STAT3005 Assignment 3
Lam Hoi Chun (1155192755)

Exercise 3.1

Note that Chatterjee's Correlation is a strong correlation, so

$$\rho_C = 0 \iff X \perp\!\!\!\perp Y$$

Therefore, we need to prove that $X \perp\!\!\!\perp Y \Rightarrow \rho_S = 0$, where

$$\rho_S = \frac{\text{Cov}(R_X, R_Y)}{\sqrt{\text{Var}(R_X) \cdot \text{Var}(R_Y)}}$$

Consider

$$\begin{aligned}
\text{Cov}(R_X, R_Y) &= \text{E}\left\{ \left[R_X - \text{E}(R_X)\right]\left[R_Y - \text{E}(R_Y)\right] \right\} \\
&= \text{E}\left\{ \left(R_X - \frac{n-1}{2}\right)\left(R_Y - \frac{n-1}{2}\right) \right\} \\
&= \text{E}(R_X R_Y) - \frac{n-1}{2} \cdot \text{E}(R_X) - \frac{n-1}{2} \cdot \text{E}(R_Y) + \left(\frac{n-1}{2}\right)^2 \\
&= \text{E}(R_X) \cdot \text{E}(R_Y) - \frac{n-1}{2} \cdot \text{E}(R_X) - \frac{n-1}{2} \cdot \text{E}(R_Y) + \left(\frac{n-1}{2}\right)^2 \quad \because X \perp\!\!\!\perp Y \Rightarrow R_X \perp\!\!\!\perp R_Y \\
&= \left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}\right) - \left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}\right) - \left(\frac{n-1}{2}\right)\left(\frac{n-1}{2}\right) + \left(\frac{n-1}{2}\right)^2 \\
&= 0
\end{aligned}$$

So, it is proved that $X \perp\!\!\!\perp Y \Rightarrow \rho_S = 0$, i.e. it is not possible to have $\rho_S \neq 0$ and $\rho_C = 0$

Exercise 3.2

Ex 3.2 1(a).

We can compute the rates of return and the and number of duplicated value from $r_2, \ldots, r_n$ using the following R code:

```
r = head(diff(s)/lag(s), -1)
sum(duplicated(r))
```

It will return the number of duplicated value = 57, i.e. ties exist.

Ex 3.2 1(b).

The required trace plots of S_i and r_i can be reproduced using the following R code:

```
par(mfrow=c(1,2))
ts.plot(s, main="Stock Price", xlab=expression(i),
        ylab=expression(S[i]), col="blue")
ts.plot(r, main="Rate of Return", xlab=expression(i),
        ylab=expression(r[i]), col="blue")
```

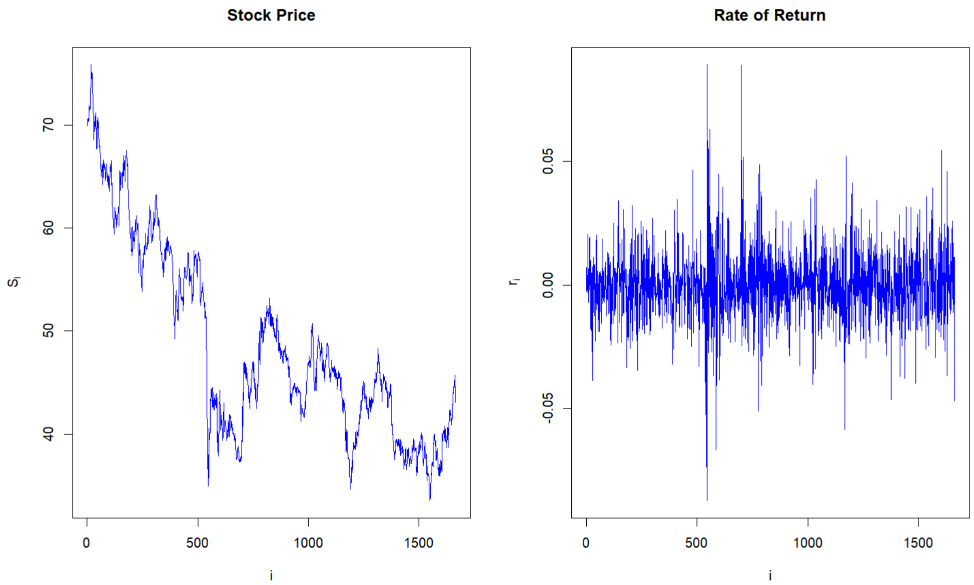

Figure 1: Trace Plots of s_i and r_i

Ex 3.2 1(c).

