Lawrence Ouyang
504128219
Math 191
Homework 1

1a.

$$\begin{aligned}
var(x+y) &= E[(x+y)^2] - (E[x+y])^2 \\
&= E[x^2+2xy+y^2] - (E[x]+E[y])^2 \\
&= E[x^2] + E[2xy] + E[y^2] - E[x]^2 - 2E[x]E[y] - E[y]^2 \\
&= (E[x^2] - E[x]^2) + (E[y^2] - E[y]^2) + (2E[xy] - 2E[x]E[y]) \\
&= var(x) + var(y) + 2cov(x,y)
\end{aligned}$$

1b.

$$\begin{aligned}
\rho = cor(x,y) &= \frac{cov(x,y)}{\sigma_x \sigma_y} \\
&= \frac{cov(x,y)}{\sqrt{var(x)var(y)}} \\
&= \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}))(\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}))}} \qquad (1)
\end{aligned}$$

$$using\ Cauchy-Schwartz\ Inequality:$$

$$\begin{aligned}
(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}))^2 &\leq \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}) \\
|\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})| &\leq \sqrt{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}))(\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}))}
\end{aligned}$$

$$applying\ to\ (1):$$

$$|\frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}))(\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}))}}| \leq \frac{\sqrt{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}))(\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}))}}{\sqrt{(\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x}))(\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y}))}} = 1$$

$$-1 \leq \rho \leq 1$$

1c. For $\rho = -1$, this means that the points $(x_1, y_1), \ldots, (x_n, y_n)$ is perfectly negatively linear. In other words, the points $(x_1, y_1), \ldots, (x_n, y_n)$ perfectly fit a decreasing linear function.

2a. Since X_1 and X_2 are standard normal random variables, they have the special condition that

$$\mathbb{E}[X_1] = \mathbb{E}[X_2] = \mu_1 = \mu_2 = 0$$

$$\begin{aligned}
\mathbb{E}[Y_1] &= \mathbb{E}[3X_1 + X_2] \\
&= \mathbb{E}[3X_1] + \mathbb{E}[X_2] \\
&= 3\mathbb{E}[X_1] + \mathbb{E}[X_2] \\
&= 0
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[Y_2] &= \mathbb{E}[X_1 - X_2] \\
&= \mathbb{E}[X_1] - \mathbb{E}[X_2] \\
&= 0
\end{aligned}$$

2b. From (1a):

$$\begin{aligned}
Cov(Y_1, Y_2) &= \mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2] \\
&= \mathbb{E}[(3X_1 + X_2)(X_1 - X_2)] \\
&= \mathbb{E}[3X_1^2 - 2X_1X_2 - X_2^2] \\
&= 3\mathbb{E}[X_1^2] - 2\mathbb{E}[X_1X_2] - \mathbb{E}[X_2^2]
\end{aligned}$$

Since X_1 and X_2 are independent, then $\mathbb{E}[X_1X_2] = \mathbb{E}[X_1]\mathbb{E}[X_2]$.

$$\begin{aligned}
Cov(Y_1, Y_2) &= 3\mathbb{E}[X_1^2] - 2\mathbb{E}[X_1]\mathbb{E}[X_2] - \mathbb{E}[X_2^2] \\
&= 3\mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] - \mathbb{E}[(X_2 - \mathbb{E}[X_2])^2] \\
&= 3(\mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2) - \mathbb{E}[X_2^2] + \mathbb{E}[X_2]^2 \\
&= 3 - 1 = 2
\end{aligned}$$

2c. The probability density function of a standard normal random variable X is:

$$f_X(X) = \tfrac{1}{\sqrt{2\pi}} e^{(-\frac{1}{2}X^2)}$$

Thus:

$$\begin{aligned}
f(Y_1, Y_2) &= f(Y_1)f(Y_2) \\
&= \frac{1}{2\pi} e^{-\frac{1}{2}(Y_1^2 - Y_2^2)}
\end{aligned}$$

3. Since (Y_1, Y_2) are joint normal, we can assume that if the Pearson correlation coefficient is 0, then Y_1 and Y_2 are independent.

