

GPEC 435 — Fall 2020

Topics in International Trade

# Ground Rules for the Empirical Exercises

October 5, 2020

Instructor: Marc-Andreas Muendler [muendler@ucsd.edu](mailto:muendler@ucsd.edu)  
Teaching Assistant: Jacob Orchard [jdorchard@ucsd.edu](mailto:jdorchard@ucsd.edu)

## Documentation, Replicability, Transparency

Essential principles for empirical work are documentation, replicability, and transparency. Your empirical exercises must adhere to them.

Each empirical exercise has three deliverable files:

1. A log file that shows all steps from initial input data to final output (.log);
2. a program file that underlies the log file (.do for Stata);
3. a results file (.pdf).

While **documentation** is arguably the foremost principle for conclusive scientific work and credible policy analysis, for the purposes of the empirical exercises a structured code file with meaningful remark lines and a log file that properly documents the crucial data transformations will suffice. You do not need to write up a verbal documentation of your data construction and methods for empirical analysis in the results file for the empirical exercises. (You will have to write up a proper verbal documentation for your gravity project paper.)

For full **replicability**, your code must start from the basic input files provided on the course web page for each empirical exercise. Your accompanying log file must exhaustively document the steps from the provided input files to the final output of results.

**Transparency** requires that you make your code and log files readable to outsiders. Meaningful remark lines specify which data you read from which source and what intention your code pursues with subsequent or immediately preceding commands. Examples of proper documentation of crucial data transformations are: a tabulation of matched and unmatched observations after a data merge, a count of observations that are appended or dropped when data matrices are stacked or rows removed, a report of observation counts when variable (field) contents are altered. Statistical software such as Stata or SAS produces informative log files by default. If you call other software, such as Python from within Stata or SQL from within SAS or operating-system commands, it is your responsibility to make sure that your log file provides the relevant information. At times, the elegance of compact coding can get in the way of transparency of data preparations and statistical methods. Err on the side of transparency.