3/11/2019

HW 5

Ian Douglas

Setup:

Read in data

```
df <- read.table("~/Desktop/Grad School/Columbia/Spring 2019/Computational
Stats/R/HW/Datasets/ebay.txt",header=T)
#Define variables
sp5 <- as.vector(df$SP500)
ebay <- as.vector(df$EBAY)
```

Part A

$$Return_{Day_i} = \frac{Price_{Day_i} - Price_{Day_{i-1}}}{Price_{Day_{i-1}}}$$

```
#Since the data is in chronologically descending order, index i PLUS 1 is
#actually day i MINUS 1
rtrn.sp5 <- NULL
rtrn.ebay <- NULL
for (i in 1:(nrow(df)-1)) {
  rtrn.sp5[i] <- ((sp5)[i] - (sp5[i+1]))/(sp5[i+1])
  rtrn.ebay[i] <- ((ebay)[i] - (ebay[i+1]))/(ebay[i+1])
}
returns <- cbind(matrix(rtrn.sp5,ncol=1),matrix(rtrn.ebay,ncol=1))
colnames(returns) <- c("S&P 500", "Ebay Stock")
rownames(returns) <- df$DATE[1:60]
returns

##                  S&P 500   Ebay Stock
## 3/31/2005 -0.0006940859 -0.011408862
## 3/30/2005  0.0137725681  0.044044321
## 3/29/2005 -0.0075961440  0.005291005
## 3/28/2005  0.0024414813  0.011549296
## 3/24/2005 -0.0009466709 -0.011692650
## 3/23/2005  0.0006998319  0.001114827
## 3/22/2005 -0.0101961513 -0.006919458
## 3/21/2005 -0.0049342244  0.031990860
## 3/18/2005 -0.0004705052 -0.030730897
## 3/17/2005  0.0018012407 -0.009868421
## 3/16/2005 -0.0080818201 -0.015915835
## 3/15/2005 -0.0075238435  0.016173246
## 3/14/2005  0.0056246250 -0.046274510
## 3/11/2005 -0.0075832127 -0.015697375
## 3/10/2005  0.0018558256 -0.026065163
## 3/9/2005  -0.0101850865 -0.023255814
## 3/8/2005  -0.0047987856 -0.024826928
```

```
## 3/7/2005   0.0026102183  0.003353293
## 3/4/2005   0.0096243608  0.006266570
## 3/3/2005   0.0003222927 -0.013786546
## 3/2/2005  -0.0002726349 -0.011977454
## 3/1/2005   0.0056580259 -0.006069094
## 2/28/2005 -0.0064142252  0.014204545
## 2/25/2005  0.0093067822 -0.003068209
## 2/24/2005  0.0078938529  0.012667304
## 2/23/2005  0.0056073504  0.018004866
## 2/22/2005 -0.0145057798 -0.032030146
## 2/18/2005  0.0006995628 -0.006086142
## 2/17/2005 -0.0079233934 -0.007550238
## 2/16/2005  0.0001818002  0.008079625
## 2/15/2005  0.0032997828  0.012808349
## 2/14/2005  0.0006969219  0.027415621
## 2/11/2005  0.0069255896  0.010465403
## 2/10/2005  0.0042114447  0.028231422
## 2/9/2005  -0.0085752308  0.005857634
## 2/8/2005   0.0004826415  0.038894034
## 2/7/2005  -0.0010889172 -0.003821824
## 2/4/2005   0.0110430376 -0.017353017
## 2/3/2005  -0.0027656953 -0.020050761
## 2/2/2005   0.0031780463  0.011163865
## 2/1/2005   0.0068908886 -0.043803681
## 1/31/2005  0.0084602513  0.004932182
## 1/28/2005 -0.0027159338 -0.019347037
## 1/27/2005  0.0004088342  0.004616132
## 1/26/2005  0.0048441900  0.028485757
## 1/25/2005  0.0040042965 -0.028286998
## 1/24/2005 -0.0035277899 -0.042765834
## 1/21/2005 -0.0064147829  0.032641306
## 1/20/2005 -0.0077830209 -0.191363416
## 1/19/2005 -0.0094901253 -0.031211808
## 1/18/2005  0.0096748050  0.011121673
## 1/14/2005  0.0060045013  0.019281077
## 1/13/2005 -0.0086301255 -0.037668998
## 1/12/2005  0.0039814369  0.022987409
## 1/11/2005 -0.0060995589 -0.023017426
## 1/10/2005  0.0034227232  0.006849315
## 1/7/2005  -0.0014311089  0.003767188
## 1/6/2005   0.0035058374 -0.042560866
## 1/5/2005  -0.0036277934 -0.003683407
## 1/4/2005  -0.0116714362 -0.024537727
```

##Part B

```r
#Correlation between the two lists of returns
cor(returns[,1],returns[,2])
```

```
## [1] 0.3094859
```

Output shows a correlation of .3094

##Part C

```
#Estimate the SE of the correlation coefficient using 1000 bootstrap samples
library("boot")
cor.boot <- function(data, i){
    dat <- data
  dat2 <- dat[i,]
  cor.boot <- cor(dat2[i,1],dat[i,2])
    return(cor(dat2[,1], dat2[,2]))
}

corrs <- boot(returns, cor.boot, R=1000)
SE <- sd(corrs$t[,1])/sqrt(corrs$R)
SE

## [1] 0.003274843
```

####Output shows a SE of .003

Part D

```
# Report the BCa 95% CI for the correlation coefficient:
boot.ci(corrs, type="all")

## Warning in boot.ci(corrs, type = "all"): bootstrap variances needed for
## studentized intervals

## BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
## Based on 1000 bootstrap replicates
##
## CALL :
## boot.ci(boot.out = corrs, type = "all")
##
## Intervals :
## Level      Normal              Basic
## 95%   ( 0.1057,  0.5117 )   ( 0.1264,  0.5279 )
##
## Level     Percentile            BCa
## 95%   ( 0.0910,  0.4925 )   ( 0.0778,  0.4846 )
## Calculations and Intervals on Original Scale
```

Output shows the 95% BCa CI to be (.0473, .4705)

Part E

Calculate $z_r \pm 1.96\sqrt{\frac{1}{n-3}}$ where $z_r = \frac{1}{2}log\sqrt{\frac{1+r}{1-r}}$

```
zr <- NULL
for (i in 1:corrs$R) {
  zr[i] <- (.5)*log(sqrt((1+mean(corrs$t[i,1]))/(1-mean(corrs$t[i,1]))))
}
zr.mean <- mean(zr)
```

```r
int <- 1.96*sqrt(1/(nrow(returns)-3))
Fisher.ci <- c(zr.mean-int,zr.mean+int)
Fisher.ci

## [1] -0.09720736  0.42200933
```

####Output shows a CI from -0.096 to .424

Part F

```r
tanh(Fisher.ci)

## [1] -0.09690233  0.39862183

# [1] -0.09528203  0.39999647
```

####Compared to the CI in part D, the re-scaled Fisher CI is larger. It also crosses zero, bringing into question the statistical significance of the correlation. more