477/577 In-class Assignment 3 : Fitting ARIMA(p,d,q) to Wine Data

(due Tue 4/7/2020) 8pm

Submit your code file (.R or .txt extension).

- You must submit the code file (.R or .txt extension) on Brightspace.
- (1point) The file name must be

TS-A3-FirstLastname.R (or .txt)

- (1point) The code file must have title, and visible separator for questions. You can use the template posted on the website.
- Each block of the code must have brief comment of what it's doing.
- You are encouraged to collaborate with your classmate, but your submission must be your own work.
- **You can consult me by emailing nmimoto@uakron.edu. If your question is too complex for email, you can call me or webEx me too. Inquire by email.**

Here is the code to load the data from the web.

```
D  <- read.csv("https://nmimoto.github.io/datasets/wine.csv")
D1 <- ts(D, start=c(1980,1), freq=1)
plot(D1, type='o')
```

Now your "D1" in R contains monthly wine sales in Australia.

Preliminary Analysis

1. Does "D1" look like stationary time series? State your graphical observations and conclusions drawn from p-values form **Stationarity.tests()**.

2. Take difference of D1 using diff(), plot it and check the stationarity. State your graphical observations and conclusions drawn from p-values form **Stationarity.tests()**.

3. Should we take any transformation before differencing? Why or why not? If yes, use Box-Cox power transformation. Pick your value of λ.

For all problems below use your value of λ if you decided to use the transformation.

ARIMA(d=1) analysis

4. Use **auto.arima()** function to find best ARRIMA(p,d,q) model with constraint that d=1. What is the suggested model? (use stepwise=FALSE, approximation=FALSE option.) Does it pass the residual test for model adequacy? Copy and Paste the output from auto.arima() and Randomness.tests(). (not the plot)

5. Now we search for better ARIMA model without the guidance of AICc. Start with ARIMA(15, 1, 15) with the drift model, use Arima() function to estimate parameters. Reduce p and/or q if the last parameter in AR or MA is not significant. Stop if the LAST parameter of both AR and MA term is significant. Remove the drift if not significant.
   What is your final model? Compare AICc of your final model to the ones you got from (4). Which one is lower? Why did this model was not suggested in (4)? Does this final model pass the residual adequacy test?
   (Only include the output of your final model. Model pram + Residual p-values.)

ARIMA(d=0) with Linear Trend analysis

6. Another model we can fit this data is d=0 with linear trend model. Use auto.arima() with d=0 and xreg=time(D1) option to find best ARMA(p,q) model to go on top of the linear trend. Don't forget to use the same λ as before. What is your linear trend model? Does this final model pass the residual adequacy test?
   (Model pram + Residual p-values.)

7. Is the slope estimate you get in (6) consistent with the drift term you had in (5)?

8. Use the following code to fit the regression line outside of auto.arima(), and test the regression residuals for stationarity. Is the estimate consistent with (6)? (Replace lambda with your lambda)

```
D2 <- BoxCox(D1, lambda)
Reg <- lm(D2~time(D2))
summary(Reg)
plot(D2)
abline(Reg, col="red")
Stationarity.tests(Reg$residuals)
```

Compare the two model

9. Now we need to compare the models you got from (5) vs (6). Perform 12-step forecast using the model from (5). What is the 95% CI for the next observation? Include the numbers here.

10. Perform 12-step forecast using the model from (6). What is the 95% CI for the next observation? Include the numbers here.
    (Remember that your forecast() needs xreg. See slide (6-5))

11. Perform Rolling 1-step prediction of last 42 observations retrospectively using model from (5). Report prediction rMSE. Compare that with sigma-hat from the model.

12. Perform Rolling 1-step prediction of last 42 observations retrospectively using model from (6). Report prediction rMSE. Compare that with sigma-hat from the model.

13. Which model do you like better and why? Model from (5) or (6)? Write down your mathematical model, and list estimates for all parameters. You can type using following notation. (you may not need to use some of them)

Observation: Y_t, Y_{t-1}, Y_{t-2}, …
X_t, X_{t-1}, X_{t-2}, …
e_t, e_{t-1}, e_{t-2}, …
WN(0, sigma^2)
a+bt
(phi_1) (phi_2) …
(theta_1) (theta_2) …
Difference operator: (del)
Backwards operator: B