SM2 Computational Statistics
Hilary Term 2019

Week 6 Practical

The data that we will consider are daily log-returns from the Financial Times Stock Exchange 100 Index (FTSE 100) from January 1984 until February 2019. The FTSE 100 is a share index of the 100 largest companies listed on the London Stock exchange. The data are downloaded using the package quantmod, which needs to be first installed.

```
#install.packages("quantmod") # if needed
library("quantmod")

ftse100 <- new.env()
getSymbols("^FTSE", env = ftse100, src = "yahoo",
           from = as.Date("1984-03-01"), to = as.Date("2019-02-04"))
```

```
## [1] "FTSE"
```

```
FTSE <- ftse100$FTSE
head(FTSE)
```

```
##            FTSE.Open FTSE.High FTSE.Low FTSE.Close FTSE.Volume
## 1984-03-01    1042.1    1047.1   1042.1     1046.1           0
## 1984-03-02    1048.8    1060.7   1048.8     1060.7           0
## 1984-03-05    1053.7    1053.7   1053.7     1053.7           0
## 1984-03-06    1053.7    1064.8   1050.9     1064.5           0
## 1984-03-07    1063.1    1063.1   1055.6     1055.6           0
## 1984-03-08    1055.6    1056.0   1050.9     1055.8           0
##            FTSE.Adjusted
## 1984-03-01        1046.1
## 1984-03-02        1060.7
## 1984-03-05        1053.7
## 1984-03-06        1064.5
## 1984-03-07        1055.6
## 1984-03-08        1055.8
```

We are interested in the daily log-returns y_t of the closing index, defined as

$$y_t = \log c_t - \log c_{t-1}$$

where c_t is the closing index at time t.

```
returns = diff(log(na.omit(FTSE$FTSE.Close)))
returns = returns[2:length(returns)]
names(returns)=c("FTSE logReturn")
head(returns)
```

```
##            FTSE logReturn
## 1984-03-02   0.0138600809
## 1984-03-05  -0.0066212882
## 1984-03-06   0.0101974722
## 1984-03-07  -0.0083959024
## 1984-03-08   0.0001895169
## 1984-03-09   0.0040644008
```

Note that `returns` is not a standard R data frame. It is an object of the class `xts`, specifically created to store and manage time-series data. Information on the `xts` class can be found at this url. An important difference in the behaviour of `xts` objects is in the use of the "[ ]" operator. We can for example access the data by entering the date directly, or a range of dates. Plotting `xts` objects also automatically adds the dates as an x label.

```
class(returns)
```

```
## [1] "xts" "zoo"
```

```
print(returns["2000-01-04"])
```

```
##              FTSE logReturn
## 2000-01-04    -0.03888374
```

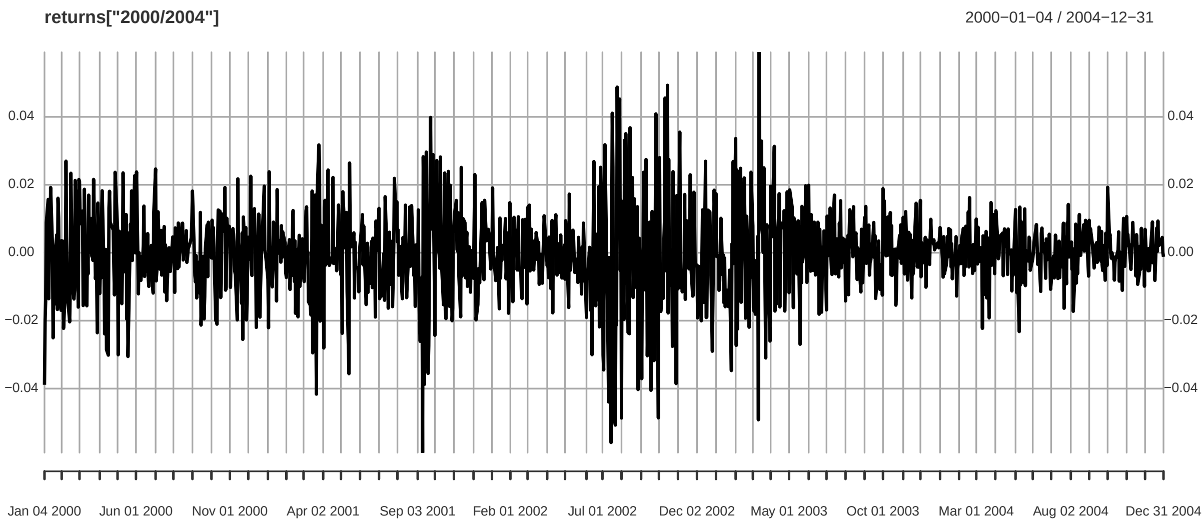
```
print(returns["2016-06-22/2016-06-30"])
```

```
##              FTSE logReturn
## 2016-06-22    0.005541438
## 2016-06-23    0.012207163
## 2016-06-24   -0.031966029
## 2016-06-27   -0.025824599
## 2016-06-28    0.026101445
## 2016-06-29    0.035154245
## 2016-06-30    0.022419351
```

```
print(returns["2019-01-25/"])
```

```
##              FTSE logReturn
## 2019-01-25   -0.001438166
## 2019-01-28   -0.009161870
## 2019-01-29    0.012782709
## 2019-01-30    0.015636803
## 2019-01-31    0.003925069
## 2019-02-01    0.007334356
```

```
plot(returns["2000/2004"])
```



