Econometrics - Exam 2
Fall 2019
Prof. Benjamin Zweig

(1) Consider a linear regression model where X_1 is the variable of interest. That is, you want to get an unbiased estimate of the marginal effect of X_1 on Y. You believe that the following covariances exist
$Cov(X_1, X_2) > 0 \qquad Cov(X_1, X_3) > 0 \qquad Cov(X_1, X_4) = 0$
$Cov(X_2, X_5) \neq 0 \qquad Cov(Y, X_2) \neq 0 \qquad Cov(Y, X_3) = 0$
$Cov(Y, X_4) < 0 \qquad Cov(Y, X_5) \neq 0$
Which variables (aside from X_1) must be included in the regression? Why?

(2) Assume that a true model is given by:
$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + \epsilon_i$
All relevant assumptions are satisfied for this model except that you do not have any information on W_i so you estimate:
$Y_i = \beta_0 + \beta_1 X_i + \mu_i$
You get an estimate, $\beta_1 = 1.2$ and believe that $Cov(X, W) > 0$ and $Cov(Y, W) < 0$.
How can you use this information to draw a useful conclusion about the effect of X on Y? Go into detail.

(3) You propose the following model for the cross-sectional determinants of LSAT scores:
$LSAT_i = \beta_0 + \beta_1 prep\_course_i + \beta_2 study\_hrs_i + \beta_3 GPA_i + \beta_4 private\_sch_i + \beta_5 male_i + \beta_6 income_i + \beta_7 white_i + \epsilon_i$

Where the variables are defined as above, and:
$prep\_course_i$: indicator dummy taking the value of 1 if individual i participated in a prep course course and zero if otherwise
$study\_hrs_i$: number of hours studying
$private\_sch_i$: indicator dummy taking the value of 1 if individual i attends a private university and zero if public school
$male_i$: indicator dummy taking value of one if individual i is male and zero if female
$income_i$: household income level of individual i in thousand USD
$white_i$: indicator dummy taking value of one if individual i identifies as "white/Caucasian" and zero if otherwise

(a) Provide an intuitive interpretation of each of the coefficients in your model, assuming you are able to estimate the equation with the data you have collected.
Note that you do not need numeric estimates of the β's to do this, so do not refer to Table 1 for this question.)

You decide to jump right in and find the following parameter estimates (Table 1) via OLS regression.

$prep\_course_i$	6.5	
$study\_hrs_i$	7.2	
GPA_i	4.8	
$private\_sch_i$	1.8	
$male_i$	6.8	(Table 1)
$income_i$	0.5	
$white_i$	5.6	
$constant$	89.0	
R-squared	.2236	

(b) Provide interpretation of the estimated parameters and R-squared of the model.

(c) Based on the second model, what would your estimates predict as a raw SAT score for an individual who:
- did not take a prep course
- studied four (4) hours per day on average
- had a GPA of 4.0
- attends public university
- is male
- whose parents income is $40,000 per year (be careful with units here)
- and is white?

(d) In addition to those used above, what are some additional data fields (variables) you would be interested in collecting/using in a multiple linear regression to explain LSAT scores? Name at least three.

(4) You are running a regression examining the relationship between time spent studying and salary of first job and collect data on a large sample of college students across all major universities in the US. You first run a regression with the log of salary of first job as the dependent variable and the log of time spent studying as the independent variable. You get a coefficient on the log of time spent studying of 0.4. Then, you run the same regression, but you also include dummy variables for each university. The second regression gives you a coefficient on the log of time spent studying of 0.1.

How do you interpret each coefficient? Why are they different? What kind of interesting conclusions can you draw?

(5) A study on the effect of walking and weight lifting on weight loss produced the following OLS estimate of a regression of Loss, number of pounds lost per month by participants in the study during a 6 month period on the independent variables Walk, a variable which is the distance the person walked per week; Lift, a dummy variable that takes the value 1 if the person lifted weights 3 time a week and 0 otherwise; Walk*Lift, an interaction term.
$Loss = 1.21. + 1.47Walk + 0.6Lift + 2.2Walk * Lift$

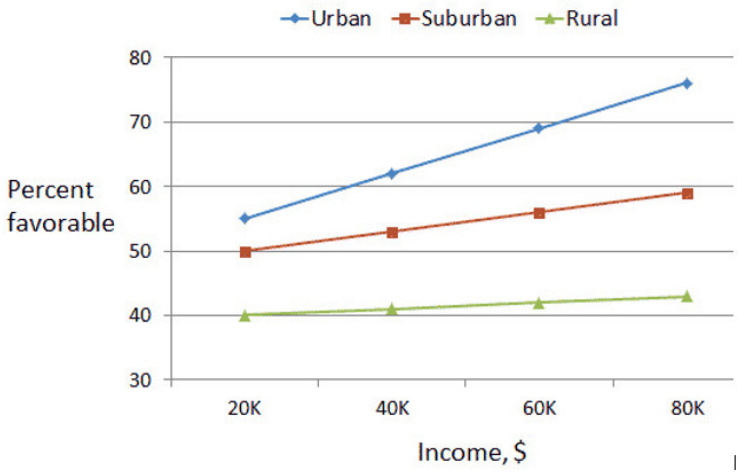
(a) Describe the base (reference) group in this model.

(b) Interpret in numbers and words the coefficient on the variable $Walk$.

(c) Interpret in numbers and words the coefficient on the variable $Lift$.

(d) What is the estimated average weight loss if a person both walked and lifted weights?

(6) Consider the interaction plot shown below. The plot shows the effect of income and community type (urban, suburban, rural) on attitudes toward gun control.



What conclusions would you draw from the plot?
I) The effect of income varies, depending on community type.
II) The regression equation should include an interaction term.
III) The regression equation should not include an interaction term.

(A) I only

(B) II only

(C) III only

(D) I and II

(E) I and III