ECO 5375-701 Prof. Tom Fomby
Eco and Bus Forecasting Fall 2016

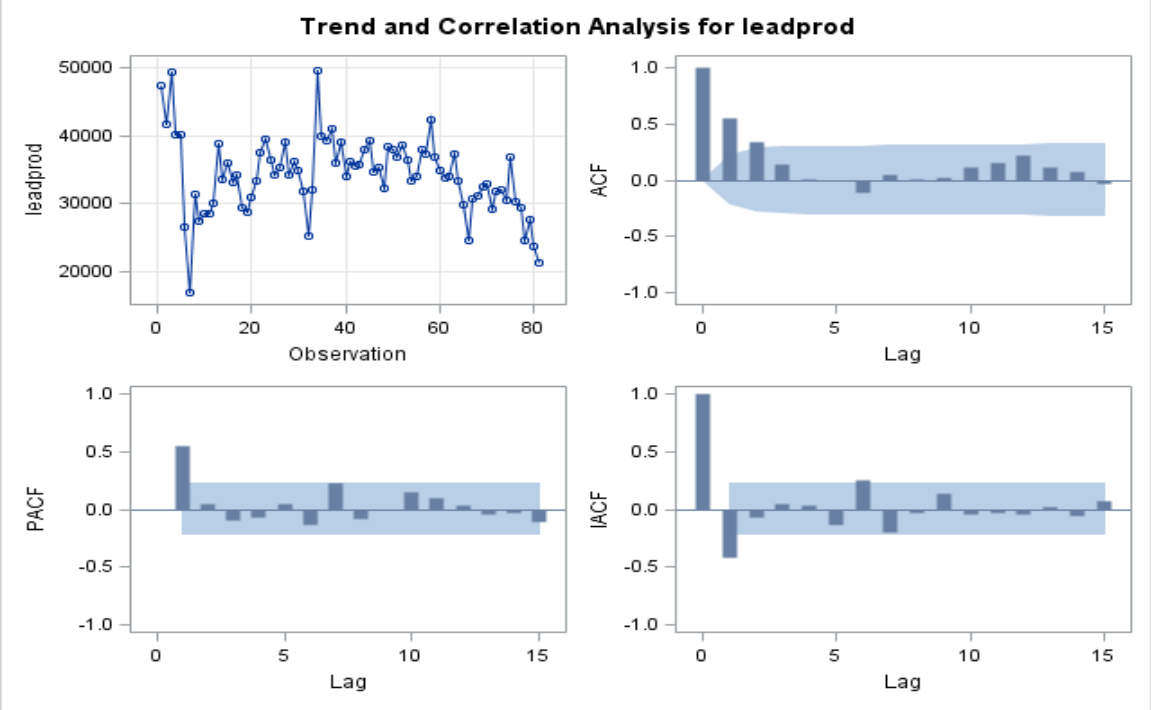
EXERCISE 5
KEY

Purpose: To learn how to use the **sample autocorrelation function** and **sample partial autocorrelation function** as well as the **P-Q Box** to help determine the **best** Box-Jenkins model for the Lead Production data. Also we will use the estimation of **overfitting models** and the **significances of the overfitting coefficients** to further confirm our choice of the **best** Box-Jenkins model for the Lead Production data. This homework is due **Thursday, September 29.**

Download the **Leaddata.txt** data and cut and paste it into a SAS program that you might call Leadprod.sas. Add the necessary SAS code to complete the rest of this exercise.

(i) Print out and hand in with this exercise the ACF and PACF of the lead production data and a plot of the data. Make a tentative **choice** of p and q for a Box-Jenkins model of the data. Thoroughly explain your reasoning. Consult the document ACF_PACF_Table.pdf.

Answer: The sample ACF appears to be damping out while the sample PACF has one spike in it and then cuts off. This is indicative of the AR(1) model (p=1, q=0). See ACF_PACF_Table.pdf and the below output:



Given these graphs, let's examine our tentative choice of model by constructing the P-Q Box. See the next part of this exercise.

(ii) Complete the below P-Q Box (a total of 6 cells). You can assume the data is already stationary. Given the statistics that you report, which model do you prefer? Explain your reasoning. For information, see Stats.pdf on the course website.

Answer: Here is the P-Q Box for the Lead Production data. The AR(1) model (p=1, q=0) has the smallest goodness-of-fit measures (AIC, SBC) and the lag = 24 Box-Pierce Q statistic has a p-value = 0.7491 > 0.05 indicating that the residuals of the AR(1) model are white noise. This makes the AR(1) model a good tentative choice for further examination vis-à-vis overfitting.

P|Q	0	1	2	3
0	1628.977 1631.372 53.39 (0.0005)	1608.91 1613.699 26.16 (0.2936)	1601.337 1608.52 16.20 (0.8060)	
1	1598.93 1603.719 18.15 (0.7491)	1600.633 1607.816 17.04 (0.7614)		
2	1600.531 1607.714 16.61 (0.7843)		Legend: AIC SBC Q-statistic (lag=24) (p-value)	
3				

(iii) Conduct overfitting exercises for the model preferred by the P-Q Box.

Overfitting Model 1 is ARMA(**2**,**0**).
The overfitting coefficient is **0.07346**.
The T-statistic of the overfitting coefficient is **0.62**.

Therefore the overfitting coefficient from this model is
statistically (significant/ **insignificant**). Circle one alternative.

Overfitting Model 2 is ARMA(**1**,**1**).

The overfitting coefficient is **0.09095**.
The T-statistic of the overfitting coefficient is **0.46**.
Therefore the overfitting coefficient from this model is
statistically (significant/ **insignificant**). Circle one alternative.

(iv) Given all of the above results what is your final choice for p and q?

P = **1** Q = **0**.

In the below space, write out the final chosen model in both "deviation-from-mean" form and "intercept" form.

Deviation from Mean Form:

$$(y_t - 34558.3) = 0.59452(y_{t-1} - 34558.3) + \hat{a}_t$$
$$(1229.1) \quad (0.09554) \quad (1229.1)$$

AIC = 1598.93
SBC = 1603.719
Q = 18.15
(0.07491)

Intercept Form:

$$y_t = 14012.79 + 0.59452 y_{t-1} + \hat{a}_t$$
$$(0.09554)$$

AIC = 1598.93
SBC = 1603.719
Q = 18.15
(0.07491)