Harvard Extension Data Science

Dynamic Modeling and Forecasting in Big Data

Instructor: William Yu

Assignment 9

Part A. Rolling Beta and Rolling Window Validation

- In H09a_rolling, we learned how to estimate model's beta/parameter as time goes by. This is similar to the spirit of Kalman filter for a time-varying parameter.
- Robert Shiller, Yale economist, won the Nobel Prize for his contribution partly on stock market research. Let's look at his research and data in the Assignment 9 folder: ie_data.xlsx. You can also see his website here: http://www.econ.yale.edu/~shiller/data.htm
- In tab "Data1" of the ie_data.xlsx, I selected four monthly variables:
  - **return**: Annual compounded real return *in the next 10 year*
    - *Ex -- In 2014/3, the return is compound annual growth rate (CAGR) of stock S&P500 prices between 2014/3 and 2024/3.*
  - **cape**: Cyclically adjusted price earnings ratios (*over the past 10 years*)
    - *Ex -- In 2014/3, the cape is price-earning ratio based on the data between 2004/3 and 2014/3.*
  - **trcape**: Cyclically adjusted total return price earnings ratios (*over the past 10 years*)
  - **ecy**: Excess CAPE yield
  - You can go to tab "Data" to see how these variables are being calculated.
- Let's build some fundamental long-term stock return models for the U.S. in which *return* is the dependent variable and *cape, trcape, and ecy* are potential predictors.

- **(1) OLS model (1881.1 to 2014.3)**
- Run four univariate least square model for the whole sample period, i.e.:
  - Model 1: lm(return ~ cape, data)
  - Model 2: lm(return ~ trcape, data)
  - Model 3: lm(return ~ ecy, data)
  - Model 4: lm(return ~ trcape + ecy, data)
  - Explain which model is the best
  - Note that this is in fact an out-of-sample regression model because the dependent variable – return is Year 1 to Year 10 while the explanatory variables are at Year 0.

- **(2) Rolling OLS model**
- Use the sample period from 1881.1 to 1999.12 as the trainset and use "rollapply" function to estimate alpha and beta in the testset **with the length of rolling window of 1428 months**

from 2000.1 to 2014.3. Meanwhile, predict the return and calculate the RMSE for all the rolling testset.

- Explain which model has the smallest testset errors.
- If you are interested in this subject, feel free to develop any other models for this assignment or your project.

Part B. GARCH Model prediction

- Use GARCH models to analyze the volatility of Telsa stock returns.
    - getSymbols("TSLA", src="yahoo")
- Following H10_grarch.R and use the three following models for Telsa stock returns.
    - (1) GARCH (1,1) model without mean
    - (2) GARCH (1,1) model with mean equation having ARMA(1,1) process
    - (3) GARCH (1,1) -in-mean model with mean having ARMA(1,1) process
- Show which model is the best model and copy and paste the conditional variance prediction over the next 100 days.