PS1

Issac Li

1/30/2017

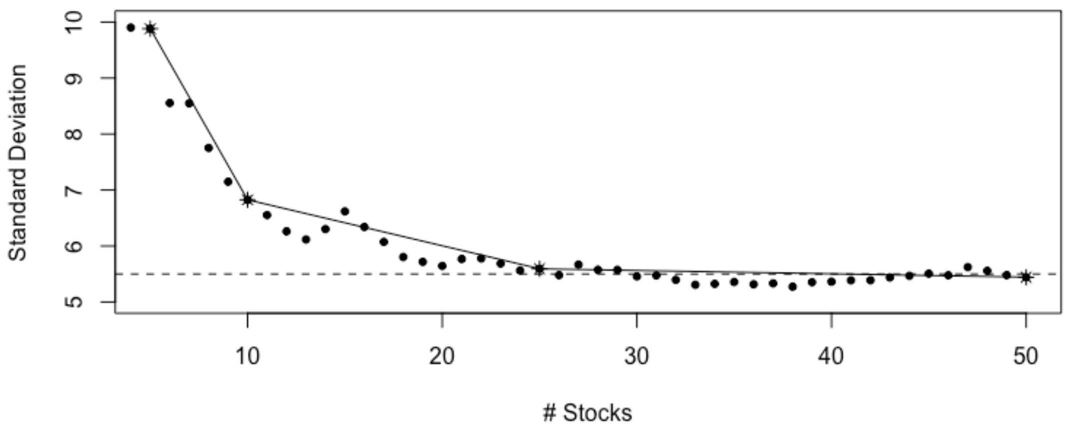
Part I

The first is a simple data exercise, which examines how the variances of portfolios comprised of randomly selected stocks are reduced as a result of diversification. There are also some simple sample statistics and regressions on the data you are asked to perform and interpret. In order to proceed, you need Microsoft Excel and the file "Problem_Set1_2017.xls", which can be downloaded from the course website.

a) Using the file `Problem_Set1_2017.xls`, form equal-weight portfolios using the first 5, first 10, first 25, and all 50 stocks. Calculate the sample mean and standard deviation of returns for each of the four equal-weight portfolios. Plot estimated standard deviations as a function of the number of stocks in the equal-weight portfolio. Comment on the shape of the function. Are the results consistent with what you would expect theoretically? Eye-balling the graph, does it look like adding more and more stocks will diversify away all of the standard deviation? Why or why not?

```
require(data.table)
require(ggplot2)
ps1=read.csv("/Users/lizhuo/Documents/MGT595/Problem_Set1_2017.txt",sep='\t',
fileEncoding="UTF-16LE",skip = 0,header = T)
ps1=ps1[,1:(58-6)]
colnames(ps1)[1]<-"Date"

# There is no missing value

returns=colMeans(ps1[,c(-1,-52)],na.rm = T)


results=matrix(NA,nrow=50,ncol=3)
colnames(results)<-c("Mean","SD","Var")

V=cov(ps1[,c(-1,-52)],use = "complete.obs")


for (i in seq(1,50)){
  w=rep(1/i,i)
  results[i,1]=i
  results[i,2]=w%*%returns[1:i]
  results[i,3]=sqrt(w%*%V[1:i,1:i]%*%w)
}
```

```
plot(results[c(5,10,25,50),1],results[c(5,10,25,50),3],type = "l",lty=1,ylab
="Standard Deviation",xlab="# Stocks",ylim=c(5,10))
abline(h=5.5,lty=2)
points(results[,1],results[,3],pch=20)
points(results[c(5,10,25,50),1],results[c(5,10,25,50),3],pch=8)
```



The function looks like the the first quadrant part of a/x, where a is a positive real number. The starred points are the 4 equally-weighted portfolios and they are connected by the solid line. The solid points at x correspond to SD of the portfolio of the first x stocks. The result is consistent with what I woudl expect theoretically. Adding more stocks will not reduce the SD to zero as the points converge to the dotted line. The reason is that the variance due to individual stocks are still present.

