DB 1 – Week 7

A researcher uses a sample of 200 quarterly observations on Y\_t , the number (in 1000s) of unemployed persons, to model the time series behavior of the series and to generate predictions. First, he computes the sample autocorrelation function with the following results:  
  
(k, rho)   –>   (1,0.83); (2,0.71); (3,0.60); (4,0.45); (5,0.44); (6,0.35); (7,0.29); (8, 0.20); (9,0.11); (10, -0.01)

What do we mean by the simple autocorrelation function?

Autocorrelation is similar to the R-squared in that it measures the correlation between the present (dependent variable) and the past (independent variable) lags times. It is logical to assume that the most recent lag time is the most correlated to the dependent variable.

When an autocorrelation function diminishes over time, as seen above, it is the first hint that the correct model to use would be the simple autoregressive model. In addition, the largest time lag value is the first time lag, which points to the autoregressive model. In my mind, the data is acting in a very normal way such that the most recent lag time is the best predictor of the present value.

Next, a sample partial autocorrelation function is determined. It is given by:  
  
(k, theta)   –>   (1,0.83); (2,0.16); (3,-0.09); (4,0.05); (5,0.04); (6, -0.05); (7, 0.01); (8,0.10); (9,-0.03); (10,-0.01)

What do we mean by the sample partial autocorrelation function (PACF)? Why is the first partial autocorrelation equal to the first autocorrelation coefficient (0.83)?

The PACF values are coefficients that equal the lagged values. The PACF values are similar to the ACF values in that they fall between -1 & 1, and indicate which time lag correlates the highest with the response variable, which is the present value of the variable.

I am just as stumped as John in regard to why the PACF and ACF are the same, but here is what I have found.

In regression, this partial correlation could be found by correlating the residuals from two different regressions:

(1) Regression in which we predict y from x1 and x2, (2) regression in which we predict x3 from x1 and x2.  Basically, we correlate the “parts” of y and x3 that are not predicted by x1 and x2 (PennState).

What I take this to mean is that the first lag regressed on the present value will equal the partial correlation coefficient, but this really only make sense for this example…

Does the above pattern indicate that an autoregressive or moving average representation is more appropriate? Why?

From this example, the AR model is appropriate based on the fact that the model recedes overtime and there is only one major spike in the data. In addition, it is the first order lag.

https://onlinecourses.science.psu.edu/stat510/?q=node/62