If you want to convert the data into a matrix and extract the dates, you can use the following functions.

```
returns.mat = coredata(returns)
dates = index(returns)
class(returns.mat)
```

```
## [1] "matrix"
class(dates)
## [1] "Date"
```

1 Question 1

1. Exploratory analysis
   - Plot the whole dataset over time
   - Plot the data over the year 1987
   - Plot the data over 2008 and 2009
   - Plot an histogram of the whole dataset
   - Plot a normal qqplot of the whole dataset
   - Compute some summary statistics of the data
2. Consider first that we ignore the temporal structure, and assume that the log-returns $Y_1, \ldots, Y_n$ are iid from some unknown distribution F. Let F^{-1} be the associated quantile function. For a given quantile $\theta(q) = F^{-1}(q)$ with $q \in (0, 1)$, we consider the plug-in estimator

   $$\hat{\theta}_n(q) = F_n^{-1}(q)$$

   where F_n is the empirical cdf.
   - Calculate the estimates $\hat{\theta}_n(q)$ for $q = 0.01, 0.25, 0.5, 0.75, 0.99$.
   - Using the nonparametric bootstrap, provide 99% confidence intervals for $\theta(q)$ and $\exp(\theta(q))$ for $q = 0.01, 0.25, 0.5, 0.75, 0.99$.
3. We now want to estimate the probability that the FTSE100 closing index increases by more than 1% in a day that is

   $$\eta = \mathbb{P}(Y_t \geq \log(1.01))$$

   - Propose an unbiased and consistent estimator for η, and calculate the estimate.
   - Using the nonparametric bootstrap, provide a 99% confidence interval for η
   - Repeat the above steps to estimate the probability that the index increases by 5% in a day and provide a 99% confidence interval

2 Question 2

We now want to identify periods of high volatility in the data. To do so, we consider a two-state homogeneous hidden Markov model $(X_1, Y_1, \ldots, X_n, Y_n)$ with state-space $\mathcal{X} = \{1, 2\}$ where 1 corresponds to a normal volatility, and 2 to high volatility and observation space $\mathcal{Y} = \mathbb{R}$. The emission density is given as

$$g_1(y) = \mathcal{N}(y; 0, \sigma_1^2)$$
$$g_2(y) = \mathcal{N}(y; 0, \sigma_2^2)$$

where $\sigma_1 = 0.01$ and $\sigma_2 = 0.05$. We define the state transition probabilities as $A_{1,1} := \mathbb{P}(X_t = 1|X_{t-1} = 1) = \mathbb{P}(X_t = 2|X_{t-1} = 2) := A_{2,2} = 0.99$, and uniform initial probability mass function.

- Derive a Viterbi algorithm to find the MAP estimate of the hidden states. You can adapt the code from the lectures, but some modifications may be necessary to ensure numerical stability. Explain why.
- Create a copy of the `xts` object `returns`, named `returns.hv`, and add a column "HighVolatility" to this object, with value 1 if $\hat{x}_t = 2$, 0 otherwise
- Plot the log-returns and the estimated periods of high volatility on the same figure.
- Look at some of these periods and try to explain the high volatility.
- Plot a normal qqplot of the data in state 1 and another one for data in state 2. Comment on ways to improve the goodness-of-fit of the hidden Markov model.
- Change the parameters $A_{1,1}, A_{2,2}, \sigma_1, \sigma_2$ of the HMM, and study the sensitivity of the MAP estimates to the setting of the parameters.