FEB23001(X) Assignment 3

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. 10 pages (not including a possible appendix). All essential figures and tables need to be in the main text. 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 "`A1-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 `GDPEA` contains quarterly observations on gross domestic product for the Euro Area (Millions of Chained 2010 Euros) over the period Q1 1995 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("forecast")    # for prediction
install.packages("urca")        # for ADF test
install.packages("tseries")     # for Jarque-Bera test
install.packages("tsDyn")       # for TAR models
install.packages("strucchange")# for nonlinearity test
```

and load them

```
library(readr)      # import data
library(ggplot2)    # for plotting
library(forecast)   # for AR model
library(urca)       # for ADF tests
library(tseries)    # for Jarque-Bera test
```

```
library(tsDyn)       # for TAR models
library(strucchange)# for nonlinearity test
```

1. (1 point) Import the time series data. Create a time series of quarterly growth rates of GDP by taking the first differences of the log-levels and multiply by 100. Denote this series by Δy_t. Are the quarterly growth rates normally distributed?

2. (2 points) Repeat question (1.), but using the sample Q1 1995 to Q3 2008 and (b) Q1 1995 to Q4 2019. What are the most striking differences compared to (1.)? What does this tell you about the observations in Q4 2008, Q1 2009, and in 2020? Also plot the time series over the full sample to argue about those observations.

   To obtain the subsamples you can specify, for example, the end date as

   ```
   T2 <- which(Data$DATE == '2008-07-01')
   ```

   to stop in Q3 2008.

3. (3 points) Estimate an AR(p) model for Δy_t selecting the lag length based on AIC (with a maximal lag length of 8). Evaluate the model by inspecting the properties of the residuals by applying tests for normality and no-autocorrelation. What kind of patterns do you observe in the residuals around Q1 2020 to Q4 2020?

4. (2 points) Test whether the observations in Q4 2008, Q1 2009, Q1 2020, Q2 2020, and Q3 2020 are innovation outliers. What do you conclude from the test results?

   You can construct dummy variables for the outliers by

   ```
   z       <- matrix(0,T,5)
   z[which(Data$DATE == '2008-10-01'),1] <- 1   # Q4 2008
   z[which(Data$DATE == '2009-01-01'),2] <- 1   # Q1 2009
   z[which(Data$DATE == '2020-01-01'),3] <- 1   # Q1 2020
   z[which(Data$DATE == '2020-04-01'),4] <- 1   # Q2 2020
   z[which(Data$DATE == '2020-07-01'),5] <- 1   # Q3 2020
   ```

5. (2 points) Repeat question (3.) but treating the observations Q1 2009, Q1 2020, Q2 2020 and Q3 2020 as outliers. Estimate an AR(2) model treating these observations as innovation outliers. Evaluate the model by inspecting the properties of the residuals testing for normality and no-autocorrelation. Compare the properties of the residuals to those obtained in question (3.), also comparing the plotted residuals.

6. (2 points) Compare the in-sample fit of the previously estimated AR(1) model and the estimated AR(1) model treating the observations Q1 2009, Q1 2020, Q2 2020, and Q3 2020 as innovation outliers based on the value of the log likelihood and the AIC. Based on those in-sample fit measures which model would you select?

7. (2 points) The CSV file `IEA` contains quarterly observations on investment in growth rates for the Euro Area over the period Q1 1995 until Q4 2022. We denote the series by x_t. Import the time series and plot the data together with Δy_t. What do you observe? What is the correlation between the two series?

8. (3 points) Estimate an AR(p) model for x_t selecting the lag length based on AIC (with a maximal lag length of 8). Next, estimate an AR(p) model for x_t including Δy_t as an additional regressor. Select the lag length of the model based on AIC (with a maximal lag length of 8). Evaluate the two models in the usual way, by considering the properties of the residuals. Is the second model an improvement over the AR(p) model for x_t?

9. (3 points) Estimate a threshold model for x_t with an AR(1) model in both regimes, with Δy_t as the threshold variable, and with the threshold fixed at zero. Discuss the differences between the AR(1) models in the two regimes based on the estimated coefficients. Also, explain the interpretation of setting the threshold fixed at zero for the specified threshold variable.

To estimate the threshold model you can use the function `setar` of the `tsDyn` package.

10. (2 points) Estimate the same threshold model but now with the threshold as an unknown parameter (instead of fixing it at 0). Discuss the differences between the results for the threshold model with fixed threshold at zero and the threshold model with unknown threshold.

11. (2 points) Estimate a logistic smooth transition model (including an intercept and one lag of x_t and using Δy_t as threshold variable). Estimate the smoothing parameter (γ) and threshold parameter (c). Discuss the differences between the results for this model and the threshold model with unknown threshold.

12. (2 points) Use the *SupF* approach to test for threshold nonlinearity in the threshold model, but with the threshold as an unknown parameter (instead of fixing it at 0). Discuss the test results.

    You can use the function `Fstats` and `sctest` of the `strucchange` package to test for nonlinearity (use the type `supF`).

13. (4 points) Use the threshold model with unknown threshold and the logistic smooth transition model to obtain fitted values. Evaluate the relative accuracy of these in-sample forecasts based on mean squared prediction errors. Which of the nonlinear models delivers more accurate forecasts? Are potential improvements in forecast accuracy statistically significant? Discuss also the limitation of comparing the models only based on in-sample fit.