DATA 405 Assignment 4

Rin Meng 51940633

Question 1

The binomial probability mass function (PMF) for the number successes k (insect surviving) is given by

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

where n is the number of trials (insect released), p is the probability of success (insect surviving).

Applying the Maximum Likelihood Estimate (MLE) to the binomial PMF, we have

$$L(p) = \binom{100}{56} p^{56} (1-p) 1 - p^{44}$$

Now we minimize the log-likelihood function $l(p) = \log L(p)$ to find the MLE of p.

$$l(p) = \log \binom{100}{56} + 56 \log p + 44 \log(1-p)$$

$$\frac{\partial l}{\partial p} = \frac{56}{p} - \frac{44}{1-p} = 0 \Leftrightarrow \frac{56}{p} = \frac{44}{1-p}$$

$$56 - 56p = 44p \Leftrightarrow 56 = 100p$$

$$p = 0.56$$

Therefore, the maximum likelihood estimate for the probability that an insect survives is p is 0.56.

Now we find the probability that more than 50 insects survive in a new experiment.

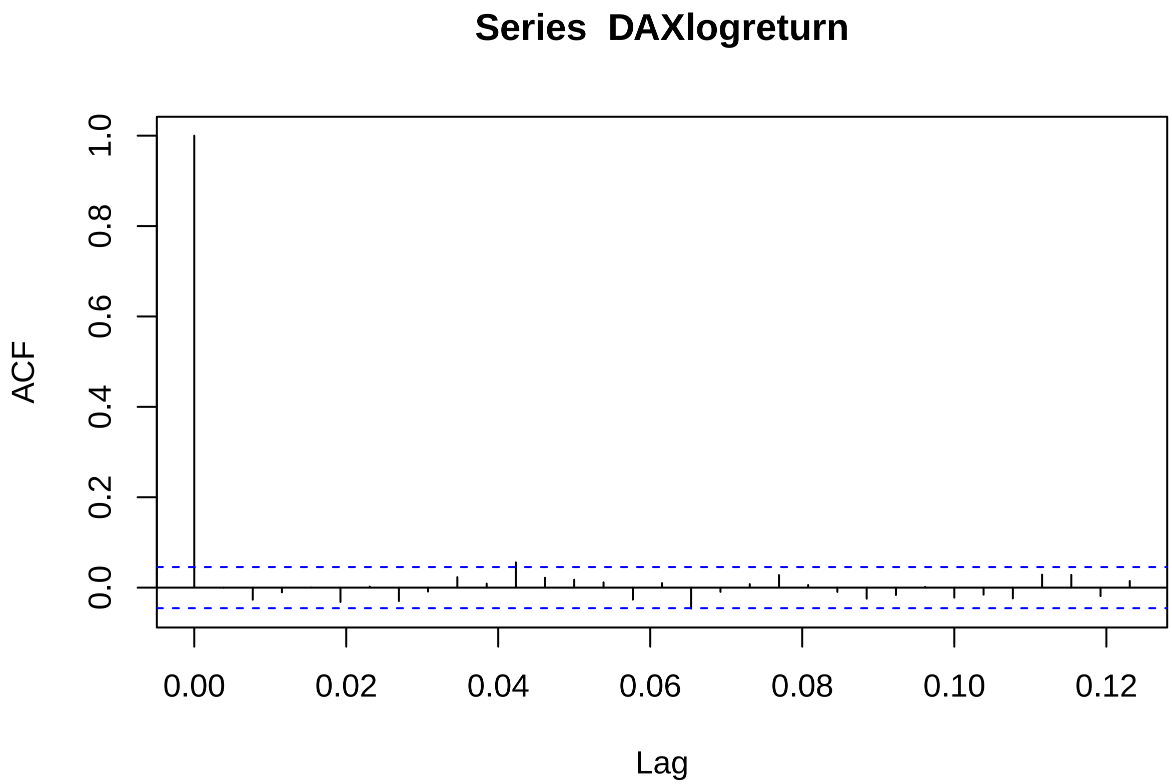
```r
pbinom(50, 100, 0.56, lower.tail = FALSE)
```

```
## [1] 0.8659234
```

