

## Homework 2

*Due: Monday, Oct 23, 11:59pm*

*Submit online by canvas*

The book by Carmona refers to “Statistical Analysis of Financial Data in R”  
by René A. Carmona, **2nd edition**.

**I.** This problem analyzes the daily Hang Seng Index data.

1. Create a vector *DHSIRET* containing the daily raw returns. Recall that the raw return on a given day is the difference between the value on that day and the day before divided by the value on the previous day. Compute the mean and standard deviation of daily raw returns.
2. Conduct an exploratory data analysis using the function `eda.shape`. Comment.
3. Fit GPD to the returns, give detailed procedure including how to specify the tails and check the model fitting. Discuss the fitting results.
4. Generate a sample of size 10,000 from the GPD fitted above. Call this sample *S\_DHSIRET*, produce a Q-Q plot of *DHSIRET* against *S\_DHSIRET*, and comment.
5. Compute the VaR at the level of 0.005
  - (a) using empirical quantile;
  - (b) assuming that the daily raw return is normally distributed;
  - (c) using the Monte Carlo sample *S\_DHSIRET* that you generated.

Explain the differences and similarities between the estimates of VaR so obtained.

6. Compute the expected shortfall at the level of 0.005
  - (a) using empirical conditional mean;
  - (b) assuming that the daily raw return is normally distributed;
  - (c) based on the fitted GPD distribution.

Explain the differences and similarities between the estimates of expected shortfall so obtained.

Note: The HSI data can be downloaded from Canvas. The necessary commands are available in “gpd\_code.R”.

**II.** Parts 1 & 2 of Problem 3.7 on page 187 of the book by Carmona. (cf. Lect 6)

Note: The data UTILITIES can be downloaded from Canvas. You can use the following command to load the data:

```
UTILITIES<-read.csv("UTILITIES.csv")
X <- UTILITIES[ , 1 ]
Y <- UTILITIES[ , 2 ]
```

Hint for part 2: if  $X$  and  $Y$  are jointly normal, then any linear combination of them is also normal.

**III.** (cf. Lect 6 p.29-48)

1. Generate a sample of size  $n = 1024$  from the distribution of  $N(0,1)$  (call  $X$  the vector containing the sample values). Construct a vector  $Y$  with each entry being the square of the corresponding entry of  $X$ . Produce a scatterplot of all the points  $(x_i, y_i)$ . Give a guess of how big the sample correlation might be (no need to report). Compute the sample correlation  $\hat{\rho}(X, Y)$  using R (command: `cor`). Is it close to your guess? (no need to report)
2. Generate a sample from the distribution of  $N(3,1)$  of size  $n = 1024$  (call  $XX$  the vector containing the sample values). Construct a vector  $YY$  with each entry being the square of the corresponding entry of  $XX$ . Produce a scatterplot of all the points  $(xx_i, yy_i)$ . Give a guess of how big the sample correlation might be (no need to report). Compute the sample correlation  $\hat{\rho}(XX, YY)$  using R. How was your guess? (no need to report)
3. Compute the correlation coefficient of two random variables  $A \sim N(\mu, \sigma^2)$  and  $B = A^2$ .  
Hint: Recall that if  $A \sim N(\mu, \sigma^2)$ , then  $A$  can be written as  $A = \mu + \sigma Z$  where  $Z \sim N(0, 1)$ . Also, if  $Z \sim N(0, 1)$ , then  $E(Z) = E(Z^3) = 0$ ,  $E(Z^2) = 1$ , and  $E(Z^4) = 3$ .
4. Consider the setting of the above question with  $\mu = 0$ . Are  $A$  and  $B$  uncorrelated? Are  $A$  and  $B$  independent?