$$\begin{aligned}\rho_{y_1,y_2} = cor(Y_1, Y_2) &= \frac{cov(X_Y)}{\sigma_{Y_1}\sigma_{Y_2}} \\ &= \frac{\mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2]}{\sigma_{Y_1}\sigma_{Y_2}}\end{aligned}$$

For the correlation to be 0, we must show that the following is true:

$$\mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2] = 0$$

$$\begin{aligned}\mathbb{E}[Y_1Y_2] - \mathbb{E}[Y_1]\mathbb{E}[Y_2] &= \mathbb{E}[(X_1 - X_2)(X_1 + X_2)] - \mathbb{E}[X_1 - X_2]\mathbb{E}[X_1 + X_2] \\ &= \mathbb{E}[(X_1^2 - X_2^2)] - (\mathbb{E}[X_1] - \mathbb{E}[X_2])(\mathbb{E}[X_1] + \mathbb{E}[X_2]) \\ &= \mathbb{E}[X_1^2] - \mathbb{E}[X_2^2] - \mathbb{E}[X_1]^2 + \mathbb{E}[X_2]^2 \\ &= \mathbb{E}[X_1 - \mathbb{E}[X_1]^2] + \mathbb{E}[X_1]^2 - \mathbb{E}[X_2 - \mathbb{E}[X_2]^2] - \mathbb{E}[X_2]^2 - \mathbb{E}[X_1]^2 + \mathbb{E}[X_2]^2 \\ &= \mathbb{E}[(X_1 - \mathbb{E}[X_1])^2] - \mathbb{E}[(X_2 - \mathbb{E}[X_2])^2] \\ &= \mathbb{E}[X_1^2] - \mathbb{E}[X_1]^2 - \mathbb{E}[X_2^2] + \mathbb{E}[X_2]^2 \\ &= 1 - 1 = 0\end{aligned}$$

4a. Take $h(t) = \mathbb{E}[(X + tY)^2]$

$$\begin{aligned}h(t) &= \mathbb{E}[(X + tY)^2] \\ &= \mathbb{E}[X^2 + 2XtY + t^2Y^2] \\ &= \mathbb{E}[X^2] + \mathbb{E}[2XtY] + \mathbb{E}[t^2Y^2] \\ &= t^2\mathbb{E}[Y^2] + 2t\mathbb{E}[XY] + \mathbb{E}[X^2]\end{aligned}$$

The discriminant of this polynomial is:

$$4\mathbb{E}[XY]^2 - 4E[Y^2]E[X^2]$$

We can note that the sign of h is positive and thus its vertex must be greater than 0, so we know that the roots of this polynomial must be imaginary and thus:

$$\begin{aligned}4\mathbb{E}[XY]^2 - 4E[Y^2]E[X^2] &\leq 0 \\ \mathbb{E}[XY]^2 - E[Y^2]E[X^2] &\leq 0 \\ \mathbb{E}[XY]^2 &\leq E[Y^2]E[X^2]\end{aligned}$$

4b. Let us take our same function $h(t) = \mathbb{E}[(X + tY)^2]$ but have $X \to X - \bar{X}$ and $Y \to Y - \bar{Y}$.

$$\begin{aligned} h(t) &= \mathbb{E}[(X-\bar{X})^2 + 2(X-\bar{X})t(Y-\bar{Y}) + t^2(Y-\bar{Y})^2] \\ &= var(X) + 2tcov(X,Y) + t^2 var(Y) \end{aligned}$$

Since $var(X) \geq 0$, then $var(X + tY) \geq 0$, then $h(t) \geq 0$.
This tells us that there can only be either 1 root or 2 imaginary roots, thus the discriminant of this quadratic must follow the condition:

$$\begin{aligned} (2cov(X,Y))^2 - 4var(X)var(Y) &\leq 0 \\ 4cov(X,Y)^2 &\leq 4var(X)var(Y) \\ |cov(X,Y)| &\leq \sqrt{var(X)var(Y)} \\ \frac{|cov(X,Y|)}{\sqrt{var(X)var(Y)}} &\leq 1 \\ |\frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}| &\leq 1 \\ -1 \leq \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}} &\leq 1 \\ -1 \leq \rho &\leq 1 \end{aligned}$$

