CFRM 543 Homework 5

Joanna Nowak

Problem 5.)

Estimates of correlation for DJIA stocks:

Column1	SIZE	BOOK	E2P	P2B
AA	0.182777	-0.00827	-0.12173	0.155019
VZ	0.16699	-0.03689	-0.1273	0.146391
BA	0.157602	-0.01485	-0.14779	0.069945
CAT	0.066068	-0.04402	-0.31235	0.151962
JPM	0.120436	0.042484	-0.0105	0.063825
CVX	0.052949	0.012033	-0.05735	0.071169
KO	0.189657	0.074213	-0.02743	0.036835
DIS	0.206332	0.109174	-0.04948	0.134959
DD	0.152364	-0.08623	0.066976	0.296186
XOM	0.126579	0.033732	-0.02939	0.090983
GE	0.075973	0.06492	-0.1985	0.01703
HPQ	0.16327	0.11518	-0.14874	0.004477
HD	0.086035	0.013235	-0.0887	0.050826
INTC	0.077135	0.016484	-0.03112	0.047222
IBM	0.163401	-0.0286	-0.05373	0.117282
JNJ	0.075007	-0.02311	-0.00465	0.08907
MCD	0.142682	0.171028	-0.14601	0.072278
MRK	0.108015	0.071968	-0.07254	0.033792
MMM	0.108469	-0.022	-0.1323	0.107219
BAC	0.144086	-0.01312	-0.20958	0.08137
PFE	0.03107	0.052437	-0.05087	-0.01158
PG	0.091717	-0.00432	-0.07007	0.065956
T	0.097352	0.034625	-0.09959	0.042517
UNH	-0.12276	-0.14151	-0.16007	0.121799
UTX	0.051781	-0.03917	-0.27824	0.263419
WMT	0.166228	0.043584	-0.02295	0.03569
MSFT	0.142175	0.019466	-0.05323	0.076163
CSCO	0.024965	0.034544	0.111562	-0.00153
TRV	-0.01453	-0.0265	-0.03086	-0.00822

Robust estimates of correlation for DJIA stocks:

Column1	SIZE	BOOK	E2P	P2B
AA	0.134431	-0.02649	-0.1506	0.145596
VZ	0.162615	0.043628	-0.07395	0.104152
BA	0.07741	-0.01235	-0.05138	0.071249
CAT	0.039462	-0.06319	-0.2567	0.13978
JPM	0.140962	0.057113	0.01418	0.057527
CVX	0.102319	0.051044	-0.0725	0.072087
KO	0.212473	0.066331	-0.02751	-0.00118
DIS	0.156481	0.125889	0.010691	0.098709
DD	0.165877	-0.10438	0.027377	0.296357
XOM	0.119521	0.033971	-0.01639	0.045642
GE	0.037986	0.090206	-0.06945	0.00667
HPQ	0.179043	0.088938	-0.14309	0.114628
HD	0.086001	-0.01217	-0.10641	0.110446
INTC	0.114824	0.033538	-0.05298	0.081551
IBM	0.158253	-0.02996	0.06922	0.138891
JNJ	0.142277	-0.0112	-0.02551	0.085157
MCD	0.078662	0.119764	-0.04586	0.030064
MRK	0.088707	0.105028	-0.05537	-0.0113
MMM	0.158212	0.024315	-0.12385	0.084301
BAC	0.051199	-0.08232	-0.06427	0.146061
PFE	0.004261	0.086229	-0.00351	-0.01414
PG	0.013605	-0.02377	-0.06329	0.159564
T	0.134902	0.099331	-0.07209	0.014493
UNH	-0.13468	-0.12702	-0.10011	0.070712
UTX	0.039557	-0.05438	-0.16646	0.210232
WMT	0.169961	0.058336	-0.04212	0.036579
MSFT	0.108238	0.014202	-0.02743	0.060467
CSCO	0.120504	0.007055	-0.05387	0.069394
TRV	0.113363	0.012572	0.032802	0.085812

