

Assignment 5

GCM 2010

Instructions. Answer the following, and be prepared to present and discuss your answers in class. This assignment is due April 10.

1. On page 428, Singer and Willett introduce data from Wheaton, Rozell, and Hall (1997), who examined the link between stressful life experiences and onset of depression. The authors state “our first step is to select a specification for the main effect of TIME.” They fit the completely general specification, and compare it against a range of polynomial specifications. At the bottom of page 428, they announce their choice of a cubic model, and give some model comparison statistics. However, full details of the analysis are not given.

What I would like you to do is the following:

- (a) Load the file *depression_pp.csv*. It includes a number of variables that actually are not necessary to perform the analysis. In particular, the file includes square and cubed values of the re-centered time variable (*age_18*). In using HLM, for example, this might be necessary, but in R, using the `glm` package, it is not. (If you are not sure why it is not, check out the meaning of the statement $I(\text{age_18}^2)$ in the context of a model specification!)
- (b) Fit the general model and linear and polynomial models up to the 4th order (quartic).
- (c) Compute the chi-square difference tests, using the `anova` command, and explain why Singer and Willett chose the cubic model over the quadratic and quartic.
- (d) Construct a table like the one in the lecture notes, that includes both the AIC and BIC statistics for all the models. (Hint: With the BIC, be *careful* about the value of N that you use, and consult your textbook if you need advice.) What models do the AIC and BIC select?