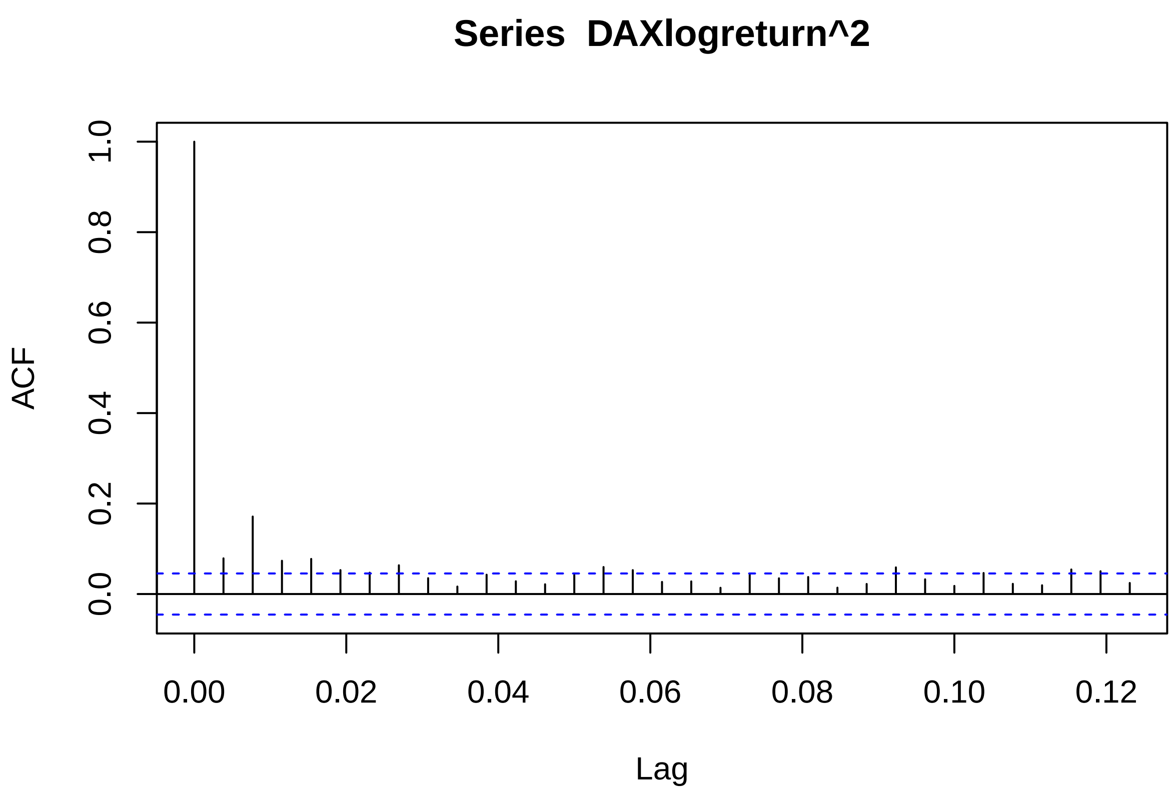
Therefore, the probability that more than 50 insects survive in a new experiment is 0.866.