Programming Assignment:

```
#1. The histograms are attached.
#Create Histograms
pdf("Ticker_Histograms.pdf", width = 8.5, height = 11);
par(mfrow = c(4,3))
for (i in 1:nrTickers) {
   hist(RET[,i], main = tickers[i], xlab = 'Log Return', col = "red");
}
dev.off()

#2. Collect the means and variance into a vector SampleMean and SampleVar:

SampleMean = matrix(, nrow = nrTickers, ncol = 1);
SampleVar = matrix(, nrow = nrTickers, ncol = 1);

for (i in 1:nrTickers) {
   SampleMean[i] = mean(RET[,i]);
   SampleVar[i] = var(RET[,i]);
}

colnames(SampleMean) = c("SM of Log Return");
rownames(SampleMean) = tickers;
colnames(SampleVar) = c("SV of Log Return");
rownames(SampleVar) = tickers;

#install.packages('gplots');
```

```
#library('gplots');
pdf("SampleMeanVar,pdf", width = 8.5, height = 11);
par(mfrow = c(4,3));
textplot(SampleMean);
textplot(SampleVar);
dev.off();

#Other Correlation Measurements (put in test_correlations())
#Maximal
#install.packages('acepack');
Max_cor = matrix(, nrow = nrTickers, ncol = nrTickers);
colnames(Max_cor) = tickers;
rownames(Max_cor) = tickers;

for (i in 1:nrTickers) {
   for(j in 1:nrTickers) {
      transfVars = ace(RET[,i],RET[,j]);
      Max_cor[i,j] = cor(transfVars$tx,transfVars$ty)[1];
      Max_cor[i,j] = round(Max_cor[i,j],3);
   }
}
print('Max_cor:'); print(Max_cor);
list_CorMats[['Maximal']] = Max_cor;

#Hoeffding's D
#library('Hmisc');
H_cor = hoeffd(RET)$D;
H_cor = round(H_cor,3);
print('Hoeffding\'s D:'); print(H_cor);
list_CorMats[['Hoeffding']] = H_cor;

#Distance
D_cor = matrix(, nrow = nrTickers, ncol = nrTickers);
colnames(D_cor) = tickers;
rownames(D_cor) = tickers;

#library('energy');
for (i in 1:nrTickers) {
   for(j in 1:nrTickers) {
      D_cor[i,j] = dcor(RET[,i],RET[,j]);
      D_cor[i,j] = round(D_cor[i,j],3);
   }
}
print('dCor:'); print(D_cor);
list_CorMats[['dCor']] = D_cor;

#MIC
#library('minerva');
MIC_cor = mine(RET,n.cores = 4)$MIC
```

```
MIC_cor = round(MIC_cor,3);
print('MIC_cor:'); print (MIC_cor);
list_CorMats[['MIC']] = MIC_cor;


#3. Modified test_correlations() to have a parameter Instr for the tickers being processed.
Pearson, Spearman, Maximal, Distance, and MIC all approximate to 1, indicating that they are
trivial relationships, wtih a near perfect linear relationship. We can conclude that
SPXS is the leveraged inverse ETF(short/bear) and SPXL is the leveraged ETF(long/bull).
From our results, we see that neither of these ETFs are earning or losing the expected
increase/decrease from being leveraged ETFs, and they are following SPY almost perfectly.

#4. The figures are attached.

#5. Strongest Positive Correlations:
   GLD - GOLD
   SPY - ^FTSE
   SPY - OIL

    Strongest Negative Correlations:
   SPY - ^VIX
    ^VIX - ^TNX
    ^VIX - ^FTSE

#6. A lagging indicator due to the extremely large dataset can be a reason why certain
correlations are so weak. The moving average is widely distributed giving inaccurate points.
With so many weak correlations in our figures, including a lagging variable would make
our points more accurate and help predict future returns. We can also note the high chance
of causation with certain correlations. An increase in a certain area does often cause
    certain
stocks to increase or decrease.

#7. The figure is attached.
```