Economics 312 Spring 2020
Project #3 Assignment Due: 11:59pm, Mon., February 17

Partner assignments

Shisham Adhikari	Matt Wan
Blaise Albis-Burdige	Roy Shannon
Isabelle Caldwell	Amrita Sawhney
Anne Cao	Lauren Rabe
Aditya Gadkari	Grisha Post
Catherine Gong	Evian Oosthuizen
Lirui Jiao	Max Nobel
Jonathan Li	Paul Nguyen
Charlie Lyu	Peter Mulgrew
George Zhao	

If you haven't yet done so, look at the project guide and sample report on the class Web site to get an idea of how your team should approach working together and what is expected in your report. The key point is that your report should be a narrative document (with an introduction and a conclusion) that incorporates the answers to the questions posed in the assignment, not a discrete set of answers to individual questions. We haven't talked yet about internal and external validity. In brief, internal validity is whether the assumptions of your estimator are reasonable and external validity is whether the results can be generalized beyond your specific sample/population.

This project uses a subset of a data set that accompanies Stock and Watson's popular econometrics textbook. A link to the data set is on the assignment page.

The data are taken from the following paper, to which the authors apply an amiable (Amy-able?) and appealingly alliterative appellation:

Daniel Hamermesh and Amy Parker, "Beauty in the Classroom: Instructors' Pulchritude and Putative Pedagogical Productivity," *Economics of Education Review*, 24(4), August 2005, 369–76.

(You are not expected to read this paper for this assignment; in fact, it's probably better if you don't because reading it may bias how you would approach or interpret the questions in the assignment. The reference is included to give appropriate credit and in case you are interested in the topic.)

The table below describes the variables in the data set:

Variable	Definition
eval	"Course overall" teaching evaluation score, on a scale of 1 (very unsatisfactory) to 5 (excellent)
beauty	Rating of instructor physical appearance by a panel of six students, averaged across the six panelists, shifted to have mean zero.
age	Instructor's age
female	= 1 if instructor is female, 0 otherwise
minority	= 1 if instructor is non-white, 0 otherwise
nnenglish	= 1 if instructor is not a native English speaker, 0 otherwise
intro	= 1 if the course is introductory (usually large lectures), 0 otherwise
onecredit	= 1 if the course is single-credit elective (yoga, aerobics, dance, etc.), 0 otherwise

1. Data and summary statistics

a. Examine the data using the data browser. Do there seem to be multiple observations from the same instructor? How might this violate the OLS "SLR" assumptions? Based on casual examination of the data, discuss how much of this dataset might be affected.

b. Compute and present the summary statistics for four variables. The *female* variable and the ones below it in the list are what we call "dummy" (or indicator or binary) variables. How can we interpret the mean of these variable? The table above says that the *beauty* variable is "shifted to have mean zero." What does this mean (no pun intended … for once)? It this verified by your sample mean? Is there anything else notable about the summary statistics?

2. Simple regression. Perform a simple regression of course evaluations on beauty and use the results to answer the following questions:

a. What is the interpretation of the intercept term in the regression? What is the interpretation of the slope coefficient? If Professor Average has average beauty, what is our best prediction of her evaluation score? If Professor Gorgeous has a beauty score two standard deviations above the mean, what is our prediction of his score? If Professor Ugly has a beauty score two standard deviations below the mean and gets evaluations of 3.9, would you consider him to be a better or worse teacher than would be expected based solely on his appearance? Why?

b. If we increase the value of *beauty* by one standard deviation (of *beauty*), by how many standard deviations (of *eval*) does the conditional expectation of *eval* go up along our regression line? (This measure is more common in non-economics social sciences and is often—misleadingly for our purposes—called a "beta" coefficient.) Why is this

measure a useful complement to the regression slope coefficient in understanding the relationship between *beauty* and *eval*? Does your estimated "beta" value seem "small" or "large"?

c. Assess the strength of your regression relationship using (1) the *R*-square statistic and (2) the standard error of the regression (*s* or $\hat{\sigma}$, reported in Stata as the "root mean squared error" or RMSE) in comparison to the standard deviation of the dependent variable. (And why is that comparison relevant?) Retrieve the fitted (predicted) values and residuals from this regression. What is the range of your fitted values compared to the range of actual values of *eval*? Why do you think this is true? Is this a problem? Examine the residual series to see if there are any obvious outliers. Create a graph with both the actual values of the dependent variable and the fitted values (measured on the vertical axis) plotted against the regressor (*beauty*) on the horizontal axis. Based on your graph, does the linear functional form seem reasonable or is there an obvious non-linear form (such as quadratic or log) that would fit the data better? Graph the residuals on the vertical axis of a scatter plot with the regressor on the horizontal axis. Is there evidence of heteroskedasticity in that the variance of the error may be related to the magnitude of the regressor?

d. The Union of Ugly Reed Professors (UURP) is campaigning for the abolition of traditional instructor evaluations because of beauty-bias. You have been asked by Reed to comment (positively or negatively) on this question based on your econometric results. Make your argument for or against UURP's position.

3. Sub-sample regressions. Some have argued that our society seems to place more emphasis on women's appearance than on men's. Perform separate regressions for the sub-samples of female and male professors. Interpret the results. Do they support the supposition that appearance is more important for women's evaluations? (We will soon learn how to perform a statistical test of this hypothesis; it can't be done from separate regressions.)

4. Effects of age. Another frequent presumption is that young people are generally considered to be more beautiful than older people. Is this true in this sample? Calculate and interpret the correlation coefficient between beauty and age. Suppose that older professors are generally more experienced and because of this tend to be more effective teachers, other things being equal. Explain how omitting age from the regression might bias the estimated effect of beauty on evaluations in this case: Would the bias be upward or downward? Add age to the regression to see if your expectation of bias is confirmed. Is age an important determinant of evaluations?

5. Other variables. For the other four variables in the data set, what (if any) effects would you expect on course evaluations? Run a regression of *eval* on all of the other variables and interpret the results. Are your hypotheses above confirmed, refuted, or are the results inconclusive? Did including these regressors have any effect on your conclusion about the

effect of beauty or age? How effective is your regression at explaining course evaluation scores?

6. Averaging multiple courses by the same instructor. Assume that any observations with the same values for *age*, *female*, and *beauty* are different course sections that were taught by the same instructor. Why might this be a violation of the standard regression assumptions (and which one)? Sort the dataset by these three variables, then use the collapse command to create a new dataset with one observation per instructor and with the evaluation variable (and the other regressors) set equal to the average evaluation for each instructor across all courses. (Use "save as" to be sure that you don't save this over your original data set!) How many observations are in this dataset compared with the original one? Re-run the *eval* regression (with all regressors) and discuss how or whether the results change, and why. Is this what you expected? What are the advantages and disadvantages of the one-observation-per-instructor regression model relative to the one-observation-per-course model? Which regression do you think is more reliable, and why?