Emprical Exercises - 3
Forecasting the Price of Owner-Occupied Apartments Using ARMA

Applied Statistics and Econometrics II
Spring 2018, NYU
Ercan Karadas

(Due April 20)

Motivation: House prices are a hotly debated topic. Their tendency to undergo booms and busts is well-described in the news with frequent stories about recent developments, regional differences, etc.. The price of housing inherently depends on expectations of the future price. As buying a house or an apartment is typically the most important economic decision for individuals and families, forming reasonable expectations about their future developments is crucial. Moreover, house prices are important for the overall economic activity, and thereby employment, as booms and busts can have a major impact on aggregate private consumption and savings.

An important use of econometric time series models is to make out-of-sample forecasts of future economic developments upon which economic decisions can be made. While univariate dynamic time series models, such as an autoregressive model, have limitations, they are easy to estimate, serve as a good benchmark, and they have shown to be hard to beat in practice.

Exercise 1: Based on a sample of quarterly data covering the period 1990:1 - 2017:3, the aim of this assignment is to estimate a univariate autoregressive (AR) model and use the model to forecast out-of-sample (log) changes in a price index for owner-occupied apartments in the US.

Data: Download the dataset `timeSeriesUSA.csv` from the course webpage on NYU Home. The dataset contains the following quarterly time series for the US economy between 1990:1 - 2017:3:

EXNOM	Effective (nominal) exchange rate (1980=100).
CANOM	Current account (current prices).
INVBUSS	Business machinery investments (constant prices).
GDPREAL	Total GDP (constant prices).
BOND	Bond yield (pro anno yield in ratio).
HOUSE	House price index (2000=1).
CONNP	Consumption deflator (2000=1).
URATE	Unemployment rate (in ratio).
TT	Terms of trade, export price/import price (2000=1).

[1] Select the time series you want to analyze. It must contain KP and at least one more variable. It could be a raw series or a transformation, such as the real house price or

the (ex post) real interest rate. Discuss if it is preferable to analyze the log of your chosen time series.

[2] Draw a graph of your chosen time series. If the time series does not look stationary from a visual inspection, take first differences and construct a new graph.

[3] Draw the autocorrelation function, ACF, and the partial autocorrelation function, PACF, for your time series. What does that suggest about the dynamic properties of the time series?

[4] Now estimate the relevant candidate ARMA models.

[5] Construct a table with information criteria and other relevant information. Which model do you prefer?

[6] Reestimate the preferred model and look at graphical output for the residuals, e.g. residual ACF, PACF, and histogram/density. Also calculate the available misspecification tests. What do you conclude concerning the fit of the model?

[7] Finally reestimate the model but retain 24 observations for forecasts. Fore- cast the 24 periods and compare the forecasts with the actual values.

[8] In your own words, characterize the nature of a univariate time series fore- cast. What do you conclude regarding the forecast performance of the model?

Hints:

- You are only required to assess stationarity of the variables based on the graphs of the data.
- For the graphical analysis, use any transformations of the variables you find are relevant (among the transformations proposed above, or any other ones), and explain your choices.
- When using the OLS or ML estimator, carefully discuss which assumptions are required and explain if they are fulfilled or not. If not, what are the consequences on your estimation results? Can you improve on this?
- If you come up with several model specifications, select the one that seems the most relevant to you, and justify your decision. Remember that in most cases, the simpler the better.
- If you perform some statistical tests, be precise about the hypotheses you test, why you test them, and what your conclusions are.
- Consider whether the model used for out-of-sample forecasts should be estimated based on the entire sample period.

Exercise 2: It is important to have an idea on how the appearance of an ARMA time series depends on the coefficients. This exercise simply asks you to generate a number of time series in R and look at them!

For example, to generate $T = 100$ observations from the ARMA(2,1) model

$$y_t = 0.5y_{t-1} - 0.2y_{t-2}\epsilon_{t-1} + 0.1\epsilon_{t-1}$$

the following R code could be used:

```
arma = arima.sim(list(order=c(2,0,1), ar=c(.5,-.2), ma=.1), n=100)
acf2(arma)
```

The second line gives both ACF and PACF of the generated ARMA process. But note that you need to install **astsa** package first in order to use `acf2()` function.

[1] First focus on first order autoregressive models

$$y_t = \theta_1 y_{t-1} + \epsilon_t$$

Look at the properties of the time series for $\theta_1 \in \{-0.8, 0, 0.5, 0.9, 1.0, 1.05\}$. Why do you think that the theoretical ACF is not reported for some values of θ_1?

Note that the estimated ACF is poor for some values of θ_1. Try to increase the number of observations to e.g. $T = 1,000$ or $T = 10,000$.

[2] Now look at first-order moving-average processes for $\alpha_1 \in \{-0.8, 0, 0.8\}$

$$y_t = \epsilon_t + \alpha_1 \epsilon_{t-1}$$

[3] Look at some ARMA processes of your own choice. Look at an ARMA(1,1) model with $\theta_1 = \alpha_1$. Explain what happens?

[4] Generate $T = 100$ observations from the ARMA(2,1) model above *w*ithout using built-in R functions.

Formalities:

[1] You must hand in a report that (i) presents your graphical analysis, (ii) describes the econometric model, (iii) outlines the modeling progress (e.g., the approach you have taken, the alternative models you have tried, etc.), (iv) presents your preferred model including interpretation and statements on economic and statistical significance, and (v) discusses the potential weaknesses of the model.

[2] The report must be a maximum of 10 pages of pdf produced from an R markdown file. You must hand in your R markdown file and pdf files together. You can use the R markdown template that I have posted on course webpage. Please name your files as follows: `empEx3-surname(s).Rmd`

[3] If you prefer, you are allowed to work in groups of up to three persons (not necessarily in the same exercise class as yours). The requirements and assessment criteria are the same for assignments written by one, two, or three persons.