Question 2

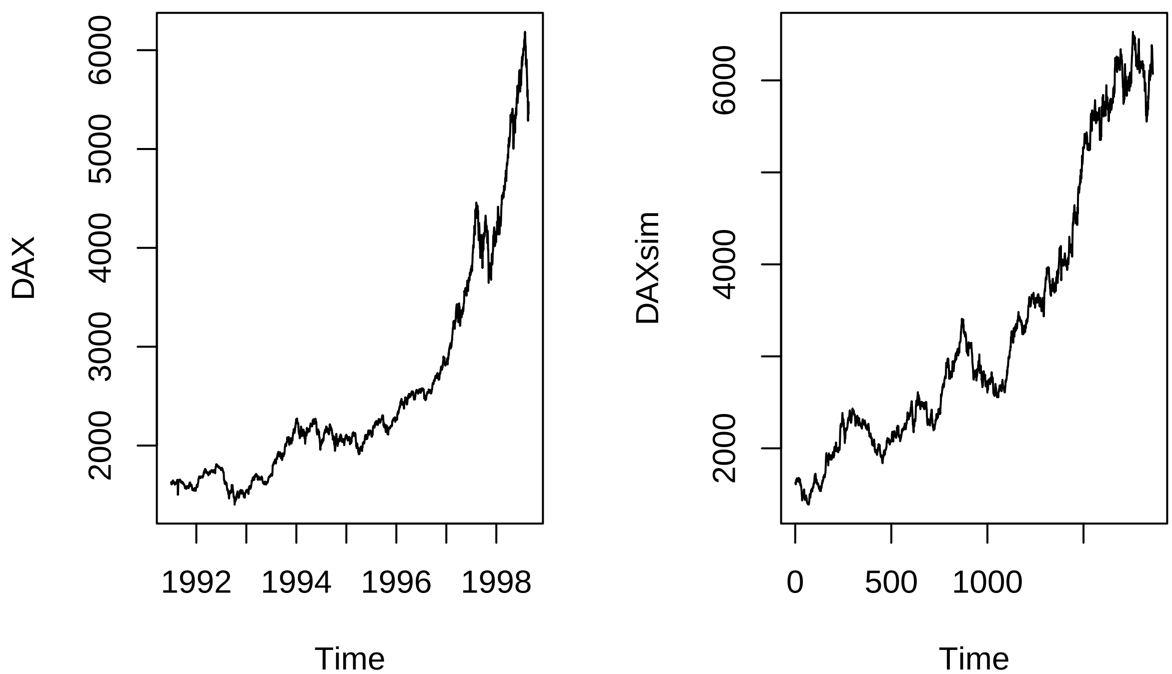
```r
DAX <- EuStockMarkets[,1] # extract the data
DAXlogreturn <- diff(log(DAX)) # (a) calculate the log returns
drift <- mean(DAXlogreturn) # (b)
acf(DAXlogreturn) # (c) no substantial autocorrelations present
```




```r
acf(DAXlogreturn^2) # (d) autocorrelations up to lag 2 present
```



```r
out <- arima(DAXlogreturn^2, order=c(2, 0, 0)) # (e)
phi <- out$coef[1:2]; xbar <- out$coef[3]; s2 <- out$sigma2
n <- length(DAX)
# Simulate from this model # (f)
y <- numeric(n)
y[1:2] <- diff(log(DAX))[1:2] # starting values for process
Z <- rnorm(n) # standard normals used in ARCH
for (i in 3:n) {
s <- sqrt(xbar + phi[1]*y[i-1]^2 + phi[2]*y[i-2]^2)*Z[i]
y[i] <- s
}
# y contains log returns, but an initial value is needed to
# re-accumulate the prices, and the drift term must be added in:
y <- c(log(DAX[1]), y + drift)
DAXsim <- exp(cumsum(y)) # simulated prices
par(mfrow=c(1, 2)) # (g) compare trace plots of real and simulated data.
ts.plot(DAX)
ts.plot(DAXsim)
```



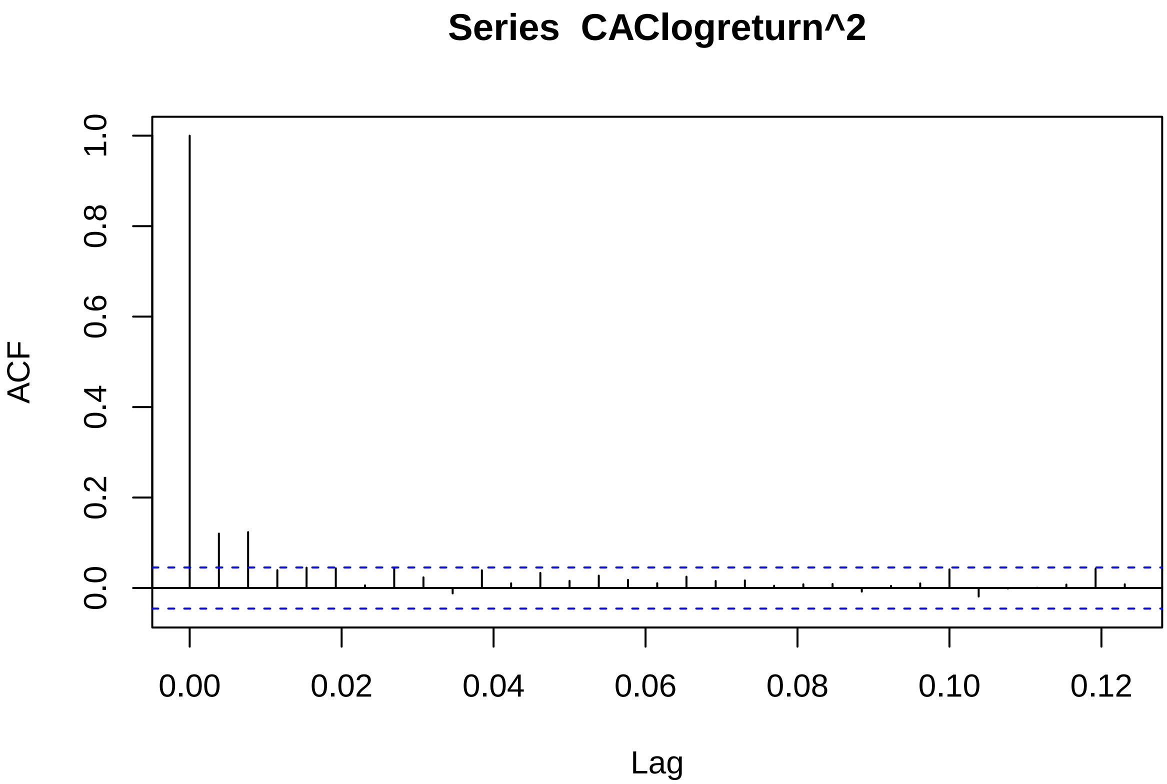
(c)
There are no autocorrelations that are significant enough.

