DSBA/ECON 6112
Graduate Econometrics

Replication Project: Part 1

The following paper is posted to Canvas, along with the dataset:

Angrist, Joshua, Eric Bettinger, and Michael Kremer. 2006. "Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia." *American Economic Review*, 96 (3): 847-862.

Instructions:

- Read the paper.

- Briefly summarize the paper's main question and data, being sure to address:
  - What is the main question that the authors are trying to answer?
  - Where do the data come from? (Who is included in the sample, etc.?)

- Replicate the results in Table 1
  - The table (values) in your final write-up must match your do-file and log file.
  - Format the table as in the published paper (rounding, parenthesis around standard errors, etc.)
  - If there are any discrepancies between your results and those reported in the paper, you should highlight these and put a note explaining that this is the case. (These are generally due to typos, so there will be no more than a few in a single table.)

- You may form small groups and consult with 1-2 other classmates, as long as you (1) include their name on your assignment, (2) submit your own *unique* answers and Stata files, and (3) do not consult any other classmates (but you may consult me).

You must submit the following files in Canvas:
- Do File
- Log File
- PDF or Word document that contains no more than 2 pages: one page for the formatted table and (no more than) one page for the written summary.

Note: The do-file and log file must show all work with the data, including opening the dataset, any data manipulations, and all subsequent code to produce the final output for your tables and responses.

Tips for Stata work:
- To examine mean differences in some outcome between two subgroups, we can regress the outcome on a subgroup dummy variable.

- To use only part of the sample, we use "if" statements with relational operators (=, <, >). For multiple restrictions, we use "&" between them.

Things to keep in mind:

- You are being asked to replicate part of a research paper *to practice doing research*, so treat the assignment as such.
- Your grade will be primarily based on the accuracy of all of the numbers in replicating the research paper, but formatting is also important and will count towards your grade.
- You must use Stata to calculate all of the numbers in the table, but you may use any software you'd like to actually put the numbers together in a table. (For instance, in the example I provide on Canvas, the code in the do-file writes output to Excel, where the table can be formatted and copied into Microsoft Word.)

Grading rubric:

	Points Possible	**Evaluation criteria**
Do File	5	Complete, executes without error, produces log file
Log File	5	Complete, produced by do-file, clean (easy to read/find values), matches table values
Summary	5	Describes main research question, data source, and sample
Table 1		
Formatting	5	Follow formatting instructions
Column 1	7	Reported numbers match paper and Stata files
Column 2	7	Reported numbers match paper and Stata files
Column 3	9	Reported numbers match paper and Stata files
Column 4	9	Reported numbers match paper and Stata files
Column 5	7	Reported numbers match paper and Stata files
Column 6	7	Reported numbers match paper and Stata files
TOTAL	**66**	