Homework 3

Joyce Yu Cahoon

6.3

Let $\boldsymbol{Y}_t$ be a vector of excess returns of N assets. Consider the multivariate linear regression model:

$$\boldsymbol{Y}_t = \alpha + \beta Y_t^m + \boldsymbol{\epsilon}_t$$

where $\boldsymbol{\epsilon}_t \sim N(0, \boldsymbol{\Sigma})$ and $cov(Y_t^m, \epsilon) = 0$.

1. Derive the MLE for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. You do not need to derive the MLE for $\boldsymbol{\Sigma}$ since this part is hard. You just take for granted that $\hat{\boldsymbol{\Sigma}}$ is the MLE.
2. Show that the MLRT for the null hypothesis H_0: $\boldsymbol{\alpha} = 0$ is:

$$T_2 = T[\log(|\hat{\Sigma}_0|) - \log(\hat{\Sigma})]$$

   where $\hat{\boldsymbol{\Sigma}}_0$ is the MLE under H_0. Give the expression for $\hat{\boldsymbol{\Sigma}}_0$.

6.4

Consider the multifactor model:

$$\boldsymbol{Y}_t = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{X}_t + \boldsymbol{\epsilon}_t$$

with observable factor $\boldsymbol{X}_t$ where $\mathbb{E}\boldsymbol{\epsilon}_t = 0$ and $cov(\boldsymbol{X}_t, \boldsymbol{\epsilon}_t) = 0$.

1. Based on the 20 stock portfolios over a period of 60 months on the 3 factors, it was computed that $|\hat{\boldsymbol{\Sigma}}_0| = 2.375$ and $|\hat{\boldsymbol{\Sigma}}| = 1.624$. Test if the multifactor model is consistent with the empirical data, $H_0 : \alpha = 0$.
2. Suppose that the beta's of the GE stock over the S&P500 index (X_1), the size effect X_2, and the book-to-market effect X_3 are respectively 1.3, 0.3, -0.4. Assume further that over the last 10 years the average risk-free interest is 4%, the average return of the S&P500 is 11%, the average difference of returns between the small large capitalization is 3%, and the average difference of returns between the high and low book-to-market is 2%, what is the expected return of the GE stock using the Fama-French model?

6.5

Consider the multi-factor model:

$$\boldsymbol{Y}_t = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{X}_t + \boldsymbol{\epsilon}_t$$

with observable factor $\boldsymbol{X}_t$.

1. Suppose that the CAPM holds and over the last five years, the average of the risk-free interest rate is 3.5% and the average return of the CRSP value-weighted index is 12.5%. If the market β of a stock (with respect to the index) is 1.3, what is the expected return of the stock?
2. Based on 15 stock portfolios over a period of 60 months regressed on five factors without knowing the risk-free interest rate, it is computed that $|\hat{\boldsymbol{\Sigma}}_0| = 2.425$ and $|\hat{\boldsymbol{\Sigma}}| = 1.742$. Test if the multifactor model is consistent with the empirical data, $H_0 : \alpha = 0$.

3. Suppose that the strict multi-factor model is correct so that var($\boldsymbol{\epsilon}$) = $\boldsymbol{\Sigma}_0$ is a diagonal matrix and $\boldsymbol{X}_t$ and $\boldsymbol{\epsilon}_t$ are uncorrelated. Show how to estimate the covariance matrix of $\boldsymbol{Y}$ based on the past T days' data:
$$\{(\boldsymbol{X}_t, \boldsymbol{Y}_t) : t = 1, \ldots, T\}$$

6.8

Use the Fama-French 100 portfolios in the last five years to construct 3 common factors via the PCA based on the correlation matrix. Report the variance explained by each principle components. Now, regress each of the Fama-French 100 portfolios on these 3 principal components and report the distribution (histogram) of the residual variances. Report also the distribution of the variance of these 100 portfolios over the same time period.

