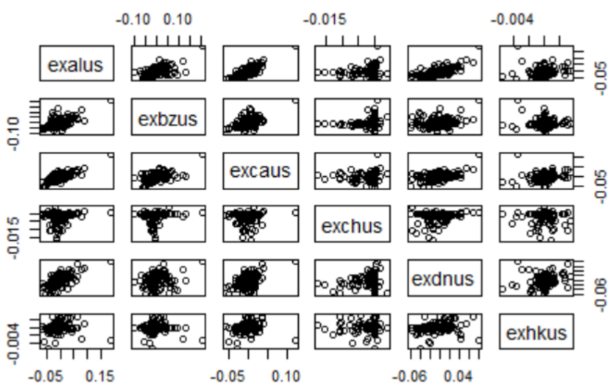
Homework 7

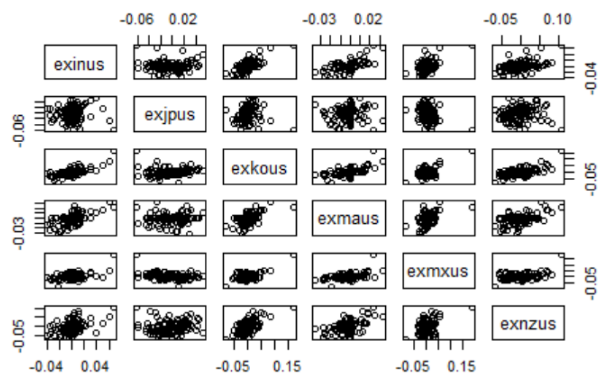
Group 29: Jayoung Kang, Hye-min Jung, Jeong Lim Kim

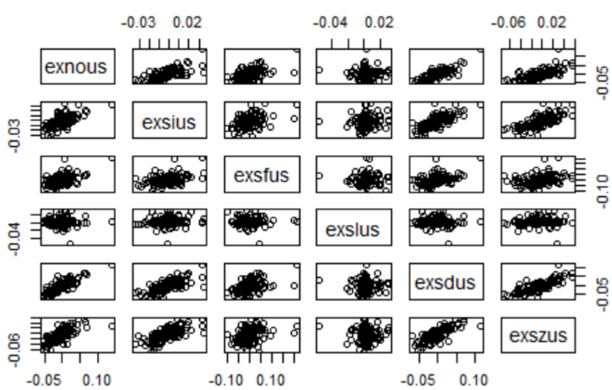
What are the latent factors of international currency pricing? And how do these factor move against US equities? We're going to investigate underlying factors in currency exchange rates and regress the S&P 500 onto this information.

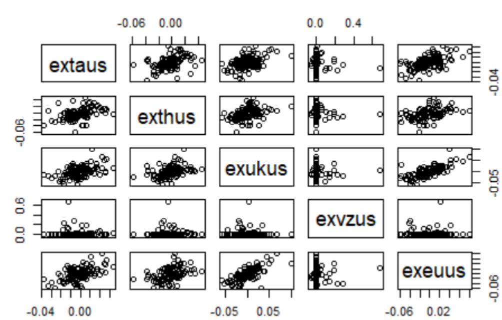
[1] Discuss correlation amongst dimensions of fx. How does this relate to the applicability of factor modelling?

- There seems to be some correlation between variables, which indicates that factor analysis may be useful. Factor analysis decorrelates data by finding principal components that can explain the reason for the correlation.
- The correlation can be seen between similar geographical regions or between countries with similar levels of development.
  - EU and UK
  - Sweden and Switzerland
  - Australia and Canada



