PSTAT274 Time Series

Lab Worksheet - Week 6

Alex Bernstein

This lab is due at 11:59pm on Friday, February 23th, 2024 and should be submitted as a pdf document via Gradescope.

Just like last week, make sure you have `astsa` and `forecast` loaded. We will be using the `sales` dataset, which is in the `astsa` package:

```
# load library
library(astsa)
library(forecast)
#sales data
head(sales)
```

Question 1: Exercise 3.35, Part 1

The `sales` Let S_t represent the monthly data in sales ($n = 150$)

1. Fit an ARIMA model to S_t and discuss your model fitting process. Discuss:

    a. Your exploratory data analysis
    b. transformations if necessary
    c. initial identification of dependence orders and degrees of differencing
    d. parameter estimation
    e. residual diagnostics and model choice

Note that there maybe more than one valid model, and it is up to your interpretation, so use your best judgement and defend it in your writeup. If you use `auto.arima`, you must compare your choice of top 3 models; set `trace=True` when you call it to identify them, and fit them individually with `sarima`. If you include a drift constant, make sure to report that as well (it will be in the output). Also, recall that `auto.arima` does not account for any transformations you make on the data.

Question 2: Exercise 3.35, Part 2

Now, we will also be using the `lead` dataset, also in `astsa`:

```
#lead data
head(lead)
```

Use the cross-correlation function (`ccf`) and lag plots between to argue that a regression of ∇S_t on ∇L_{t-3} is reasonable. (Use `lag2.plot(x,y)`, where `x` and `y` are time-series, and note that the first one, in this case `x` is the one that is lagged). In this case, even if you used a transformation, stick to the simple difference-series ∇S_t, and make sure you make a reasonable case for stationarity as we need stationarity of both ∇S_t and ∇L_t in order for a lagged regression to make sense. In `ccf`, note that the lagged series is the first one you enter, i.e. `ccf(L,S)` gives you the correlations of L_{t+h} and S_t.

Question 3: Exercise 3.35, Part 3

We are now going to fit a lagged model against another model. Note that because both `sales` and `lead` are 150 elements long, we can't simply fit a regression of ∇S_t on ∇L_{t-3} as the series are not aligned properly. In `R`, we can address this by using `ts.intersect` to create a dataframe with the lagged series, which we can then fit a model on, as follows:

```
df = ts.intersect(d.sales,d.lead_3 = lag(d.lead,3), dframe=TRUE)
fit_model <- lm(d.sales~d.lead_3,data = df,na.action=NULL)
```

Alternatively, we can just use the `dynlm` library, which takes care of all this for us:

```
#install.packages("dynlm")
library(dynlm)
fit_model_dyn<-dynlm(d.sales~L(d.lead,3))
```

1. Fit this model and report your model summary
2. Once you have this model fitted, examine the residuals of your fit using the `resid` function, in order to come up with a final model:

$$\nabla S_t = \beta_0 + \beta_1 \nabla L_{t-3} + x_t$$

where x_t is some ARMA (not ARIMA- we've already differenced both our original series) process and explain how you decide on your model for x_t. Discuss your results. (See Shumway and Stoffer (2017) Example 3.45, p. 147 for help on coding this)

References

Shumway, R. H., and D. S. Stoffer. 2017. *Time Series Analysis and Its Applications: With r Examples*. Springer Texts in Statistics. Springer International Publishing. https://books.google.com/books?id=sfFdDwAAQBAJ.