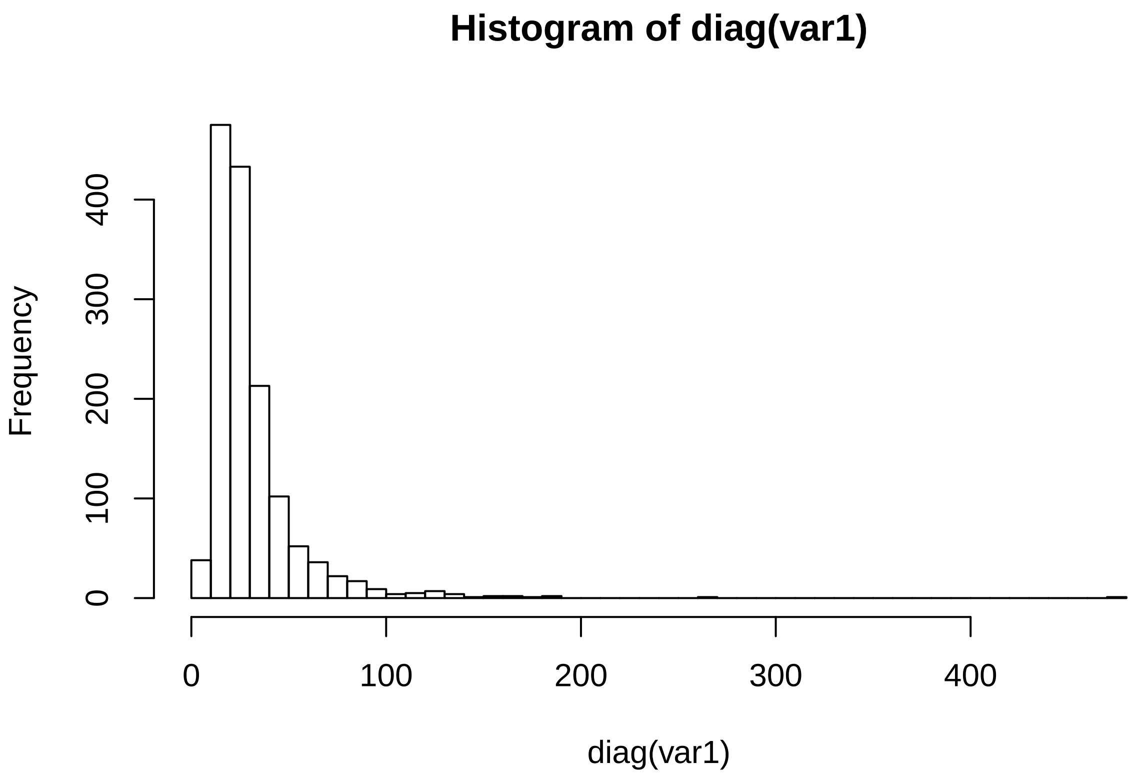
```
FF100 <- matrix(scan("~/workspace/st790-financial-stats/hw3/100_cleaned.txt"),
                ncol=101, byrow=T)
dates <- FF100[,1] # Time period used: 1/02/12 -- 08/31/18
# get SPX information from quantmod
# start <- as.Date("2013-01-02")
# end <- as.Date("2018-08-31")
# getSymbols("SPY", from = start, to = end) # SPY
# SPX <- SPY[,4]
# names(SPX) <- "spx"
# daily <- log(dailyReturn(SPX)+1)*100
# saveRDS(daily, "~/workspace/st790-financial-stats/hw3/spx.rds")
X <- readRDS("~/workspace/st790-financial-stats/hw3/spx.rds")
# match dates
D <- time(X)
D <- paste0(substr(D, 1, 4), substr(D, 6, 7), substr(D, 9, 10))
D <- as.numeric(D)
ind <- rep(0, length(D))
for(i in 1:length(ind)){
  ind[i] = (1:1427)[dates[i] == D]
}
dates <- dates[ind]
FF100 <- FF100[ind, ]

# PCA Analysis
Y <- FF100[,2:101]
resid <- resid(lsfit(X, Y)) # take the return of SP500 out
# cor(cbind(X,resid))
# variance explained of the first three principle components
p <- 100
eigen(cor(resid))$values[1:3]
```

```
## [1] 37.225922  8.688453  4.163726
```

