David Gabel April 3, 2019 Econometrics

Writing Assignment

Economists are often involved in collecting, processing and analyzing data. Perhaps an economist's greatest contribution to his/her work place is his/her ability to formulate a hypothesis on how markets and/or individuals behave, test the hypothesis, and then discuss the implications of the investigation.

Recent research by Chetty *et al* addresses college attendance and income mobility in the United States: "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility". The paper reports a tremendous variation in the mobility rate across colleges. The authors write, at page four, that "The lack of observable predictors of mobility rates underscores the value of directly examining students' earnings outcomes by college as we do here, but leaves the question of understanding the production and selection technologies used by high-mobility-rate colleges open for future work." Your assignment is to develop and estimate a model that attempts to explain the variation in mobility rates across colleges. Does your model and estimates enhance our understanding of policies that should be pursued that will enhance intergenerational mobility?

There are other questions that can be analyzed using the data. Among other findings, the authors of the paper concluded that "These aggregate trends mask substantial heterogeneity across colleges within selectivity tiers. Most importantly, the fraction of students from low-income families at the institutions with the highest mobility rates--for instance, SUNY-Stony Brook and Glendale Community College—fell sharply over the 2000s. These changes in low-income access were not strongly associated with significant changes in students' earnings outcomes, implying that these colleges have significantly lower mobility rates for more recent cohorts. In short, the colleges that offered many low-income students pathways to success are becoming less accessible over time." Develop and estimate a model that attempts to explain "the fraction of students from low-income families at the institutions with the highest mobility rates." You should try to answer the question raised in the paper, "A key question for future research is why access has fallen specifically at colleges with both high success rates and high access." What are the policy implications of your research?

Another possible topic is the issue of how do post-college earnings vary as a function of tuition, expenditures per pupil, major, quality-of-student body as measured by performance on a standardized test, etc.

A fourth feasible topic is identifying and measuring the factors associated with a lack of mobility, or downward mobility (in contrast with identifying the factors that explain upward mobility). Can you identify the factors that explain why a student falls from the top to the bottom income quintile?

You are at liberty to use the data to explore other questions. I recommend that you check-in with me before heading down your own path.

You may need to look for external data. For example, it might help to know how the acceptance rates at colleges have changed over time. Or you may want to add data on educational infrastructure, such as county library expenditures per capita, or library books per capita.

The map at http://www.equality-of-opportunity.org/ will help you visualize how mobility varies by area of the country.

The data is available at http://www.equality-of-opportunity.org/data/ You should pick-up the zip file.

You may find a need to combine two or more data sets (e.g., you may want to combine information from mrc_table10.dta with data from mrc_table1.dta. This can be done in Stata by executing the commands:
use "mrc_table10.dta"
describe
sort super_opeid
joinby super_opeid using "mrc_table1.dta"
describe
* you will see that the number of observations has declined and the number of variables has increased.

Income mobility is studied on a regular basis by U.S. Census Bureau. See, for example, "Dynamics of Economic Well-Being: Movements in the U.S. Income Distribution, 2004-2007," *Current Population Reports,* P70-124, March 2011.

You might find some useful data at https://collegescorecard.ed.gov.

Deliverables

To complete the project you must carry out the following tasks:

1. Formulate a hypothesis regarding the relationship between three or more economic variables. When you formulate your hypothesis, think seriously about what you have learned in your economic theory classes.
2. Plot the data that you will use in your regression and use the graphs to suggest a functional form (e.g., linear, log-linear, or quadratic).
3. Summarize descriptive statistics. What do the values imply?
4. Estimate the coefficients.
5. Are you missing any key variables? Is this likely to lead to omitted variable bias?
6. Plot the residuals. Are the residuals consistent with the assumptions of classical regression? If not, can you modify the model so that the residuals are consistent with the assumptions?
7. Conduct hypothesis tests.

8. Write a four to nine page paper (this page count does not include the pages associated with the Stata output). The format of the paper can be either a short paper or a memorandum. The paper should include:
    a. A description of the issue that you are analyzing.
    b. State the hypothesis that is being tested.
    c. Report your estimation results.
    d. Report the results from your hypothesis test. Include at least one hypothesis test where you are testing if two or more variables, but not all slope coefficients, are equal to some value (e.g., zero).
    e. Discuss the implications of your results. What are the policy implications of your research?
9. Use a grammar utility program to check your sentence structure and spelling.
10. Attach your log file to the paper.

Submission Format

Your paper must be submitted electronically through SafeAssign. SafeAssign is a program available through Blackboard. I will use SafeAssign to check for plagiarism. Plagiarism is a violation of the Queens College academic standards and policy and will not be tolerated.

Due Date

The due date for the papers in May 6, 2019.