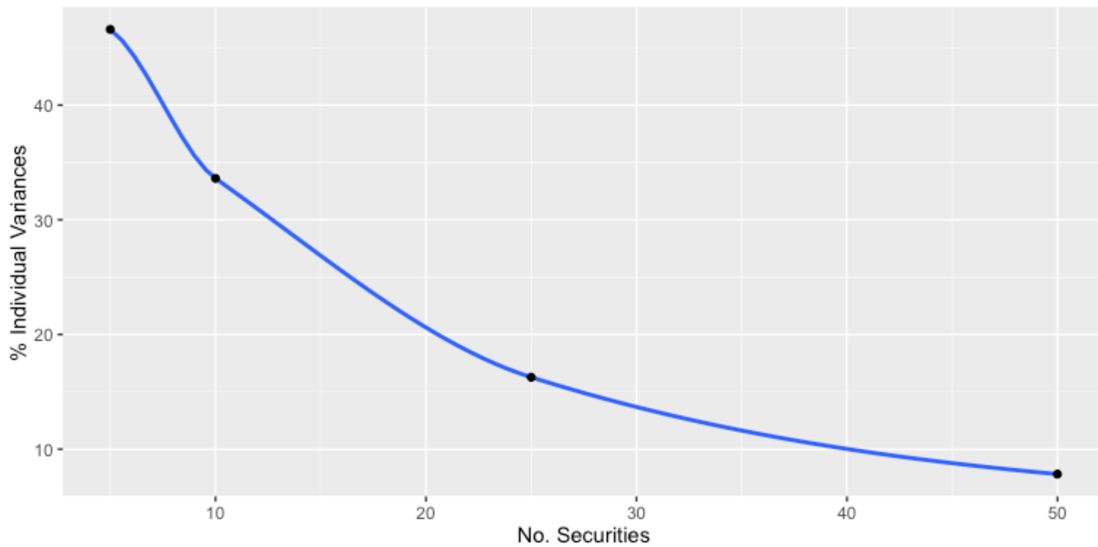
b) For all four equal-weight portfolios, decompose the estimated portfolio variance into its two components (the contributions of variances and covariances). Hint: you do not have to estimate the pairwise covariances in order to compute the decomposition. Plot the percentage of the portfolio's variance due to the variances of individual security returns as a function of the number of stocks in the portfolio. Comment on the shape of the function. Are the results consistent with what you would expect theoretically?

```
p4=results[c(5,10,25,50),]
table_1b=matrix(NA,nrow = 4,ncol = 4)
for (i in 1:4){
  n=p4[i,1]
  table_1b[i,1]=n
  table_1b[i,4]=(p4[i,3])^2
  table_1b[i,2]=(rep(1/n,n))^2%*%diag(V[1:n,1:n])
  table_1b[i,3]=table_1b[i,4]-table_1b[i,2]
}
colnames(table_1b)=c("#Stocks","Var","Cov","Tot.")
rownames(table_1b)=c(1,2,3,4)
x=table_1b[,"#Stocks"]
```

```r
y=(table_1b[,"Var"]/table_1b[,"Tot."]*100)
table_1b
```

```
##   #Stocks       Var      Cov     Tot.
## 1       5 45.507439 52.13603 97.64347
## 2      10 15.659422 30.92863 46.58806
## 3      25  5.089585 26.18864 31.27823
## 4      50  2.316261 27.29259 29.60885
```

```r
qplot(x,y, geom='smooth',xlab = "No. Securities",ylab="% Individual
Variances")+geom_point(data=data.frame(cbind(x,y)),aes(x,y))
```



The function ,again, looks like the the first quadrant part of a/x, where a is a positive real number. This makes sense because as the no of stocks increase, the variance due to covariance decreases. As the no of stocks approximate infinity, then the covariance approximates zero and then all the variance are due to individual variance.

c)

Suppose instead of equal-weighted portfolios, we computed value-weighted (e.g., weighted by market capitalization) portfolios. Would you expect the value-weighted portfolios to exhibit more or less variance relative to equal-weighted portfolios? What might it depend on? (Do not make any calculations here, just answer what it will depend on.)

Theoretically, the value-weighted portfolios are expected to exhibit less variance relative to equal-weight portfolios since 1) $\prod_i^n w_i$ is maximized when $w_1 = w_2 = w_3 ... = w_n$ and 2) securities with larger market value tend to be less volatile. However, in reality, it depends on the ratios of varinaces and market values of individual secruities. If stocks with larger market value have big variance then portfolios might have bigger variances if weighted by market value instead of equally.

d)

Compute the test statistics for whether the mean return of each of the four equal-weight portfolios you calculated in part b) is different from zero. What statistical distribution do these test statistics follow? Do you reject or fail to reject the null hypothesis that each of the mean returns on the portfolios is different from zero?

This test statistics follow the t-distribution, which has mean of 0.

```
p5=rowSums(1/5*ps1[,1:5],na.rm = T)
p10=rowSums(1/10*ps1[,1:10],na.rm = T)
p25=rowSums(1/25*ps1[,1:25],na.rm = T)
p50=rowSums(1/50*ps1[,1:50],na.rm = T)

t1=t.test(p5)
t2=t.test(p10)
t3=t.test(p25)
t4=t.test(p50)

table_1d=matrix(NA,nrow=4,ncol=2)
table_1d[1,]=c(t1$statistic,t1$p.value)
table_1d[2,]=c(t2$statistic,t2$p.value)
table_1d[3,]=c(t3$statistic,t3$p.value)
table_1d[4,]=c(t4$statistic,t4$p.value)

colnames(table_1d)<-c("T-value","P-value")
rownames(table_1d)<-c("P5","P10","P25","P50")
table_1d

##      T-value P-value
## P5  6177.850       0
## P10 6140.570       0
## P25 5972.864       0
## P50 5309.974       0
```