The required autocorrelation plots of S_i and r_i can be reproduced using the following R code:

```
par(mfrow=c(1,2))
acf(r, main=expression("ACF of {"*r[i]*"}"), col="red")
acf(r^2, main=expression("ACF of {"*r[i]^2*"}"), col="red")
```

[Pic]

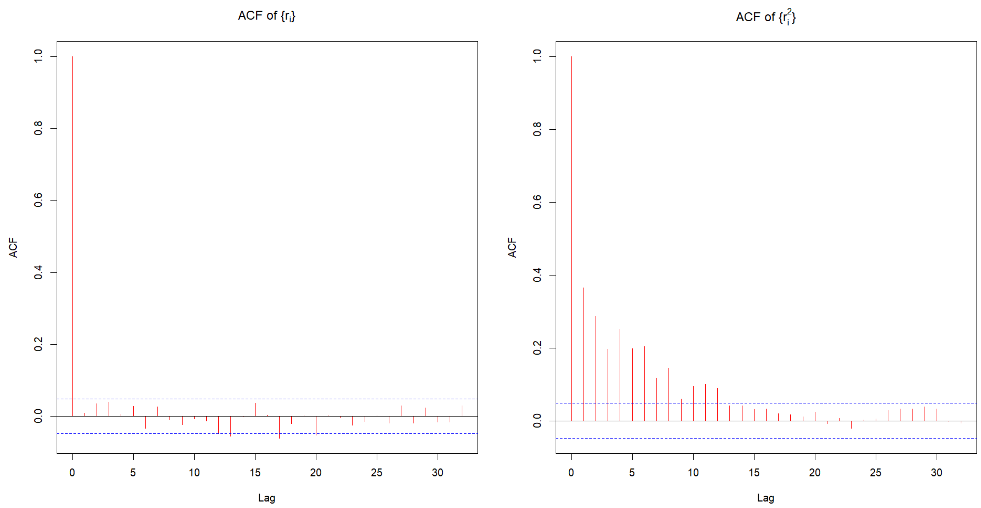

Figure 2: ACF Plots of s_i and r_i

Ex 3.2 2(a-c).

The estimates, p-value for testing independence and the computation time for each testing procedures can be computed using the following R code:

```r
# Estimates
for(i in 1:3){
    result[1,i] = cor(x,y,method=c("p","s","k")[i])
}
result[1,4] = cov.bd0(x,y)/(cov.bd0(x,x)*cov.bd0(y,y))
result[1,5] = cor.c1(x,y)

# p-value and computation time
for(i in 1:5){
    t0 = Sys.time()
    result[2,i] = cor.test0(x,y,method=c("p","s","k","b","c")[i])
    t1 = Sys.time()
    result[3,i] = difftime(t1, t0, units="secs")
}
```

All the values below are rounded off to the 5 decimal places from the R output.

	ρ_{P}	ρ_{S}	ρ_{K}	ρ_{BD}	ρ_{C}
(a) Estimate of correlation	0.00929	0.03032	0.02128	0.00055	0.04560
(b) p-value for testing independence	0.70494	0.21653	0.19395	0.22100	0.00317
(c) Time taken for testing	0.00090	0.00104	0.02347	8.89625	0.03622

Ex 3.2 (3)

I prefer the Chatterjee's correlation because

- (Assumption of the test) It does not assume the data to be linear or monotonic, which allows it to be applied to general, i.e. oscillating trend. Importantly, from the trace plot in 1(a), it can be seen that the rate of return of the stock had a fluctuating trend. So, Chatterjee's correlation would be more accurate.

