FEB23001(X) Assignment 2

Annika Camehl

Time Series Analysis

Instructions

- This assignment should be made with groups of three students. It is not allowed to cooperate with other groups, or to copy their results.
- The results of this assignment should be written down in a solution report consisting of max. 5 pages (not including a possible appendix which can contain part of the tables and graphs). Answer each question in a separate section (clearly indicating the number of the question).
- Do not include your code!
- Reports need to be submitted via Canvas Assignments as a PDF file with the name "`A2-XXX-YYY-ZZZ-.pdf`", where `XXX`, `YYY` and `ZZZ` are the student numbers of the three group members.
- Motivate your answers! For example, answers like Yes or No will not be rewarded any points.
- Some questions may not be specified up to the last detail and require you to make some choices yourself. Explain and motivate your choices in your report.
- It is recommended to use RStudio (Desktop) for programming.

Questions

The CSV file `RGPDIC` contains quarterly observations on real gross private domestic investment for the United States (billions of Chained 2012 Dollars) over the period Q1 1947 until Q1 2023. We denote the series by y_t.

Before you start install the following packages with

```
install.packages("readr")   # for data import
install.packages("ggplot2") # for plotting
install.packages("tseries") # for Jarque-Bera test
install.packages("urca")    # for ADF test
install.packages("forecast")# for prediction
```

and load them

```
library(readr)
library(ggplot2)
library(forecast)
library(tseries)
library(urca)
```

Trends

1. (1 point) Import the time series and plot the data. Is a deterministic trend present in the data?
2. (2 points) Apply the Augmented Dickey-Fuller (ADF) test to examine the presence of a unit root in y_t. Include in the test equation an intercept term and select the number of included lags based on AIC. Is a stochastic trend present in the data?
3. (2 points) Apply the ADF test to examine the presence of a unit root in y_t including additionally a deterministic trend in the test equation. Compare the test results to the results obtained in (2.). Given the properties of the time series, which specification of the test regression is preferable (which deterministic components do you include in the test regression)?
4. (1 point) Estimate AR(p) models for $p = 1, \ldots, 12$ for y_t including an intercept and deterministic trend. Which model does the AIC indicate as the preferred choice?
5. (3 points) Estimate an AR(2) model with intercept and deterministic trend for y_t. Discuss the estimation output. What conclusions regarding stationarity can you draw from the estimated parameters?
6. (2 points) Evaluate the estimated AR(2) model based on the estimated residuals. Are the residuals following a White Noise process? Test whether they are autocorrelated (using a Ljung-Box test) and also whether they follow a normal distribution.
7. (2 points) Take the first difference of y_t. Denote the new series as Δy_t. Plot the time series and discuss all key properties of time series data for Δy_t based on visual inspection.
8. (1 point) Apply the ADF test to examine the presence of a unit root in the Δy_t. Given the properties of the time series, which deterministic components do you include in the test regression? How do you interpret your test results?
9. (1 point) Compute a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for the null hypothesis that y_t is level or trend stationary. Then, compute the KPSS test for the null hypothesis that Δy_t is level stationary. How do you interpret your test results?

   You can use the function `kpss.test` of the package `tseries` and set the option `lshort = TRUE`.

Seasonality

10. (2 points) The CSV file `RGPDICnoSA` contains quarterly observations on real gross private domestic investment for the United States over the period Q2 2003 until Q1 2023 which are not seasonally adjusted. We denote the series by y_t^*.

    Import the new time series and plot the data. Does this time series exhibit a seasonal pattern? Regress y_t^* on a time trend and four seasonal dummy variables. Discuss the regression output.
11. (2 points) Estimate an AR(4) model including an intercept and deterministic trend for y_t^*. Compute the fitted values obtained from the AR(4) model. How accurate are the fitted values compared to the actual data y_t^*? You can answer this by inspecting a plot of the fitted values and actual data.
12. (3 points) Compare the out-of sample forecasting performance of the AR(4) model for y_t^* to a random walk model with a deterministic trend based on the mean squared forecast errors. Compute one-step ahead forecast for the last 50 observations based on a rolling window. Test whether the difference in forecast performance of the two models is statistically significant. Which of the models is preferable in terms of forecasting accuracy?

ADF-test

13. (4 points) In order to understand the behavior of the ADF test better conduct a small simulation to see the dependence of rejection/no-rejection decision on the value of the autoregressive parameter in an

AR(1) model. To do so, run 1000 simulations in which you simulate a time-series with 100 observations based on the following data generating process (DGP):

$$\begin{aligned} y_t &= \phi y_{t-1} + \epsilon_t \\ y_0 &= 0 \\ \epsilon_t &\sim \mathcal{N}(0,1). \end{aligned}$$

Repeat the 1000 simulations for three different values of ϕ. Set (1) $\phi = 0.5$, then (2) $\phi = 0.9$, and finally (3) $\phi = 0.99$. Hence, you perform 3 times 1000 simulations and in each simulation you generate a time series of length 100 according to the DGP with a specific ϕ. In each simulation draw, calculate the value of the ADF test statistic. Select the setting of the ADF test accordingly based on the knowledge of the DGP. Plot the values of the test statistics for all simulation draws in a histogram. Compare those to the critical values. What do you conclude regarding the performance of the ADF test based on the value of the autoregressive parameter in an AR(1) model?

14. (4 points) Repeat the simulation for a model with a random walk. Run again 1000 simulations generating data from the following DGP

$$\begin{aligned} y_t &= y_{t-1} + \epsilon_t \\ y_0 &= 0 \\ \epsilon_t &\sim \mathcal{N}(0,1). \end{aligned}$$

In each simulation draw, calculate the value of the ADF test statistic. Select the setting of the ADF test accordingly based on the knowledge of the DGP. Plot the values of the test statistics in a histogram. Compare those to the critical values. What do you conclude regarding the performance of the ADF test? Compare you findings to those of the question before.