Final Exam QF202

To be submitted in Canvas no later than May 14, 2018

Notes:

- There are two parts in this exam, a theoretical and an applied part.
- The theoretical part can be done on paper but the paper needs to be scanned and uploaded as a pdf file using the link provided in canvas. For the applied part, please generate your report using R Markdown, as presented in class (all relevant code needs to be visible). The two files may be combined into a single pdf.
- Please note that submission in pdf format is required and that 5 points will be subtracted from any submission not conforming with the requirements.
- You have all worked with downloadable data more then once during these semester, and, at this point, you should all know that sometimes data do not come in the best format to work with. In previous assignments, we guided you on dealing with these problems, but now it is up to you. If you fail to perform the required preprocessing, points will be subtracted.
- Be very specific with your random variables and your definitions. Showcase your work.
- Communication with other students either physical or virtual is forbidden.

For instructor's use only

Problem	**Points**	**Score**
Theoretical	100	
Practical	100	
Interview	100	
Total	300	

Theoretical problems

Probability Problems

1. Based on an empirical study, when a customer makes a restaurant reservation, there is a 20% chance the reservation is canceled. Assume you work for this restaurant and this restaurant requires reservation. Tonight the restaurant accepted 52 reservations and there are 50 tables available in this restaurant. We also assume that each reservation is independent of any other reservation.

   - What is the probability that there will not be enough tables available to cover all reservations?
   - What is the probability that more than 3 tables will remain empty?

2. Consider the function $f : \mathbf{R} \to \mathbf{R}$ such that

$$f(x) = \frac{2}{\pi(1+x^2)} \mathbf{1}_{[0,\infty)}(x)$$

   Show that f is a probability density function.

3. Let X a random variable distributed as an exponential with parameter λ, $\lambda > 0$. That is, its probability density function is:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

   Prove that for every $s > 0$ and $t > 0$,

$$\mathbf{P}(X \geq s + t | X \geq s) = \mathbf{P}(X \geq t)$$

4. Suppose the simple return of a monthly bond index follows the AR(1) model:

$$r_t = 0.1 + 0.3 r_{t-1} + a_t$$

   We know the standard deviation of the white noise $\sigma = 0.025$. Assume that $r_{1000} = 0.001$. Calculate the value of the 1-step and 2-step ahead forecast at time $t = 1000$. Calculate the standard deviation of the corresponding forecast errors. Finally compute the numerical values for the 1, 2 and 3 lag autocorrelation of the return time series.

Practical Problem For this problem, use the Yahoo! Finance website (`http://finance.yahoo.com/`) or a Bloomberg terminal. Download the past 3 years of daily prices of a stock of your choice.

(a) Compute the time series of continuously compounded returns (log returns) using the series of daily close prices (if you downloaded data from Yahoo! Finance, use the Adj. Close column). Please provide plots of both time series (close prices and log returns) with proper labels and legends.
For the remainder part of this Problem, we will be using only the series of log returns.

(b) Divide your data in 2 parts: the out-of-samble set, containing only the last 5 days of data; and the in-sample set, containing all the rest. Make a plot containing only the last 5 days of the in-sample set and all of the out-of-sample set (i.e. the last 10 days of data), using different colors for each of the sets. Again, provide proper labels and legends.

(c) Using the in-sample set, build three different models: one AR, one MA, and one ARMA. Use all the tools you have learned during the semester to help you create the best possible models. You should be able to explain all the steps you use and discuss a lot. **Please remember that although the best model sometimes is a random walk, you should avoid it, since its use is uninteresting.**

(d) Use the `predict` function to perform a 5-day ahead forecast with your 3 models. Repeat the plot from part **(b)**, but now including the prediction as a third set. In total you should have 3 plots: one for each model.

(e) Compute the sum of squared errors between the predicted values of each model and the realized set of values of the out-of-sample set. Using these results, please explain which of the three models you created you consider the best.

Interview Type Problem Attached please find three data files (`qf-202-final-data-xxx.csv`, `qf-202-final-data-yyy.csv`, and `qf-202-final-data-zzz.csv`). It is known that all the data comes from ARIMA time series models (please note this information is not typically given). We also know that the models used were of the same order (e.g., they were all ARIMA(3,2,6)) but with different parameter values as well as different white noise variances. The white noise distribution is the same for all datasets. Please construct code in R that will input these files and output a csv file containing the estimated parameter values. Specifically, for whatever order you estimated the ARIMA(p,d,q) model to have you have to output μ, $\phi_1, \dots \phi_p$, $\theta_1, \dots, \theta_q$, and σ. Either output 3 files or a single csv file containing 3 rows corresponding to file 1, 2, and 3. You will be judged as follow:

1. The quality of the report you provide detailing the approach you took.
2. How far are your estimates from the known parameters used to generate the data.
3. Whether or not your submitted code can run a 4^{th} csv file of a different length and produce meaningful output.