- (Computation time) Although Bergsma–Dassios' correlation can also be used for testing non-linear and non-monotonic data, Chatterjee's correlation is still a better choice since it is more efficient in computation. To be precise, in this case, Chatterjee's correlation test is more than 200 times faster than Bergsma–Dassios'correlation test. We can also forsee that when the sample size increases, the difference in efficiency of these 2 tests will be much larger, as their complexities are in $O(n)$ and $O(n^4)$ respectively.

Ex 3.2 4(a)

The residuals $\widehat{\varepsilon}_i$ and the trace plot of $\widehat{\varepsilon}_i$ can be computed and produced respectively using the following R code:

```
residuals = tseries::garch(r, order=c(1,1), trace=0)$residuals[-1]
ts.plot(residuals, main="Residuals Plot", xlab=expression(i),
        ylab=expression(epsilon[i]), col="blue")
```

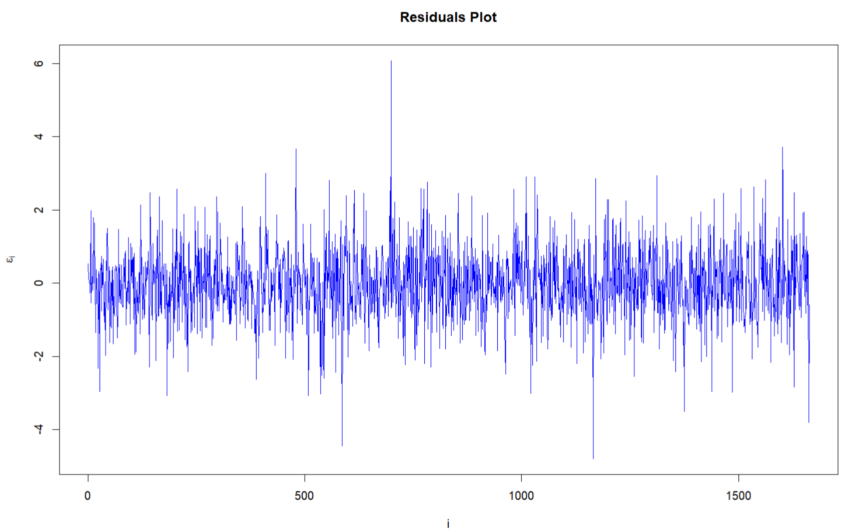

Figure 3: Residuals Plot

Ex 3.2 4(b)

Whether $\widehat{\varepsilon}_i$ are correlated at lag one by Chatterjee's correlation test will be tested using the following R code:

```
residuals_x = r[-1]
residuals_y = r[-length(r)]
cor.test.c1(residuals_x,residuals_y) # p-value = 0.0113211

rep.cor.test.c0 <- function(x,y,nRep){
  p_value <- numeric(0)
  for (iRep in 1:nRep){
    p_value <- append(p_value, cor.test.c0(x,y))
  }
  mean(p_value)
}
rep.cor.test.c0(residuals_x, residuals_y, 10000) # p-value = 0.01403572
```

Note that the value of sample Chatterjee's correlation varies because the XICOR package handles tied observations by rank(x, ties.method="random"). Therefore, I repeated the test for 10000 times to minimise the effect of the randomness in choosing the method handling tied observations.

Note that $H_0 : \rho_C = 0$ and $H_1 : \rho_C \neq 0$.

Since both tests' p-values $< \alpha = 0.05$. So we reject H_0, meaning that we cannot conclude the residuals are uncorrelated.

Ex 3.3 (1).

First, we let A_1, A_2 and A_3 be IID copies of A.

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{Cov}\big[f(A), g(A)\big] \\
&= \mathrm{E}\big[f(A)g(A)\big] - \mathrm{E}\big[f(A)\big]\mathrm{E}\big[g(A)\big] \\
&= \mathrm{E}\big[f(A_1)g(A_1)\big] - \mathrm{E}\big[f(A_1)\big]\mathrm{E}\big[g(A_2)\big] \\
&= \mathrm{E}\big[f(A_1)\big\{g(A_1) - g(A_2)\big\}\big] \\
&= \mathrm{E}\big[f(A_2)\big\{g(A_2) - g(A_1)\big\}\big] \\
&= \mathrm{E}\big[-f(A_2)\big\{g(A_1) - g(A_2)\big\}\big] \\
&= \frac{1}{2}\cdot \mathrm{E}\big[f(A_1)\big\{g(A_1) - g(A_2)\big\} - f(A_2)\big\{g(A_1) - g(A_2)\big\}\big] \\
&= \frac{1}{2}\cdot \mathrm{E}\big[\big\{f(A_1) - f(A_2)\big\}\big\{g(A_1) - g(A_2)\big\}\big]
\end{aligned}$$

