Quiz 7

04/15/2019

Instructions:
Write your name at the top right. You are to work on this quiz alone without any help from any other resource *except for a single* 8.5×11 *inch page of handwritten notes.*

Problem 1: Describe three common situations (types or structures of data) in which a linear model would have correlated errors, and that the presence of which would suggest using generalized least squares.

The following are three typical data types/ structures that may have correlations in the error, and that we have discussed in the class:

- time series data with temporal (auto-)correlations;
- spatial data with spatial correlations;
- batch collected data with correlations within each batch.

Problem 2: Suppose that for some $\boldsymbol{\beta} \in \mathbb{R}^2$ and $\boldsymbol{\epsilon} \sim N(0, \boldsymbol{\Sigma})$, the following equation holds

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{1}$$

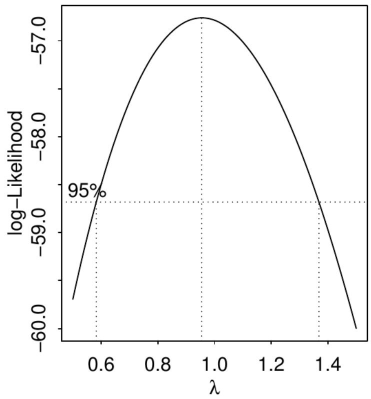
Suppose that $\boldsymbol{\Sigma}$ is of the form,

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix}. \tag{2}$$

Does the first or second observation receive more weight in weighted least squares? Why?

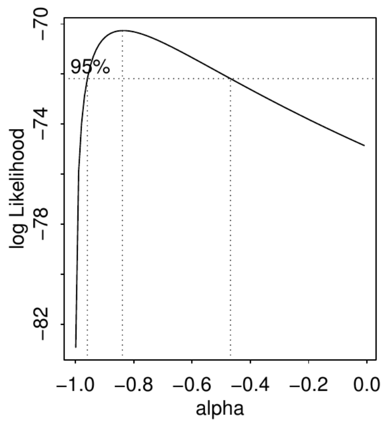
The variance of the first observation is 1 while the variance of the second observation is 10. The observations with higher variance receive less weight in the regression in weighted least squares, such that the first observation will be given more weight than the second.

Problem 3: Below is a plot of the log-likelihood versus the parameter λ with the corresponding 95% confidence interval for the Box-Cox transformation. Explain what is the null hypothesis for this confidence interval and if we can reject it at 5% significance.



The null hypothesis for the Box-Cox transformation is that the exponent λ should be equal to one — this corresponds to the transformation of the response $y \mapsto y$. Here, $\lambda = 1$ is within the 95% confidence interval, so that we cannot reject this with 5% significance.

Problem 4: Below is a plot of the log likelihood versus the parameter α and the corresponding 95% confidence interval for the shifted-log transformation. Does this plot suggest that a log transformation of the variable is appropriate without a shift? Why or why not.



The value α in the horizontal axis is the value that we will shift by in the shifted-log transformation. Particularly, the value $\alpha = 0$ is excluded from the 95% confidence interval, such that a standard log transformation doesn't appear to be reasonable (though a negatively shifted one does).