CFRM 502 Financial Data Science

Homework 7

Total marks: 29

March 7, 2022, 11:59 PM

General instructions. Submit your answers as PDF or HTML files on Canvas. The files should include your answers, working out, and explanations where appropriate. The only programming language that can be used is R. For questions where you use R, you must include your R code and its outputs generated using R Markdown or Jupyter Notebook and submitted as a PDF or HTML file. Please check that your submission on Canvas can be opened and the content of the files are complete and correct.

Question 1. Consider the dataset[1] stored in the variable `futures` obtained by downloading the data from Canvas as a Rdata file and then loading it into your R environment by running the following.

```
load("Homework 7 Data.Rdata")
```

The dataset contains 1-minute intraday spot and futures log prices of the S&P 500 index in the columns `lnspot` and `lnfuture`, respectively. The prices are from May 1993 and are for the June futures contract.

(a) Let the response variable be $\{Y_t\}$, the differences of `lnfuture`, and the predictor variable be $\{X_t\}$, the differences of `lnspot`. Fit a linear model and determine whether or not the residuals have autocorrelation. [1 mark]

(b) Fit an ARIMAX model for $\{Y_t\}$ with $\{X_t\}$ as the predictor. Then using residual diagnostics, check whether the model is a good fit or not. [3 marks]

(c) Use the fitted model to find the 95% prediction interval for the 1-step ahead forecast of the futures log return if the 1-step ahead value of the spot log return is 0.0002. [1 mark]

Question 2. Consider 2 years of daily log returns $\{Y_t\}$ of the S&P 500 index before the Black Monday stock market crash stored in the variable `ret`, and the log return on Black Monday stored in `ret.black.mon` obtained by running the following code.

```
data(SP500, package="Ecdat")
ret <- SP500$r500[(1804-2*253+1):1804]
ret.black.mon <- SP500$r500[1805]
```

Denote the observations in `ret` as $Y_1, \dots, Y_n$.

(a) Explain why an ARIMA process is not adequate to model $\{Y_t\}$ (the variable `ret`) and why an MA(1)+GARCH(1,1) process should be considered instead. [2 marks]

(b) Fit a MA(1)+GARCH(1,1) process where the white noise of the GARCH process is a standard t-distribution. Then using residual diagnostics, check whether the model is a good fit or not. [4 marks]

(c) Let time n be the day before Black Monday. The distribution of Y_{n+1} conditional on the information up to time n is $a + bT$, where $a, b \in \mathbb{R}$ and T is a standard t-distribution with ν degrees of freedom. Find a, b and ν. [3 marks]

(d) The 99.9% value-at-risk (expressed as a log return) is the 0.999-quantile of $-Y_{n+1}$ conditional on the information up to time n. Does the loss for the log return on Black Monday exceed the 99.9% value-at-risk? [2 marks]

[1] This data appears in *Analysis of Financial Time Series*, 3rd ed., R. S. Tsay, John Wiley & Sons, 2010, and was obtained from the website, Chapter 2, data set for exercise 14.

(e) Plot the estimated conditional volatility (standard deviation) according to the fitted model. [1 mark]

(f) Recall that the conditional volatility is given by the GARCH formula $\sigma_t^2 = \omega + \alpha a_{t-1}^2 + \beta\sigma_{t-1}^2$. However, σ_1 depends on quantities before the first observation. Now suppose that σ_1 is arbitrarily set to 0.013. If your fitted model using `garchFit()` in the `fGarch` package is stored in `mod`, then $\{a_t\}$ can be obtained using `residuals(mod, standardize=FALSE)`, and the coefficients of the fitted model can be obtained using `coef(mod)`. Estimate the conditional volatility in this case, and add it to the plot in (e). [3 marks]

(g) By considering the GARCH formula, explain why the plots of the conditional volatility in (e) and (f) quickly converge to the same values. [2 marks]

Question 3.

Consider a stationary GARCH(1,1) process $\{a_t\}$ given by

$$a_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \omega + \alpha a_{t-1}^2 + \beta\sigma_{t-1}^2,$$

where $\{\varepsilon_t\} \sim \mathrm{IID}(0,1)$. Time is measured in days, and it is currently day $n-1$. Show that the average predicted variance over the next T days is

$$\overline{\sigma}_{n-1}^2(T) = v + \frac{1-e^{-aT}}{T(1-e^{-a})}(\sigma_n^2 - v),$$

where $v = \frac{\omega}{1-\alpha-\beta}$ and $a = -\log(\alpha+\beta)$. [3 marks]

Question 4. Use `set.seed(343)` at the start of your code for this question. Let $\{Y_t\}$ be an ARFIMA$(1, 0.3, 0)$ process with $\phi_1 = 0.35$ and $\sigma^2 = 1$. Simulate a sample path of this process with n observations. Then fit an ARFIMA$(0, d, 0)$ process and an ARFIMA$(1, d, 1)$ process to obtain the estimate of the order of fractional integration d, which will be denoted $\widehat{d}_0$, and $\widehat{d}_1$, respectively. Repeat this 1000 times to obtain histograms of $\widehat{d}_0$ and $\widehat{d}_1$ to approximate the sampling distributions in the case where $n = 200$. Then repeat this again to obtain the two histograms in the case where $n = 1000$. Based on your results, comment on how well fitting an ARFIMA$(0, d, 0)$ process and fitting an ARFIMA$(1, d, 1)$ process does in estimating d. [4 marks]