As f and g are both increasing functions:

$$\begin{aligned}
&\text{If } A_1 > A_2 \Rightarrow f(A_1) - f(A_2) \geq 0 \text{ and } g(A_1) - g(A_2) \geq 0 \\
&\text{If } A_1 < A_2 \Rightarrow f(A_1) - f(A_2) \leq 0 \text{ and } g(A_1) - g(A_2) \leq 0
\end{aligned}$$

and their product will still be

$$\begin{aligned}
\big\{f(A_1) - f(A_2)\big\}\big\{g(A_1) - g(A_2)\big\} &\geq 0 \\
\therefore \mathrm{Cov}(X,Y) &\geq 0
\end{aligned}$$

Recall that

$$\rho_{\mathrm{P}} = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

where

$$\begin{aligned}
\mathrm{Cov}(X,Y) \geq 0, \sigma_X \geq 0 \text{ and } \sigma_Y &\geq 0 \\
\therefore \rho_{\mathrm{P}} &\geq 0
\end{aligned}$$

Ex 3.3 (2).

First, we let A_1, A_2 and A_3 be IID copies of A. By the definition of Spearman correlation, we have

$$\begin{aligned}\rho_{\mathrm{S}} &= 3 \cdot \mathrm{E}\big[\mathrm{sign}\big\{(X_1 - X_2)(Y_1 - Y_3)\big\}\big] \\ &= 3 \cdot \mathrm{E}\big[\mathrm{sign}\big\{(f(A_1) - f(A_2))(g(A_1) - g(A_3))\big\}\big]\end{aligned}$$

Ex 3.3 (3).

By the definition of Kendall correlation, we have

$$\begin{aligned}\rho_{\mathrm{K}} &= \mathrm{E}\big[\mathrm{sign}\big\{(X_1 - X_2)(Y_1 - Y_2)\big\}\big] \\ &= \mathrm{E}\big[\mathrm{sign}\big\{(f(A_1) - f(A_2))(g(A_1) - g(A_2))\big\}\big]\end{aligned}$$

Using the idea from (a),

$$\begin{aligned}\big\{f(A_1) - f(A_2)\big\}\big\{g(A_1) - g(A_2)\big\} &> 0 \\ \mathrm{sign}\big\{(f(A_1) - f(A_2))(g(A_1) - g(A_2))\big\} &= 1 \\ \mathrm{E}\big[\mathrm{sign}\big\{(f(A_1) - f(A_2))(g(A_1) - g(A_2))\big\}\big] &= 1 \\ \rho_{\mathrm{K}} &= 1\end{aligned}$$

Ex 3.3 (5).

We can compute the results using the following R code:

```
n.all = c(5, 50 ,500)
test_method = c("p", "s", "k", "c")
out = array(dim = c(length(n.all), length(test_method)),
            dimnames = list(paste0("n = ", n.all), test_method))

for(i in 1:length(test_method)){
  for(i.all in 1:length(n.all)){
    A = rexp(n.all[i.all], 1)
    X = A^3
    Y = exp(A)
    out[i.all, i] = cor0(X, Y, method=test_method[i])
  }
}
out
```

The output of the R code is:

	$\widehat{\rho_{\mathrm{P}}}$	$\widehat{\rho_{\mathrm{S}}}$	$\widehat{\rho_{\mathrm{K}}}$	$\widehat{\rho_C}$
$n = 5$	0.9997222	1	1	0.5000000
$n = 50$	0.9976577	1	1	0.9411765
$n = 500$	0.9077994	1	1	0.9940120

Person correlation is close to the theoretical value when n is small and then decrease, which may be due to the data is not linear. Spearman and Kendall correlations are close to 1, showing a strong monotonic relationship. Chatterjee correlation increases when n increases and close to 1.