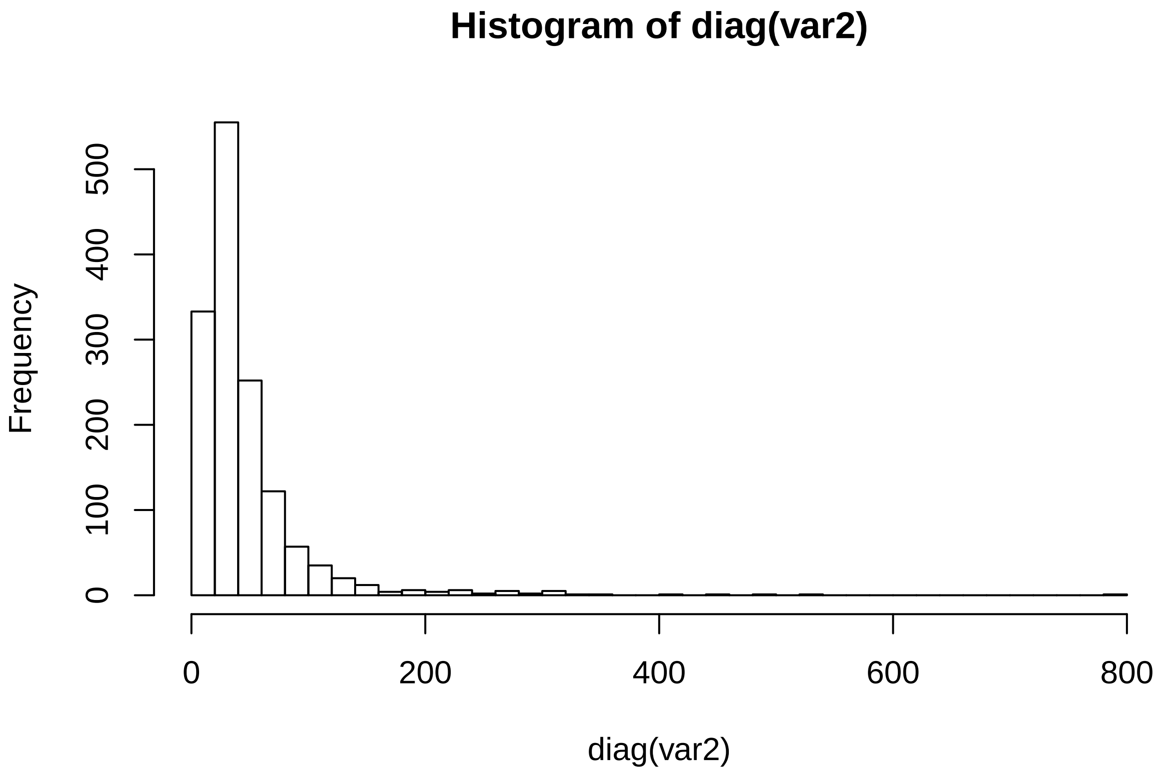
```
# proportion of total variability explained by the first 3 principle components
eigen(cor(resid))$values[1:3]/p*100
```

```
## [1] 37.225922  8.688453  4.163726
```

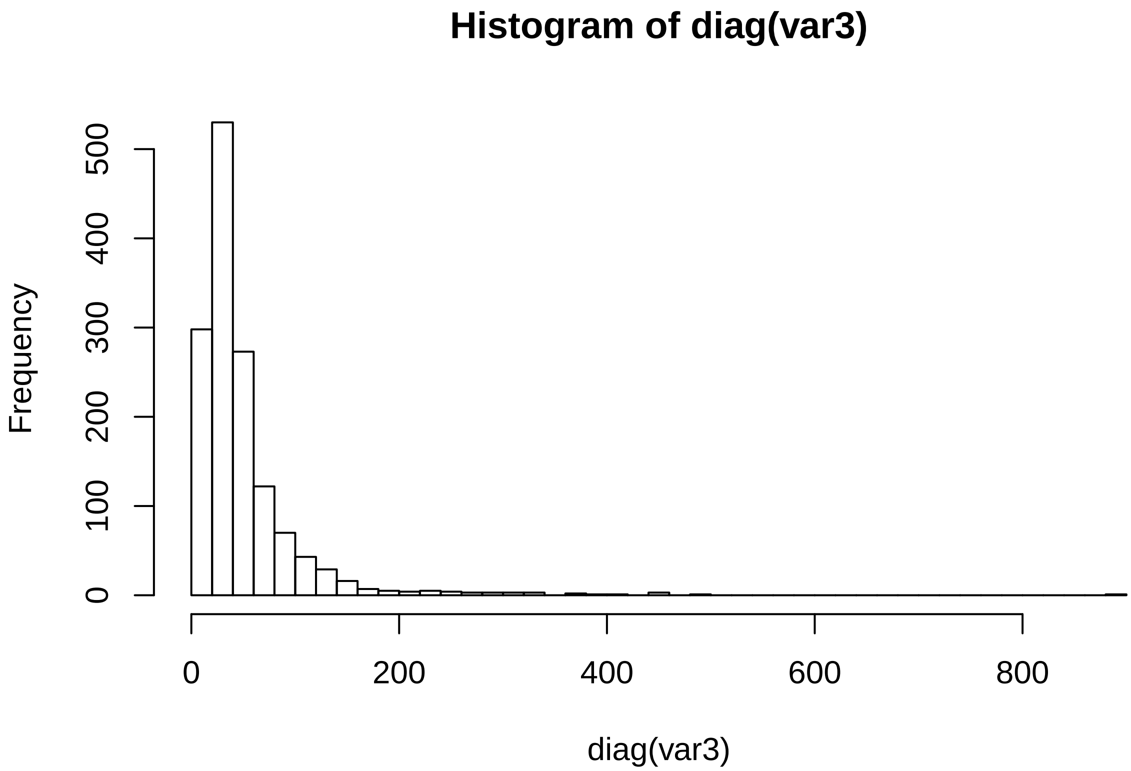
```r
# regress each port on these 3 principal components
a <- eigen(cor(resid))$vectors[,1]
a <- a / sqrt(apply(resid, 2, var))      # standardize the variables
Factor1 <- resid %*% a
resid1 <- resid(lsfit(cbind(X, Factor1), Y))
# get residual variances
var1 <- resid1 %*% t(resid1)
hist(diag(var1), breaks = 50)
```



