Introduction to R Programming

Slide Set 9: Application 4 - Time Series

Maria Ptashkina

Barcelona GSE ITFD

September 2021

Learning Objectives

- Content
  - Analysing time series data
  - Building autoregressive models
  - Forecasting
  - Simulating series
- R-specific learning objectives
  - Familiarizing with time series objects in R
  - Writing own functions
  - Simulating processes

Task 1: Load and Analyse the Series

- The dataset we will be using in this session contains quarterly data on US real (i.e. inflation adjusted) GDP from 1957 to 2013
- Load the dataset 'USMacroSWQ.RData' and quickly explore it
- Change the date column to YEAR Q format (for example, 1957 Q1)
- Adjust the column names
- It is useful to work with time-series objects that keep track of the frequency of the data and are extensible
- In `R` we use objects of the class `xts`, see `?xts`
- Store GDP as an `xts` object, and choose the data from 1960 to 2013
- Calculate GDP growth rates and declare them as an `xts` object

Task 1: Load and Analyse the Series (cont.)

- Plot log of US quarterly real GDP
- Plot real GDP growth rates
- We would often need to report the time series, its logarithm, the annualized growth rate and the first lag of the annualized growth rate series
- Write a function in R which would produce and return this set of values, and apply this function to GDP series for the period 2012:Q1 - 2013:Q1
- Note that annual the growth rate is computed using the approximation

$$AnnualRateY_t = 400 * \Delta \log Y_t$$

since $100 * \Delta \log Y_t$ is an approximation of the quarterly percentage changes

Task 1: Load and Analyse the Series (cont.)

- Observations of a time series are typically correlated. This type of correlation is called autocorrelation or serial correlation
- Compute the first four sample autocorrelations of the series `GDPGrowth`
- Plot the first 20 lags of the series `GDPGrowth`

Task 2: First-Order Autoregressive Model

- An autoregressive model relates a time series variable to its past values
- The immediate past of a variable should have power to predict its near future, thus the simplest autoregressive model uses only the most recent outcome to predict future values
- First-order autoregressive model, AR(1)

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$$

- The first-order autoregression model of GDP growth can be estimated by computing OLS estimates in the regression of $GDPGR_t$ on $GDPGR_{t-1}$

Task 2: First-Order Autoregressive Model (cont.)

- Subset the data such that you use data from 1962 to 2012 to estimate the model
- Estimate the AR(1) model of GDP growth using `ar.ols()` function
- Check that the computations done by `r.ols()` are the same as done by `lm()`
- Use `coeftest()` to obtain a robust summary on the estimated regression coefficients

Task 3: Forecasts and Forecast Errors

- Suppose that Y_t follows an AR(1) model with an intercept and that you have an OLS estimate of the model on the basis of observations for T periods
- Then you may use the AR(1) model to obtain $\hat{Y}_{T+1|T}$ using data up to period T, where

$$\hat{Y}_{T+1|T} = \hat{\beta}_0 + \hat{\beta}_1 Y_T$$

- The forecast error is

$$Error = Y_{T+1} - Y_{T+1|T}$$

- Perform the forecast for GDP growth for 2013:Q1 (remember that the model was estimated using data for periods 1962:Q1 - 2012:Q4, so 2013:Q1 is an out-of-sample period)
- Calculate the R-squared and the root mean squared forecast error

Task 4: Autoregressive Model of Order p

- An AR(p) model incorporates the information of p lags of the series
- Estimate an AR(2) model of the GDP growth series from 1962:Q1 to 2012:Q4
- Calculate the R-squared and the root mean squared forecast error
- use the AR(2) model to obtain a forecast for GDP growth in 2013:Q1 in the same manner as for the AR(1) model

5. Nonstationarity (Trends)

- If a series is nonstationary, conventional hypothesis tests, confidence intervals and forecasts can be strongly misleading
- One type of non-stationarity is when a series exhibits a trend
- A formal test for a stochastic trend has been proposed by Dickey and Fuller (1979) - you will learn about it in Econometrics classes!
- Use the DF test to assess whether there is a stochastic trend in U.S. GDP using the regression

$$\Delta \log(GDP_t) = \beta_0 + \alpha t + \beta_1 \log(GDP_{t-1}) + \beta_2 \Delta \log(GDP_{t-1}) + \beta_3 \Delta \log(GDP_{t-2}) + u_t$$

- The ADF test can be done conveniently using `ur.df()` from the package `urca`

Task 6: Simulations and Spurious Correlation

- A way to model a time series Y_t that has stochastic trend is the random walk with a drift

$$Y_t = \beta_0 + Y_{t-1} + u_t$$

  where $E(u_t | Y_{t-1}, Y_{t-2}, \dots) = 0$
- Simulate four random walks with a drift in `R` using `arima.sim()`
- Use `matplot()` function for simple plots of the columns of a matrix
- Plot only the simulated series two and three, and observe the spurious correlation
- Imagine you did not have information on the fact that the series were not related to each other, and run a regression of series two on series three. What do the results tell you?

References and Resources

- Introduction to Time Series Regression and Forecasting ▸ Tutorial
- Time Series: A Data Analysis Approach Using R ▸ Examples
- Using R for Time Series Analysis ▸ Tutorial
- ETC3550: Applied Forecasting for Business and Economics ▸ Lectures
- R Graphics Essentials ▸ Tutorial
- R Time Series Quick Fix ▸ Tutorial