Financial Econometrics Rauli Susmel
FINA 4397

Homework 3 (Due November 30)

1.1 (**ARMA Processes**). Consider the following three models that a researcher suggests might be a reasonable model of stock market prices, where ε_t is a white noise error process:

$$y_t = \varepsilon_t + .5\,\varepsilon_{t-1}$$
$$y_t = .8\ y_{t-1} + \varepsilon_t$$
$$y_t = .1 + y_{t-1} + \varepsilon_t$$

(a) What classes of models are these examples of?
(b) What would the autocorrelation function for each of these processes look like? Calculate the first three ACFs.
(c) Which model is more likely to represent stock market prices from a theoretical perspective, and why? If any of the three models truly represented the way stock market prices move, which could potentially be used to make money by forecasting future values of the series?
(d) By making a series of successive substitutions or from your knowledge of the behavior of these types of processes, consider the extent of persistence of shocks in the series in each case.

1.2 (**ARMA Processes**). You obtain the following estimates for an AR(2) model of some returns data

$$y_t = .803\,y_{t-1} + .682\,y_{t-2} + \varepsilon_t$$

where ε_t is a white noise error process.
a. Check the estimated model for stationarity.
b. What is the meaning of stationarity?
c. Without calculating the ACF and PACF, describe what are the expected patterns?
d. Produce one-, two-, and three-step ahead forecasts.

1.3 (**Significance of ACF/PACF**). Consider the following autocorrelation and partial autocorrelation coefficients estimated using 500 observations for a weakly stationary series, y_t:

Lag	acf	pacf
1	0.307	0.307
2	−0.013	0.264
3	0.086	0.147
4	0.031	0.086
5	−0.197	0.049

a. Use the Bartlett SE –i.e., 1/sqrt(T)-, to determine which, if any, of the ACF and PACF coefficients are significant at the 5% level.
b. Use the Ljung–Box statistics to test the joint null hypothesis that the first five autocorrelation coefficients are jointly zero.
c. What process would you tentatively suggest could represent the most appropriate model for the series y_t? Explain your answer.

1.4 (**ACF, PACF and Order of ARMA process**). Download a simulated dataset (sim.arima.csv) from my homepage. You can use the below code to download the data:
S_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/sim.arima.csv", head=TRUE, sep=",")

You have four simulated series: sim_1, sim_2, sim_3, sim_4.

a. Plot the ACF/PACF for sim_1.
b. Using the ACF/PACF guess the order of the ARIMA process for all four series.
c. Check the residuals to see if there is any autocorrelation left in each of your series.

1.5 (Forecasting with ARMA and SES)
10. You have estimated the following ARMA(1,1) model for some time series y_t:

$$y_t = 0.036 + .69\, y_{t-1} + 0.42\, \varepsilon_{t-1} + \varepsilon_t$$

Suppose that you have data for time to *t*−1, i.e. you know that $y_{t-1} = 3.4$ and $\hat{\varepsilon}_{t-1} = -1.3$.

a. Obtain forecasts for the series y_t for times *t*, *t*+1, and *t*+2 using the estimated ARMA model.

b. If the actual values for the series turned out to be −0.032, 0.961, 0.203 for *t* , *t* +1, *t* +2, calculate the (out-of-sample) mean squared error.
c. A colleague suggests that a simple exponential smoothing (SES) model might be more useful for forecasting the series. The estimated value of the smoothing constant, α, is 0.15, with the most recently available smoothed value, $S_{t-1} = 0.0305$. Obtain forecasts for the series y_t for times *t*, *t*+1, and *t*+2 using this model.
d. Given your answers to parts (a) to (c) of the question, determine whether ARMA or exponential smoothing models give the most accurate forecasts in this application.

1.6 (Forecasting Evaluation)
Two researchers are asked to estimate an ARMA model for a daily USD/GBP exchange rate return series, denoted x_t . Researcher *A* uses Schwarz's BIC for determining the appropriate model order and arrives at an ARMA(0,1). Researcher *B* uses Akaike's AIC which deems an ARMA(2,0) to be optimal. The estimated models are

A: $\hat{x}_t = 0.38 + 0.10\varepsilon_{t-1}$

B: $\hat{x}_t := 0.63 + 0.17x_{t-1} - 0.09x_{t-2}$

where ε_t is a white noise error term.

You are given the following data for time until day *z* -i.e., $t = z$:

$$x_z = 0.31,\ x_{z-1} = 0.02,\ x_{z-2} = -0.16$$

$$\varepsilon_z = -0.02,\ \varepsilon_{z-1} = 0.13,\ \varepsilon_{z-2} = 0.19$$

a. Produce forecasts for the next four days (i.e. for times $z+1$, $z+2$, $z+3$, $z+4$) from both models.

b. Suppose that the actual values of the series x on days $z+1$, $z+2$, $z+3$, $z+4$ turned out to be 0.62, 0.19, −0.32, 0.72, respectively. Using MAE, determine which researcher's model produced the most accurate forecasts.

1.7 (Fitting ARIMA and Seasonality) From my homepage, download the data set Real_Estate_2023.csv, which you can download with the below R line:
RE_da <- read.csv("http://www.bauer.uh.edu/rsusmel/4397/Real_Estate_2023.csv", head=TRUE, sep=",")

You are going to analyze changes in US real estate prices (USA_c). You are going to estimate the model using the information from **1990:2** to **2020:12**. This is your estimation period. You will validate your model with the **2021:1** to **2023:7** data.
a. Using the function arima, in the package forecast, fit an ARIMA(3,0,0) model –i.e., an AR(3) model.
b. Extract your residuals from your fitted model. This is the filtered USA_c series . Run the filtered series against monthly dummies. Do you see seasonality?
c. If you have seasonality, filter again the series –i.e., extract residuals from regression in b- and check if now you need to review your ARIMA model.
d. Using the AR(3) model, forecast USA_c for the period **2021:1** to **2023:7**.

1.8 (Comparison of Forecasts) Continuation from 1.7. Suppose you keep your AR(3) model for USA_c.
a. Using auto.arima propose another model for USA_c, using only data from **1990:2** to **2020:12.**.
b. Forecast USA_c for the period **2021:1** to **2023:7**.
c. Compare your forecasts from 1.7 and 1.8, using the Diebold-Mariano test. (You need to calculate forecast errors.) Which model is better?