R code:

```
setwd("C:\\Users\\joanna\\Documents\\CFRM_543\\Homework\\HW5")

library(xts)
library(robust)
dat.all <- read.csv(file="MonthlyFactorDataSet.csv",sep=",",header=TRUE,as.is=TRUE)
```

```
dat.all$DATE <- as.Date(dat.all$DATE, format<- "%m/%d/%Y")
# extract a list of tickers
tickers = unique(dat.all[,"TICKER"])
# Extract date, ticker, price, return and three factor exposures dat.all
facdat = dat.all[,c("DATE","TICKER","RET","SIZE","BOOK", "E2P","P2B")]
# Replace factor names with shorter names
names(facdat) = c("DATE","TICKER","RET","SIZE","BOOK", "E2P","P2B")


n = length(tickers)
# construct two tables 29x4 tables for the 29 tickers and 4 factors to hold non-robust
# and robust estimates of the IC
IC <- matrix(nrow=n, ncol=4, rep(0,(n*4) ) )
rownames(IC) <- tickers
colnames(IC) <- names(facdat[4:7])
RobIC <- IC
# lop throgh the list of tickers
for(i in 1:n){
  TCK = tickers[i]
  facdatTck = facdat[facdat$TICKER == TCK,]
  facdatTck.ts = as.xts(facdatTck[,4:7],order.by = facdatTck[,1])
  # for each tickerand each factor estimate the IC  and write to the tables
  for(j in 1:4){
     ret = facdatTck[,"RET"]
     factor <- facdatTck.ts[,j]
     k <- length(ret)
     # lag the returns by one time unit (month)
     x <- ret[1:k-1]
     y <- factor[2:k]
     IC[i,j] <- cor(x,y)
     FactName <- names(facdat[3+j])
     # robust estimate of corrilation
     RobIC[i,j] <- cor(x,y,method = "spearman")
  }
}

# write the rsults to the files
write.table(IC, file="IC.csv", sep=",", col.names=TRUE, row.names=TRUE)
write.table(RobIC, file="RobIC.csv", sep=",", col.names=TRUE, row.names=TRUE)
```

Homework 5 CFRM 543 Joanna Nowak

Problem 1

$$(E_1) = \frac{\partial}{\partial a} MSE(a,b) = 0$$

$$\frac{\partial}{\partial a} MSE(a,b) = \frac{\partial}{\partial a}(y-(a+x^Tb))^2 = 2(y-(a+x^Tb))^2$$

$$= 2y - 2a - 2x^Tb = 0$$

$$\frac{2a}{2} = \frac{2y - 2x^Tb}{2}$$

$$a = y - (x-\bar{x})b$$

$$= \bar{y} - bE[x]$$

$$\frac{\partial}{\partial b} MSE(a,b) = 2(y-(a+x^Tb))^2(x)$$

$$= 2yx - 2ax - 2bx^Tx$$

$$-2bx^Tx = 2yx - 2ax$$

$$b = \frac{2(yx-ax)}{2x^Tx}$$

$$b = \frac{y(x-\bar{x}) - a(x-\bar{x})}{x^Tx}$$

$$b = \frac{(x-\bar{x})(y-\bar{y})}{(x-\bar{x})^2}$$

$$b = \frac{cov(X,Y)}{Var(X)}$$

* $a = \bar{y}$

a is the intercept

Problem 2)

$$\hat{y} = a + bX = E(y) - bE(x) + \frac{cov(X,Y)}{var(x)}X$$

$$= E(y) - \frac{cov(X,Y)}{var(x)}E(x) + \frac{cov(X,Y)}{var(x)}X$$

$$= E(y) + \frac{cov(X,Y)}{var(x)}(x - E(x))$$

Problem 4

Show that the following properties are true.

P1) $E(\hat{\varepsilon}) = E(y - \hat{y}) = 0$

By linearity $E[Y|X] = \tilde{b}X + a$. In other words the regression of y on X is the conditional mean $E[Y|X]$. If $E(\hat{\varepsilon}) > 0$ then $E[Y|X] \neq a + bX$. Therfore $E(\hat{\varepsilon})$ must be $\emptyset$.

Furthermore, if we have a reason to specify that the mean of $\hat{\varepsilon}$ is something other than zero, we should build it into the systematic part of the regression, leaving in the disturbance only the unknown part of it.

(P2) $E[X \cdot \hat{\varepsilon}] = 0$

$E(\hat{\varepsilon}) = 0$ implies that $E(\varepsilon_i | X) = 0$ for all ε_i

This means that no observation of X convey information about the expected value of the error term. It means that X and $\hat{\varepsilon}$ are independent and for independent variables we have

$$E[X\hat{\varepsilon}] = E[X]\,E[\hat{\varepsilon}] = E[X]\,0 = 0$$

(P3) $\text{cov}[X, \hat{\varepsilon}) = E(X \cdot \hat{\varepsilon}) - E[X] \cdot E[\hat{\varepsilon}] = 0$

we have proved in P2 that X and $\hat{\varepsilon}$ are independent which means the covariance between the two random variables ~~are~~ is $\emptyset$.

$$\text{cov}(X\hat{\varepsilon}) = E[X\hat{\varepsilon}] - E[X] \cdot 0 = E[X]E[\hat{\varepsilon}] - E[X]E[\hat{\varepsilon}] = 0$$

Problem 3

$$\sigma_{\hat{y}}^2 = E[y - \hat{y}]^2 =$$

$$E\left[\left(y - E[y] + \frac{cov(xy)}{var(x)} \cdot (x - E(x))\right)^2\right] =$$

$$E\left(y^2 - E(y)^2 - 2(y - E(y))(x - E(x))\frac{cov(xy)}{var(x)} + \frac{cov(xy)^2}{var(x)}\right)$$

$$= var(y) - var(y)\,corr(xy)^2$$

$$= \sigma_y^2 (1 - \rho_{xy}^2)$$