, type = "qqplot")
uniPlot(matriz_logrtn, type = "box")
```




```
uniPlot(matriz_logrtn, type = "scatter")
```




```
uniPlot(matriz_logrtn, type = "histogram")
```




```
library(MVN)
library(fBasics)
uniNorm(matriz_logrtn)
```

```
## $`Descriptive Statistics`
##          n Mean Std.Dev Median    Min   Max   25th  75th   Skew Kurtosis
## ITSA4 3777    0   0.007      0 -0.108 0.075 -0.003 0.003 -1.225   34.177
## ITUB4 3777    0   0.007      0 -0.099 0.080 -0.003 0.003 -0.135   22.259
## BBDC4 3777    0   0.008      0 -0.198 0.080 -0.004 0.004 -3.842  101.520
## ABEV3 3777    0   0.005      0 -0.043 0.022 -0.002 0.002 -0.383    5.151
## BBSE3 3777    0   0.007      0 -0.060 0.087 -0.004 0.003  0.369   11.993
##
## $`Shapiro-Wilk's Normality Test`
##    Variable Statistic   p-value Normality
## 1   ITSA4      0.8502         0    NO
## 2   ITUB4      0.8742         0    NO
## 3   BBDC4      0.7937         0    NO
```

```
## 4   ABEV3      0.9517         0    NO
## 5   BBSE3      0.9191         0    NO
```

```
jarqueberaTest(matriz_logrtn$ITSA4)
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 184985.3846
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Thu Dec 08 15:46:56 2016 by user: Lucca
```

```
jarqueberaTest(matriz_logrtn$ITUB4)
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 78079.208
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Thu Dec 08 15:46:56 2016 by user: Lucca
```

```
jarqueberaTest(matriz_logrtn$BBDC4)
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 1633035.358
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Thu Dec 08 15:46:56 2016 by user: Lucca
```

```
jarqueberaTest(matriz_logrtn$ABEV3)
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 4274.4093
```

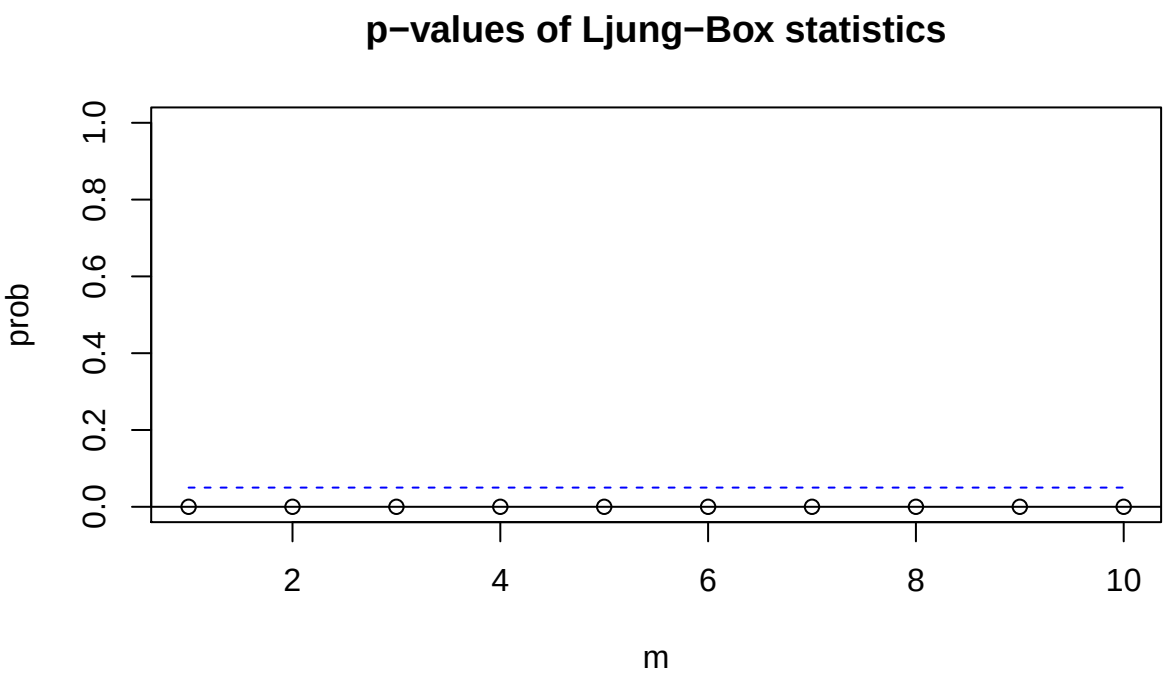
```
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Thu Dec 08 15:46:56 2016 by user: Lucca
```

```
jarqueberaTest(matriz_logrtn$BBSE3)
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 22752.1698
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Thu Dec 08 15:46:56 2016 by user: Lucca
```

```
library(MTS)
mq(matriz_logrtn, lag = 10)
```

```
## Ljung-Box Statistics:
##         m       Q(m)     df    p-value
##  [1,]     1       156      25        0
##  [2,]     2       202      50        0
##  [3,]     3       252      75        0
##  [4,]     4       265     100        0
##  [5,]     5       298     125        0
##  [6,]     6       318     150        0
##  [7,]     7       349     175        0
##  [8,]     8       408     200        0
##  [9,]     9       442     225        0
## [10,]    10       465     250        0
```


p−values of Ljung−Box statistics
prob
0.0
0.2
0.4
0.6
0.8
1.0
2
4
6
8
10
m