The University of British Columbia

Computer Science/Data Science 405/505 Modelling and Simulation

Asssignment 4

Exercises

1. 100 insects are placed in a container holding a certain amount of insecticide. After 1 hour, 44 insects have died. Assuming that the insects survive independently of each other, use a binomial distribution model, together with maximum likelihood, to estimate the probability that more than 50 insects would survive in another experiment held under identical conditions.

2. Data from the German stock exchange is in the DAX column of `EuStockMarkets`.

   (a) Store the successive differences of the log of the data in an object called `DAXlogreturn`.

   (b) Calculate the mean of the log returns. This represents a deterministic drift in the series which translates into a deterministic trend - either upwards or downwards.

   (c) Apply the `acf()` function to these data. Is there evidence of linear dependence (i.e. autocorrelation)? If so, at which lags?

   (d) Apply the `acf()` function to the squared data points. Is there evidence of autocorrelation now? If so, at which lags?

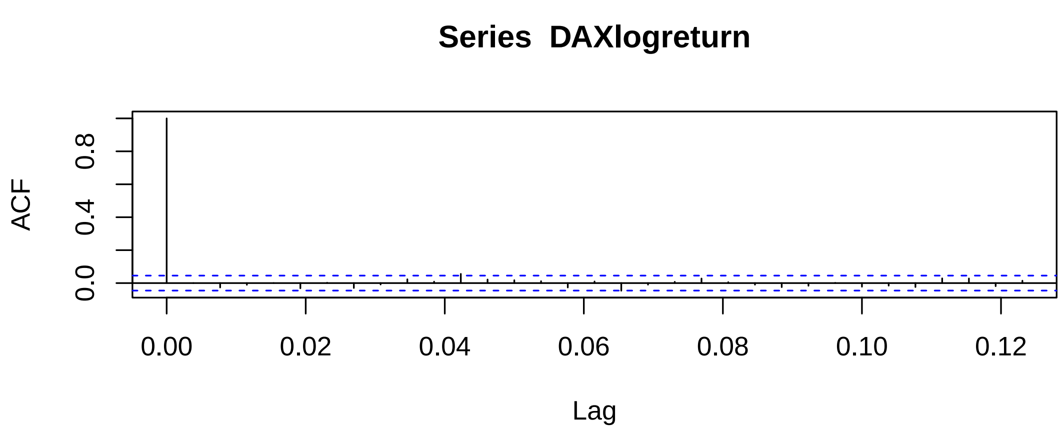
   (e) We can obtain approximate values to a_0, a_1 and a_2 for an ARCH(2) model by fitting an AR(2) model to the squared data points. Apply this technique to the `DAXlogreturn` data. Write out the fitted AR(2) model.

   (f) Using the ϕ estimates from the fitted AR(2) model, write out an approximate ARCH model for the `DAXlogreturn` data.

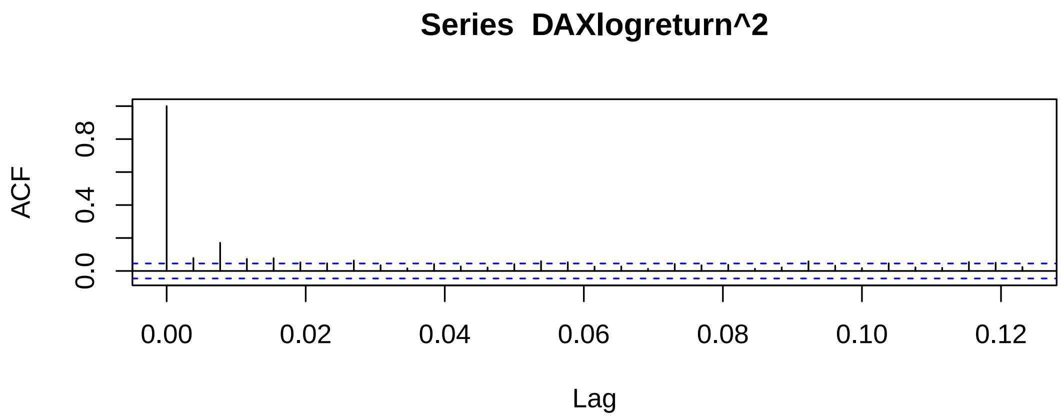
   (g) Using the fitted model, simulate a time series of the same length as the DAX series which has essentially the same properties, and plot the result, together with the original data. Make sure to include the drift term in your model.

   Detailed hint:

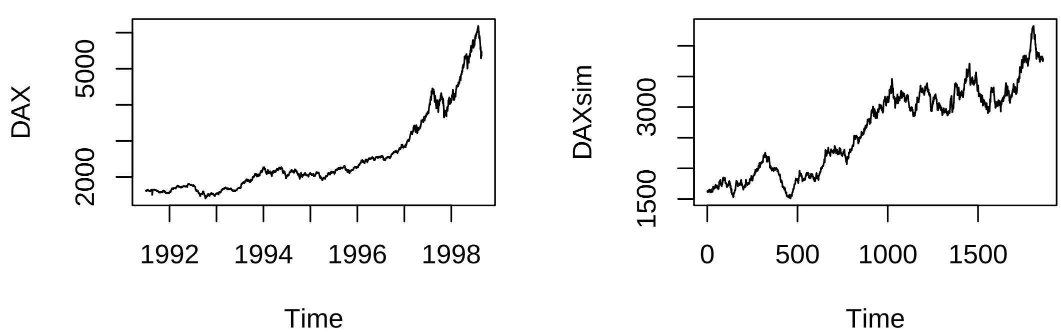
```
DAX <- EuStockMarkets[,1]  # extract the data
DAXlogreturn <- diff(log(DAX)) # (a) calculate the log returns
drift <-  mean(DAXlogreturn) # (b)
acf(DAXlogreturn) # (c)  no substantial autocorrelations present
```



```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)
```



3. Fit an ARCH(2) model to the French CAC stock market data (the 3rd column of `EuStockMarkets`). Simulate a new series of the same length with the same properties as the CAC data, and plot the simulated data, as well as the original data.

4. Containers are temporarily stored at a stockyard with capacity to store 3 containers. At the beginning of each day, precisely one container arrives at the stockyard, unless the stockyard is full; in that case, the container is taken elsewhere. Each container stays a certain amount of time before it is removed. The residency times of the containers are independent of each other. A container will be removed during a given day with probability $p = .8$ (independently of how many days the container has been stored).

   Let X_n denote the number of containers in the stockyard at the end of day n. $\{X_n\}$ is a Markov chain with state space $\{0, 1, 2, 3\}$.

   (a) Find the transition matrix.

   (b) Find the probability that there are 3 containers in the stockyard at the end of day 3, given that there were 2 containers there at the end of day 1.

   (c) Is the state space irreducible? Explain.

   (d) If there is a limiting distribution, find it.

5. Consider the time-reversible Markov chain discussed in class:

$$P_{i,j} = \frac{1}{6} \min\left(\frac{\pi_j}{\pi_i}, 1\right), \text{ for } j = i-2, i-1, i+1, i+2$$

   and 0 for $|j - i| > 2$. $P_{i,i}$ is set to ensure that the row sums of P are 1.

   Use this in an MCMC simulation of the probability distribution $\pi_j = P(X = j) = k(.7)^{j-1}$ for $j = 1, 2, \ldots$, where k is a value that could be calculated but is not needed. Simulate 20000 values and use these to estimate the probability π_3.