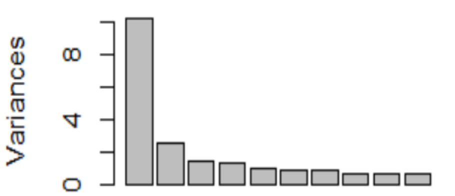




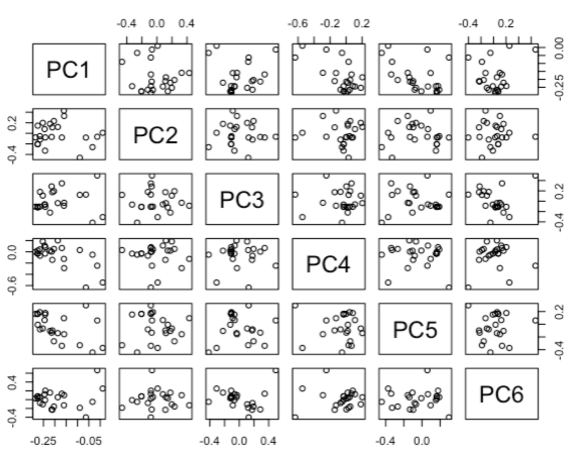


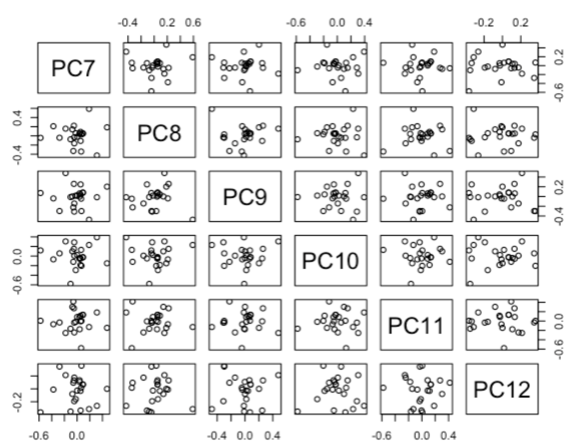
[2] Fit, plot, and interpret principal components.

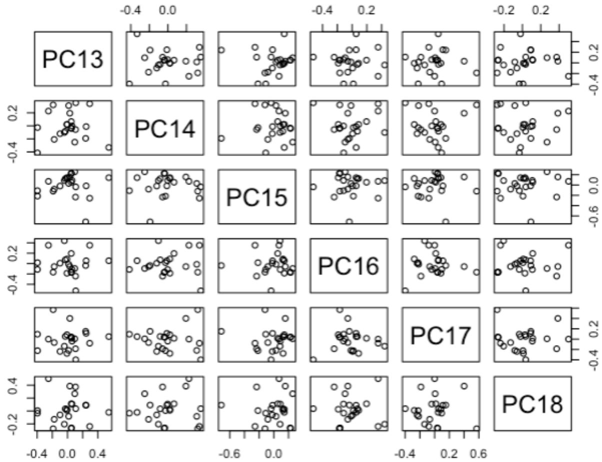
- In total there are 23 principal components
- The PC scree plot indicates that the there is a large difference between the variance of the first PC and the rest of the PCs
- Interpretation of PCs:
    - PC1:
        - The highest loadings for PC1 are Sweden, EU, Denmark, Australia, Norway, South Korea, Singapore.
        - This could be indicative of a latent factor grouping of relatively developed nations. The list of countries consist of countries that are ranked in the the top 12~40 in terms of GDP.
    - PC2:
        - The largest absolute value from negative signs are Brazil and Mexico, which could be indicative of South America
        - The largest absolute value from the positive signs are Japan and Switzerland, which could be indicative of countries that have key currencies.
        - It could be possible that this proxy could depend on the level of fluctuation in the currencies but more insight would be required to confirm this.
    - PC3:
        - The highest loadings for PC3 are Sri Lanka, Taiwan and Thailand which could be indicative of grouping by SEA countries
    - The summary of PCs indicates that 75% of variation can be explained by the first 6 PCs
    - When we pair our PCs, there is no apparent correlation.