(d) Yes there are autocorrelations up to lag 2 present. # Question 3

```r
# Load the CAC data (3rd column of EuStockMarkets)
CAC <- EuStockMarkets[, 3]

# Step 1: Calculate the log returns for the CAC data
CAClogreturn <- diff(log(CAC))

# Step 2: Check for autocorrelations in the squared log returns
acf(CAClogreturn^2) # This helps confirm the presence of ARCH effects
```



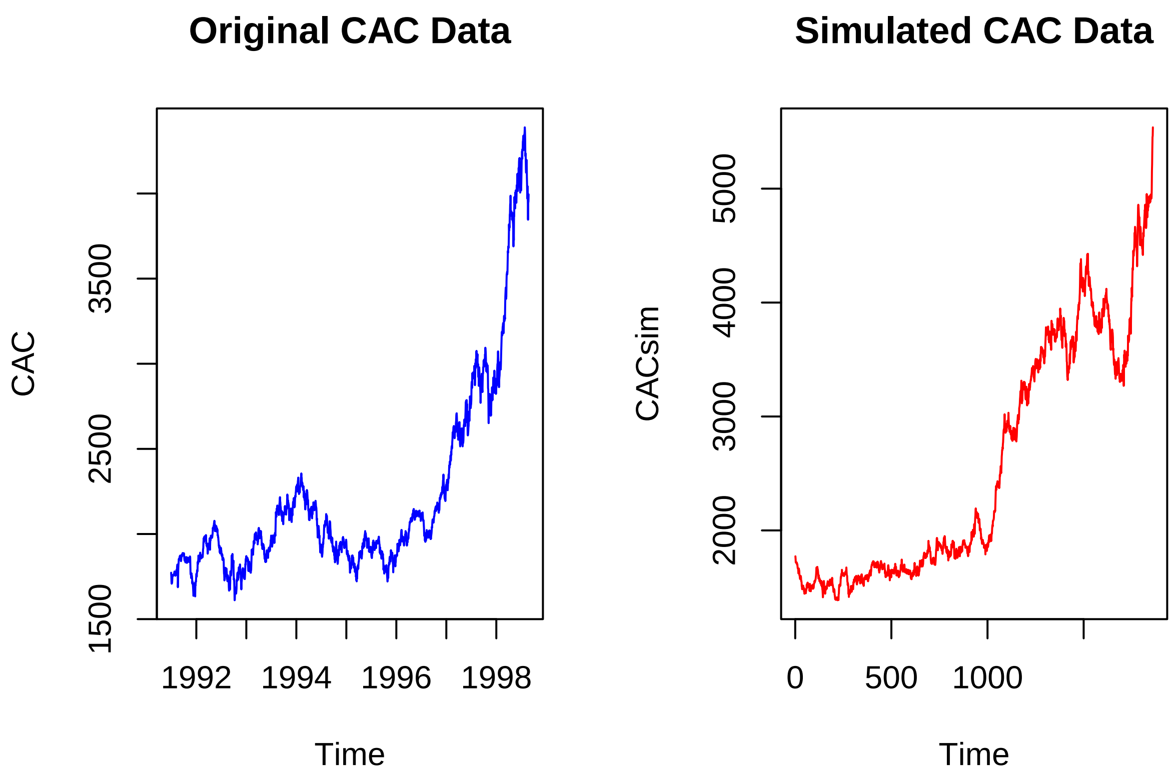

```r
# Step 3: Fit an ARCH(2) model to the squared log returns
out <- arima(CAClogreturn^2, order = c(2, 0, 0))
phi <- out$coef[1:2]    # ARCH coefficients
xbar <- out$coef[3]     # Intercept term
s2 <- out$sigma2        # Residual variance
n <- length(CAClogreturn)

# Step 4: Simulate a new series with the same properties
y <- numeric(n) # Initialize the simulated log returns
y[1:2] <- CAClogreturn[1:2] # Use the first two log returns as starting values
Z <- rnorm(n) # Generate standard normal random variables for the ARCH model

for (i in 3:n) {
  s <- sqrt(xbar + phi[1] * y[i-1]^2 + phi[2] * y[i-2]^2) * Z[i]
  y[i] <- s
}

# Add drift and re-accumulate prices to generate the simulated price series
drift <- mean(CAClogreturn) # Calculate the drift term
y <- c(log(CAC[1]), y + drift) # Add drift and initial log price
CACsim <- exp(cumsum(y)) # Convert log returns back to prices

# Step 5: Plot the original and simulated data
par(mfrow = c(1, 2)) # Arrange plots side by side
ts.plot(CAC, main = "Original CAC Data", col = "blue")
ts.plot(CACsim, main = "Simulated CAC Data", col = "red")
```