Since all of the p-values are smaller than 0.05, so we can reject the null hypothese that the true mean is equal to 0, meaning thetue mean is not equal to 0 by one-sample t-test.

e)

Choose the first stock and test whether its returns follow a normal distribution. Repeat this test for the equal-weighted portfolio of all 50 stocks as well as the market portfolio index of all NYSE, AMEX, and Nasdaq stocks (which is also included in the spreadsheet).

```
require(moments)
mkt=ps1$Market..Value.Weighted.Index.

norm_test<-function(x){
n=NROW(x)
skewness(x)
```

```
ses=sqrt(6*n*(n-1)/((n-2)*(n+1)*(n+3)))
pchisq(skewness(x)/ses,df=2)

sek=2*ses*sqrt((n^2-1)/(n-3)*(n+5))
c((kurtosis(x))^2+(skewness(x)/ses)^2,pchisq((kurtosis(x))^2+(skewness(x)/ses
)^2,df=2,lower.tail = F))
}

p1=ps1[,1]
p50=rowSums(1/50*ps1,na.rm = T)

ans1=norm_test(p1)
ans2=norm_test(p50)
ans3=norm_test(mkt)

paste("One stock: Chi-square test statistics is ",round(ans1[1],3)," with 2
degrees of freedom, yielding a p-value of ",round(ans1[2],6))

## [1] "One stock: Chi-square test statistics is  3.202  with 2 degrees of
freedom, yielding a p-value of  0.201685"

paste("All stocks: Chi-square test statistics is ",round(ans2[1],3)," with 2
degrees of freedom, yielding a p-value of ",round(ans2[2],8))

## [1] "All stocks: Chi-square test statistics is  8.159  with 2 degrees of
freedom, yielding a p-value of  0.01691563"

paste("Market index:Chi-square test statistics is ",round(ans3[1],3)," with 2
degrees of freedom, yielding a p-value of ",round(ans3[2],8))

## [1] "Market index:Chi-square test statistics is  32.298  with 2 degrees of
freedom, yielding a p-value of  1e-07"
```

The p-values of the Chi-square tests on skewness and kurtosis for all three portfolios are << 0.05, meaning that we can reject the null hypothesis that those come from normal distribution. These returns are not normally distributed.

f)

Regress the returns of the first ten stocks in the spreadsheet (TXN through VMC) on the value weighted market index (a cap-weighted index of all NYSE, AMEX, and Nasdaq stocks) return, and report intercepts ( ) and slope coefficients ( ), in a table.

1. Interpret the slope coefficients of this model.

2. Interpret the intercepts of the model, in terms of the effect of the market return on the securities' returns.
3. What does the R2 of the regressions tell you? What does it mean when the R2 is low in this case?

The intercept and coefficients are:

```
reg_data=cbind(ps1[,2:11],mkt)
table_f1=data.frame(matrix(NA,nrow = 10,ncol = 2,dimnames =
list(colnames(ps1[,2:11]),c("Intercept","Coefficients"))))
for(i in 1:10){
fit=lm(data=reg_data,formula =
as.formula(paste(colnames(reg_data)[i],"~mkt")))
table_f1$Intercept[i]=fit$coefficients[1]
table_f1$Coefficients[i]=fit$coefficients[2]
}
round(table_f1,4)
```

```
##      Intercept Coefficients
## TXN    -0.0987       1.3995
## ISRG    2.5118       1.3993
## CIEN   -1.4257       2.5154
## BHI    -0.0037       1.2665
## WDC     2.1986       1.5840
## CAH     0.3354       0.6487
## CCI     0.7690       1.5545
## BBT     0.1996       0.7211
## TAP     0.5382       0.7239
## VMC     0.3335       1.1185
```

1.

The slope coefficients (beta) in these regressions are equivalent to covariances divided by variances. The interpretation of this slope coefficient is for a one unit movement in the market return (1 percent), the return of TXN is expected to increase by 1.4 percent.

2.

The interpretation of the intercepts (alpha) in these regressions, with respect to the market return, is that, for instance for TXN, when the market has zero percent return for a given month or if we take away the effect of the market return, the expected return of TXN for that same month is -0.099 percent. ##### 3. R^2 of regression tell us how much of variance is explained by the model. This is a measure of goodness of regression fit. Thus, for instance, for TXN, the amount of variation in the monthly return of TXN that is explained by the variation in the monthly market return. The relatively low 2 for most of these regressions indicates that there is a large amount of variation in the monthly return of these securities that is not explained by the variation in the monthly return of the market.