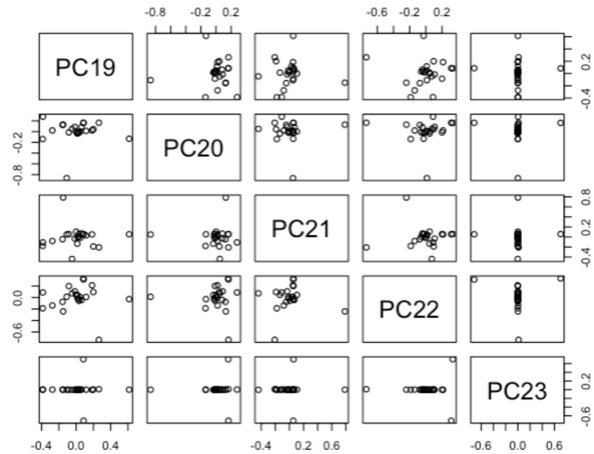



```
Importance of components:
                          PC1    PC2     PC3     PC4     PC5     PC6     PC7     PC8
Standard deviation     3.1904 1.5905 1.18680 1.14792 0.99740 0.93815 0.92009 0.82835
Proportion of Variance 0.4425 0.1100 0.06124 0.05729 0.04325 0.03827 0.03681 0.02983
Cumulative Proportion  0.4425 0.5525 0.61377 0.67107 0.71432 0.75258 0.78939 0.81923
                           PC9    PC10    PC11    PC12    PC13    PC14    PC15    PC16
Standard deviation     0.80841 0.76390 0.69185 0.65917 0.58024 0.56012 0.55254 0.50190
Proportion of Variance 0.02841 0.02537 0.02081 0.01889 0.01464 0.01364 0.01327 0.01095
Cumulative Proportion  0.84764 0.87301 0.89382 0.91271 0.92735 0.94099 0.95427 0.96522
                          PC17    PC18    PC19    PC20    PC21   PC22    PC23
Standard deviation     0.44624 0.41834 0.38808 0.33724 0.30771 0.2580 0.01557
Proportion of Variance 0.00866 0.00761 0.00655 0.00494 0.00412 0.0029 0.00001
Cumulative Proportion  0.97388 0.98149 0.98803 0.99298 0.99709 1.0000 1.00000
```



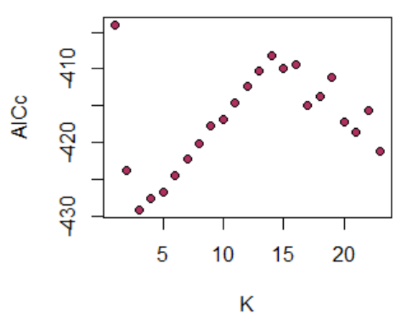


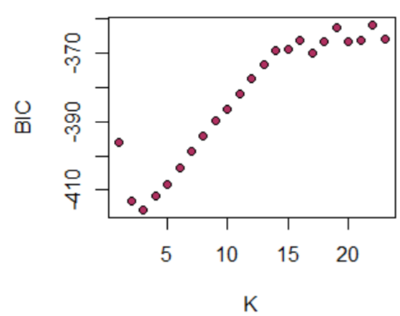




[3] Regress SP500 returns onto currency movement factors, using both ‘glm on first K’ and lasso techniques. Use the results to add to your factor interpretation.

- Using AICc and BIC, only 3 Ks are chosen. However, when using lasso techniques 8 PCs are selected. The PCs chosen by lasso are 1, 2, 3, 15, 17, 20, 21, 23
- The order of PCs for predicting SP500 returns differs from predicting the x-variables only because it considers the new Y variable.
- The PCs chosen by lasso correspond to the dips in the graph of the AICc. In predicting SP500 returns PCs 15, 17, 20, 21, 23 in addition to the first three could be useful. However, it is difficult to determine what the latent factor is based on the leading variables in the PCs because when we examine our PCs 15, 17, 20, 21, 23, there were no logical grouping.





