MSDS 413, Assignment 5
Nonstationary Time Series and Modeling Covariates (TS5)

Introduction

For this assignment, you will use the datasets and R script (TS5.R) included in the zip file (TS5.zip), posted to the Module 5 Overview page of Canvas. You will read the data files into R and conduct the requested analyses.

The instructions for submitting your assignment follow the Procedure section below.

The following list defines the data sets and their respective variables.

The monthly market liquidity measures are from Professors Pastor and Stambaugh. The data are available from Wharton WRDS and are in the file m-PastorStambaugh.txt. See `https://breakingdownfinance.com/finance-topics/equity-valuation/pastor-stambaugh-model/https://breakingdownfinance.com/finance-topics/equity-valuation/pastor-stambaugh-model/`

The following list defines the variables:

- DATE: is the month the data were collected
- PS_LEVEL: levels of aggregate liquidity
- PS_INNOV: innovations in aggregate liquidity
- PS_VWF: traded liquidity factor

The monthly Fama-Bliss bond yields have maturities of 1 and 3 years. The data are available from CRSP and are in the file m-FamaBlissdbndyields.txt.

The following list defines the variables:

- qdate: month of the yields
- yield1: 1 year yields
- yield3: 3-year yields

Your objective is to explore the time series behavior of these data sets including EDA, modeling, model diagnostics, and interpretation.

Procedure

The following steps are necessary to complete this assignment. Address each and every part and ensure that you cover all the details specified in the questions.

1. **Outlier Management** (20 points) Consider the monthly market liquidity measure of Professors Pastor and Stambaugh. The data are available from Wharton WRDS and are in the file m-PastorStambaugh.txt. Consider the variable PS level and denote the series by x_t.

   1.1. Perform EDA.

   1.2. Build a time series model for x_t (the expected value equation) using the model-building process. Write the equation of the model to be fitted (not the fitted model).

   1.3. Identify the largest outlier in the series. Refine the fitted model by using an indicator for the outlier. Write the equation of the refined model (not the fitted model).

   1.4. Further refine the model by setting the least significant parameters to zero. Write the equation of the revised model to be fitted (not the fitted model).

   1.5. Compare your model from part 1.3. with your model from 1.4.. Which is preferred and why?

2. **Box-Jenkins Methodology** (20 points) Consider the monthly Fama-Bliss bond yields with maturities of 1 and 3 years. The data are available from CRSP and are in the file m-FamaBlissdbndyields.txt. Denote the yields by y_{1t} and y_{3t}, respectively.

   2.1. Perform EDA.

   2.2. Build a time series model using the Box-Jenkins method for the log of the year three (y_{3t}) data: $x_t = \log(y_{3t})$. For simplicity, you may ignore possible outliers, but describe how you would treat outliers if they were not to be ignored.

   2.3. Fit the following model to the log earnings series:

   m <- arima(xt, order=c(0,1,1), seasonal=list(order=c(0,0,1), period=4))

   where **xt** denotes the log of the earnings. Write the equation of the fitted model. Compare this model with the model in part 2.2.. Which model is preferred? Why?

   2.4. Use the backtest procedure to compare these same two models via 1-step ahead forecasts. You may use $t = 600$ as the starting forecast origin. Which model is preferred? Why?

3. **ARIMA and Regression Errors** (20 points) Consider the monthly Fama-Bliss bond yields with maturities of 1 and 3 years. The data are available from CRSP and are in the file m-FamaBlissdbndyields.txt. Denote the yields by y_{1t} and y_{3t}, respectively. The goal is to explore the dependence of the 3-year yield on the 1 year yield.

   3.1. Fit the linear regression model $y_{3t} = \beta_0 + \beta_1 y_{1t} + e_t$ using the model-building process. Write the equation of the model to be fitted.

   3.2. Fit a linear regression model letting $d_{1t} = \triangle y_{1t}$ and $d_{2t} = \triangle y_{3t}$, where $\triangle$ is the differencing operator. Here $d_{it},\ i = 1, 2, 3$ denotes the change in monthly bond yields. Consider the linear regression $d_{3t} = \beta d_{1t} + e_t$. Write the equation of the model to be fitted. Is the model an adequate model? Why?

3.3. Based on the model refinements, describe and compare the linear dependence between the bond yields of the two linear regression models.

3.4. Fit an AR($m33\$order$) model to d_{3t} using d_{1t} as an explanatory variable using the model-building process. Write the equation of the model to be fitted.

3.5. Refine the model in 3.4. by setting the insignificant coefficients to zero. Write the equation of the fitted model. Compare this model with your best linear regression model. Which is better? Why?

3.6. Use the command **polyroot** in R to find the solutions of the characteristic equation of the refined AR($m33\$order$) model. How many real solutions are there?

3.7. Compute the inverse of the absolute values of the solutions of the characteristic equation. Show the maximum value of the inverses. Does the maximum value imply that the AR($m33\$order$) model likely contains a unit root? Why?

3.8. Compare the fit of your best linear regression model and your best AR model. Which is preferred and why?

4. **Report** (20 points) For the Fama-Bliss bond yields analyses, choose what you think is the best model's outcomes and write an executive summary that allows stakeholders to make decisions or take action.

Deliverables

Your instructor may modify these and all the following directions. See Section Submission Directions below. The assignment deliverables, each in pdf format, are as follows:

- *Only if requested by instructor*
  - The program or script
  - Logs
  - Outputs
- Mandatory
  Data analysis write-up: no programs, logs, or just code outputs; **complete EDA and model diagnostics are expected unless otherwise instructed; I will be looking for innovative interpretations in the assignments over and above the rote adherence to assignment requirements. Only partial credit will be awarded for rote adherence to assignment requirements.**.

The data analysis must follow and use the item numbering of each assignment, i.e., use the numbers, say, 1 - 5, with the sub-lettering if used. These deliverables are provided according to the instructions in the Submission Directions section below.

Submission Directions

Title Page

Include a title page with your name and the assignment designation. Leave room for instructor comments.

File Names

The assignment write-up file shall be submitted to Canvas according to the schedule in the syllabus using the item (1) naming convention below. The naming convention is case sensitive. Use letters and numbers as given. **The file name parts have no spaces or other separator characters**. TS5Lastname.pdf (submit via Canvas)

The parts are the assignment code, TS5; your last name with only the first letter capitalized; a period, and lastly, the extension "pdf". Generically,

TS5Lastname.pdf

For example: Suppose your name is Student McStats. Your filename then is:

TS5Mcstats.pdf

The analysis write-up file must be submitted for grading. Each write-up requires a title page for instructor comments. The analysis may use either R or any other statistics package you wish, or if you use more than one package, you must use the germane tables, plots, etc., in a single report. If you use more than one package, differences and similarities should be indicated.

email

ONLY IF REQUESTED email your instructor the program (script or code), log and output as separate pdf files. The R log and output may be combined. The file names shall be as follows:

- The program or script file names
  - TS5LastnameRprog.pdf
- The log file names
  - TS5LastnameRlog.pdf