Question 4

(a) Finding the transition matrix, first we consider that the state space is $S = \{0, 1, 2, 3\}$, where X_n is the number of containers in the yard at time n.

- Arrival Rule: A container arrives unless the yard is full, which is at state 3.
- Removal Rule: Each container is independently removed with probability $p = 0.8$

At state 0, $(X_n = 0)$:

- $P(0 \to 1) = 0$ (no container arrives)
- $P(0 \to 0) = 1$ (no container is removed)

At state 1, $(X_n = 1)$:

- $P(1 \to 0) = 0$ (not possible since atleast a container arrives)
- $P(1 \to 1) = (1 - p) = 0.2$ (container is removed with probability 0.8)
- $P(1 \to 2) = p = 0.8$ (container arrives with probability 0.8)
- $P(1 \to 3) = 0$ (we cannot reach 3 directly from state 1)

At state 2, $(X_n = 2)$:

- $P(0 \text{ containers removed}) = (1 - p)^2 = 0.2^2 = 0.04$

- $P(1 \text{ container removed}) = 2p(1-p) = 2 \times 0.8 \times 0.2 = 0.32$
- $P(2 \text{ containers removed}) = p^2 = 0.8^2 = 0.64$

Thus,

- $P(2 \to 1) = P(1 \text{ containers removed}) = 0.32$
- $P(2 \to 2) = P(0 \text{ containers removed}) = 0.04$
- $P(2 \to 3) = P(2 \text{ containers removed}) = 0.64$

At state 3, $(X_n = 3)$:

- $P(0 \text{ containers removed}) = (1-p)^3 = 0.2^3 = 0.008$
- $P(1 \text{ container removed}) = 3p(1-p)^2 = 3 \times 0.8 \times 0.2^2 = 0.096$
- $P(2 \text{ containers removed}) = 3p^2(1-p) = 3 \times 0.8^2 \times 0.2 = 384$
- $P(3 \text{ containers removed}) = p^3 = 0.8^3 = 0.512$

Thus,

- $P(3 \to 0) = P(0 \text{ containers removed}) = 0.512$
- $P(3 \to 1) = P(1 \text{ container removed}) = 0.384$
- $P(3 \to 2) = P(2 \text{ containers removed}) = 0.096$
- $P(3 \to 3) = P(3 \text{ containers removed}) = 0.096$

Then, the transition matrix is given by

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0.2 & 0.8 & 0 \\ 0 & 0.32 & 0.04 & 0.64 \\ 0.512 & 0.384 & 0.096 & 0.008 \end{bmatrix}$$

(b) Probability of $X_3 = 3$ given that $X_1 = 2$ is given by

$$P(X_3 = 3 | X_1 = 2) = P^2[2, 3]$$

```
P <- matrix(
  c(0, 1, 0, 0,
    0, 0.2, 0.8, 0,
    0, 0.32, 0.04, 0.64,
    0.512, 0.384, 0.096, 0.008), nrow = 4, byrow = TRUE)

P2 <- P %*% P
P2[3, 4]
```

```
## [1] 0.03072
```

Therefore, the probability of $X_3 = 3$ given that $X_1 = 2$ is 0.030.

(c) A state space is defined as irreductible if every state can be reached from every other state, possibly over multiple steps Here, we see that it can eventually transition to the other states, via arrivals and removals. Thus, the state space is irreductible.

(d) The limiting distribution $\pi = (\pi_0, \pi_1, \pi_2, \pi_3)$ satisfies:

- $\pi = \pi P$
- $\sum_{i=0}^{3} \pi_i = 1$

Then we need to find that

$$\pi_0 = 0.512\pi_3$$

$$\pi_1 = 0.384\pi_3 + 1\pi_0$$

$$\pi_2 = 0.096\pi_3 + 0.8\pi_1$$

$$\pi_3 = 0.008\pi_3 + 0.64\pi_2$$