**Group Project**

**Semester II, 2018**

This project is worth 30% of the assessment for this unit. Submit a hard copy of your assignment on or before the due date. This is a group project where groups of one, two or three are acceptable. Do NOT send the soft copies through email. However, please keep the soft copies of all of your files (e.g., MS Word, Excel, R script etc.) with your workings. They may be requested, if necessary, during marking process. **The due date for the submission is 11th of October 2018, by 16:00.** Assignment box number TBA later which is located on level 5, building 6B.

Before you start this project, make sure that you have updated R and RStudio to the latest version. Update can be done using **updateR()**command in **installr** package.

# installing/loading the latest installr package:

**install.packages**("installr"); **library**(installr)

# install+load installr

updateR() # updating R.

This project focuses on financial asset return volatility forecasting. The concept of volatility is one of the important topics in Finance because the investors, traders, market participants, and portfolio managers are interested in volatility dynamics of asset returns in their decision making. In particular, volatility is used as an input in option pricing, conditional beta, and optimal hedge ratio computations. Moreover, financial return volatility forecast is used to calculate a value-at-risk which is the most common measure of risk in practice.

The following learning objectives will be addressed in this project.

2. To understand and model the first moment (mean) and second moment (volatility) processes.

5. To understand and conduct diagnostic tests of financial time series models and generate forecasts for returns and volatility.

The data for this assignment consists of observations on the daily closing futures prices on West Texas Intermediate crude oil, heating oil #2, and natural gas. These futures contracts trade on the New York Mercantile Exchange (NYMEX) and contract specifications and trading details are available from their website (www.nymex.com). The data set for crude oil and heating oil #2 covers the period January 1, 2001 to December 29, 2017 for a total of 4435 observations. The data have been downloaded from Datastream. Futures continuously compounded price returns are calculated as, where is the closing price on day . As the full sample comprises 4435 observations, we get 4434 return observations.

1. Do each of the following tasks and provide the answers.
2. Is there any ARCH effect in all log return series (use all the available return observations)? Use 5 and 10 lags of the squared returns and 5% significance level to perform the test? Interpret the results.
3. Compute the PACF of all squared log returns (10 lags). Use all the available return observations. Interpret the results.
4. Consider the daily squared log return series of all commodity futures (i.e., oil, heating oil, and natural has). Test versus for some . Draw the conclusion. Use all the available return observations. Interpret the results.
5. Plot news impact curves for standard AR(1)-GARCH(1,1) model and AR(1)-EGARCH(1,1) model for all commodity futures under concern. Use all the available return observations. Interpret the results.

Provide R codes and/or Excel calculations and outputs in your hard-copy.

**[30%]**

1. A rolling window method of obtaining out-of-sample forecasts has been commonly used as an approach for competing models in practice. A rolling window is one where the length of the in-sample period used to estimate the model is fixed, so that the start date and end date successively increase by one observation.

The beginning and ending dates of the in-sample estimation period are then moved forward one day. The model coefficients are re-estimated, and finally, these new estimates are utilized to forecast daily one-step ahead return volatility over the out-of sample period. This procedure is repeated until the available out-of sample period is exhausted. The out-of-sample period includes the last 1250 observations of the full-sample period of 1 January 2001 to 29 December 2018 and covers the period 18 March 2013 to 29 December 2018 for each commodity.

At time period *t*, a 1-day forecast is made. Models are estimated with 3184 return observations. The estimation period is then rolled forward by adding one new day and dropping the most distant day. In this way the sample size used in estimating the models stays at a fixed length and the forecasts do not overlap. Thus there are 1250 one-day return volatility forecasts for each of petroleum futures price.

The models to be considered:

**AR(1)-GARCH(1,1) model:**

The GARCH models jointly estimate a conditional mean and a conditional variance equation. These models are very useful when analysing the data that appears to exhibit volatility clustering. The AR(1)-GARCH(1,1) is given by

|  |  |
| --- | --- |
| , | (1) |

Note: when you are estimating AR(1)-GARCH(1,1) model, please add the following sub command **solver = "hybrid"** in **ugarchfit()** function.

**AR(1)-GJR-GARCH(1,1) model:**

In financial markets, it is often the case that downward movements in the market are followed by higher volatilities than upward movements of the same magnitude. This asymmetry can be modelled using the GJR-GARCH model.

|  |  |
| --- | --- |
| , | (2) |

where if , otherwise.

Note: when you are estimating AR(1)-GJR-GARCH(1,1) model, please add the following sub command **solver = "hybrid"** in **ugarchfit()** function.

**AR model:**

This model uses an autoregressive process to model volatility. Five lagged values of past volatility, corresponding to the average number of trading days in a week, are used as drivers.

|  |  |
| --- | --- |
|  | (3) |

where

Note: **Arima()** or **arima()** commands in R can be used to estimate the AR(5) model.

**Least squares linear regression model:**

This model uses an ordinary least squares (OLS) regression to model volatility by using a one period lagged value of past volatility as driver.

|  |  |
| --- | --- |
|  | (4) |

where

**Naïve model:**

From a naïve model, the best forecast of next period's return volatility is this period's actual return volatility.

|  |  |
| --- | --- |
|  | (5) |

Note: Excel spreadsheet will be sufficient to obtain the forecasts.

**Historical mean model:**

From an historical model, the best forecast of next period’s return volatility is the average of the previous volatility.

|  |  |
| --- | --- |
|  | (6) |

Note: Excel spreadsheet will be sufficient to obtain the forecasts.

**Simple moving average model:**

Moving average (MA) methods are widely used in time series forecasting. In this exercise a moving average of length m where m=20, 60, 180 days is used to generate return volatility forecasts. These values of m correspond to one month, three months and six months of trading days respectively. The expression for the m day moving average is shown below

|  |  |
| --- | --- |
|  | (7) |

Note: Excel spreadsheet will be sufficient to obtain the forecasts.

Note: Use for a proxy of actual volatility.

1. You are trying to determine the accurate forecasting model for the energy commodity futures volatility among the models and techniques specified earlier. Use the mean squared error (MSE) as well as mean absolute deviation (MAD) loss functions for the model comparison purposes. Rank all the competing models. What is the ‘best’ and ‘worst’ model for each commodity futures return volatility?

If is a vector of one-step ahead return volatility forecasts generated from in-sample rolling windows, and is the vector of observed actual volatility, then the out-of-sample MSE and MAD of the predictor is computed as:

Note: Use for a proxy of actual volatility.

Provide R codes and/or Excel calculations and outputs in your hard-copy.

1. Explain the asymmetry phenomenon in volatility of financial data. Explain whether the forecasting accuracy increases if an asymmetry is accounted for in the model. Draw your conclusion based on the results of AR(1)-GARCH(1,1) and AR(1)-GJR-GARCH(1,1).
2. Are the conclusion obtained from using MSE and MAD forecasting accuracy measures identical? Explain discrepancies, if there is any.

Provide R codes and/or Excel calculations and outputs in your hard-copy.

1. Do the complex models like AR(1)-GARCH(1,1), AR(1)-GJR-GARCH(1,1), and AR always produce best return volatility forecasts compared to least squares linear regression, naïve, historical, and simple moving average models? Justify your answers based on your results.

Provide R codes and/or Excel calculations and outputs in your hard-copy.

1. In practice, accurate volatility forecast is important for value-at-risk (VaR) calculation. Which model assumption would you use in VaR calculation for the commodity futures under concern? Explain your answer.

**[70%]**