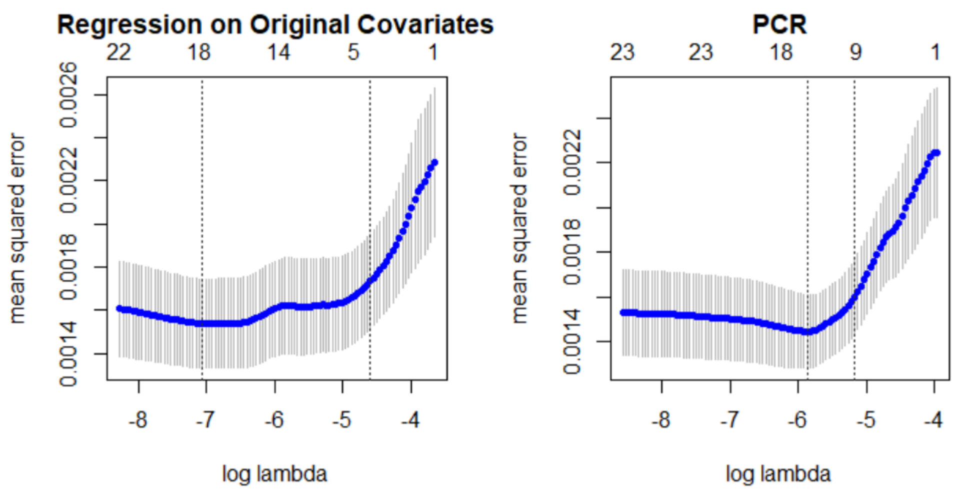
- For example, when we look at PC 23, Denmark and EU are the only nations with an absolute value above zero. But their directions differ, which makes it unclear to conclude what the underlying latent factor is.

```
exalus exbzus excaus exchus exdnus exhkus exinus exjpus exkous exmaus exmxus exnzus
   0.0    0.0    0.0    0.0   -0.7    0.0    0.0    0.0    0.0    0.0    0.0    0.0
exnous exsius exsfus exslus exsdus exszus extaus exthus exukus exvzus exeuus
   0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.7
```

[4] Fit lasso to the original covariates and describe how it differs from PCR here.

- The lasso on the original raw covariates finds a sparse model whereas PCR assumes a dense model where all x variables matter but only through the information provided through a few simple factors.
- It appears that, for this application, the individual exchange rate provides little value beyond what they tell us about underlying factors.

since you haven't simplified into linear factors
the estimation variance overwhelms any signal



Codes

```
setwd("C:/Users/13124/Desktop/Harris/2020-2 Spring/Big Data/week 8/")
fx <- read.csv("FXmonthly.csv")
fx <- (fx[2:120,]-fx[1:119,])/(fx[1:119,])
```

[1]

```
dim(fx)
par(mfrow=c(2,2))
pairs(fx[1:6])
pairs(fx[7:12])
pairs(fx[13:18])
pairs(fx[19:23])
```

[2]

```
par(mfrow=c(1,1))
(pc <- prcomp(fx, scale=TRUE)) #rotations aka what "defines" the PCs?
plot(pc, main="")
(pc2 <- prcomp(fx, scale=TRUE, rank = 2))
(z <- predict(pc))
round(pc$rotation[,1:3],1)
summary(pc)

pairs(pc)
pairs(pc[,1:6])
pairs(pc[,7:12])
pairs(pc[,13:18])
pairs(pc[,19:23])
```

[3]

```
sp<-read.csv("sp500.csv")
library(gamlr)

sp500 <- sp$sp500

zfx <- predict(pc)

zdf <- as.data.frame(zfx)

kfits <- lapply(1:23, #add one Pc each time
        function(K) glm(sp500~., data=zdf[,1:K,drop=FALSE]))

(aicc <- sapply(kfits, AICc)) # apply AICc to each fit
which.min(aicc) ## it likes 3 factors best
(bic <- sapply(kfits, BIC)) # apply BIC to each fit
which.min(bic) ## it likes 3 factors best

lassoPCR <- cv.gamlr(x=zfx, y=sp500, nfold=23)
## lasso.1se agrees with IC on first 3, then grabs a couple extra
coef(lassoPCR)

## plot
par(mfrow=c(1,2))
plot(aicc, pch=21, bg="maroon", xlab="K", ylab="AICc")
```

```
plot(bic, pch=21, bg="maroon", xlab="K", ylab="BIC")
plot(lassoPCR)

round(pc$rotation[,15],1)
round(pc$rotation[,17],1)
round(pc$rotation[,20],1)
round(pc$rotation[,21],1)
round(pc$rotation[,23],1)

## [4]
lasso <- cv.gamlr(x=as.matrix(fx), y=sp500, nfold=23)
par(mfrow=c(1,2))
plot(lasso, main="Regression on Original Covariates")
plot(lassoPCR, main="PCR")
## since you haven't simplified into linear factors
## the estimation variance overwhelms any signal
```