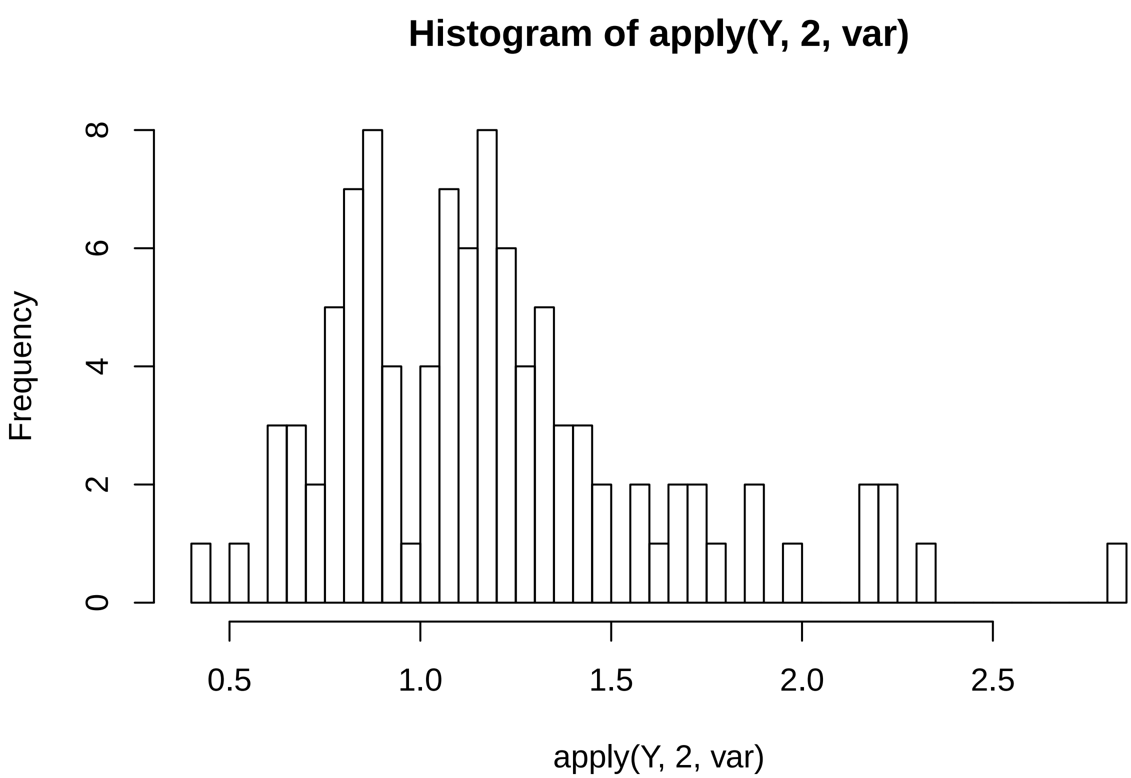

```r
# second principle component
a2 <- eigen(cor(resid))$vectors[,2]
a2 <- a2 / sqrt(apply(resid, 2, var))    # standardize the variables
Factor2 <- resid %*% a2
resid2 <- resid(lsfit(cbind(X, Factor2), Y))
# get residual variances
var2 <- resid2 %*% t(resid2)
hist(diag(var2), breaks = 50)
```




```r
# third principal components
a3 <- eigen(cor(resid))$vectors[,3]
a3 <- a3 / sqrt(apply(resid, 2, var))    # standardize the variables
Factor3 <- resid %*% a3
resid3 <- resid(lsfit(cbind(X, Factor3), Y))
# get residual variances
var3 <- resid3 %*% t(resid3)
hist(diag(var3), breaks = 50)
```




```
# dist of variance of these 100 portfolios
hist(apply(Y, 2, var), breaks = 50)
```



```r
# use Rsq
# Rsq1 <- 1-apply(resid1, 2, var)/apply(Y,2,var)
# hi st(Rsq1)
```