Intro to Econometrics Section 011 HW 4 — Due: Wednesday, Mar 9 at the beginning of class

Prepare brief and precise answers to the following questions. You are encouraged to discuss the solutions in groups but should write up the solutions independently.

1. Let C_t be consumption and X_t be a predictor of consumption. Suppose you have quarterly data on C and X. Let D_{1t}, D_{2t}, D_{3t}, and D_{4t} be dummy variables such that D_{1t} takes the value 1 in quarter 1 and 0 otherwise, D_{2t} takes the value 1 in quarter 2 and 0 otherwise, etc. Which of the following, if any, suffer from perfect multicollinearity and why?

   a) $C_t = \alpha + \beta X_t + \gamma_1 X_t D_{1t} + \gamma_2 X_t D_{2t} + \gamma_3 X_t D_{3t} + \gamma_4 X_t D_{4t} + u_t$
   b) $C_t = \alpha + \beta X_t + \gamma_1 X_t D_{1t} + \gamma_2 X_t D_{2t} + \gamma_3 X_t D_{3t} + \gamma_4 X_t (1 - D_{1t} - D_{2t} - D_{3t}) + u_t$
   c) $C_t = \alpha + \delta_1 D_{1t} + \delta_2 D_{2t} + \gamma_1 X_t D_{1t} + \gamma_2 X_t D_{2t} + \gamma_3 X_t D_{3t} + \gamma_4 X_t D_{4t} + u_t$
   d) $C_t = \alpha + \delta_1 D_{1t} + \delta_2 D_{2t} + \beta X_t + \gamma_1 X_t D_{1t} + \gamma_2 X_t D_{2t} + \gamma_3 X_t D_{3t} + u_t$

   (There's nothing special about labeling the parameters with α, β, γ_1, γ_2, δ_1, δ_2, etc. We could have labeled them β_1, β_2, ... without changing their interpretation.)

2. a) In models (a)–(d) of question 1, what are the slope coefficients of X_t in each of the 4 quarters?
   b) Suppose you estimate model (c) and wrote down the estimated slope coefficients for X_t in each of the 4 quarters. You then estimate model (d) and write down the estimated slope coefficients for X_t in each of the 4 quarters. Do your estimates change? Why or why not?

3. Determine whether the following are true or false and explain why:

   a) Adjusted R^2 can be negative.
   b) Adjusted R^2 can be larger than 1.

4. For this question you will need to load the data file `incomedata.Rda` (available from the NYU classes page) using the command `load("incomedata.Rda")`. Make sure you have the data file stored in your working directory. You will also need to load the package `AER` using the command `require("AER")`.

   The data file contains data on earnings (EARN), education (ED), and gender (GEN) for 1192 individuals in the early 1990s.

a) Estimate the regression model:

$$\text{EARN}_i = \beta_0 + \beta_1 \text{GEN}_i + \beta_2 \text{ED}_i + \beta_3 \text{GEN}_i \text{ED}_i + u_i$$

and interpret the estimated coefficients and the adjusted R^2.

b) Construct 95% confidence intervals for β_1 and β_3.

c) Do a t-test for $H_0 : \beta_3 = 0$ against $H_1 : \beta_3 > 0$ at the 5% level of significance.

d) Do a Wald test for $H_0 : \beta_1 = \beta_3 = 0$.

e) Now estimate the regression model:

$$\text{LOGINC}_i = \beta_0 + \beta_1 \text{GEN}_i + \beta_2 \text{ED}_i + \beta_3 \text{GEN}_i \text{ED}_i + u_i$$

where LOGINC is now the log of the earnings of person i. How does your answer to (c) change?