Pre-Publication Code Review Checklist

Computational Reproducibility (Required)

☐ Provide a .zip file or GitHub link with the entire project directory. See this project for an example of an organized directory structure. The folders should include:

- ☐ All necessary de-identified data for the analysis
- ☐ All code necessary for the analysis
- ☐ The raw outputs you have used for the paper
- ☐ No extraneous documentation or PII data you would not share publicly
- ☐ No outputs or datasets that are not used in the analysis

In either the /dofiles/ folder or in the root directory, include a master script (dofile or Rscript for example). The master script should allow the reviewer to change one line of code setting the directory path. Then, running the master script should run the entire project and re-create all the raw outputs exactly as supplied. Using iefolder from our ietoolkit can help standardize this in Stata.

☐ Indicate the filename and line to change:

☐ Check that all your code will run completely on a new computer:

- ☐ Install any required user-written commands in the master script (for example, in Stata using ssc install or net install and in R include code for installing packages, including installing the appropriate version of the package if necessary),
- ☐ Make sure critical settings like version, matsize, and varabbrev are set correctly.
- ☐ The master file should indicate the settings of these needed to run, or use a wrapper command like ieboilstart from ietoolkit
- ☐ All outputs should clearly correspond by name to an exhibit in the paper, and vice versa. Supplying a compiling TeX document can support this.

☐ The submission package should include these outputs in the location they are produced.
☐ Code and outputs which are not used should be removed.

☐ Approximately how long does the code take to run (ie, minutes, hours, or days)?

Ease of Use (Recommended)

☐ Analysis scripts should not include any data cleaning or variable creation, unless technically necessary for the creation of a table or graphic (eg, the separate command in Stata).

☐ Data cleaning of raw variables should not be included in the analysis package at all.

☐ Variable creation for derived or constructed measures should be included in a separate script with detailed code comments about each new variable that is generated.

☐ Analysis scripts should be completely modular. The section for each exhibit should begin with a fresh environment (use or clear in Stata, or start a fresh R session). This can also be accomplished by having a separate scripts for each exhibit.

☐ Analysis code should be well-commented and indented, such that the reader can easily identify functional chunks of code and evaluate whether they correctly implement the econometric or statistical process described. The reader should not have to figure out the process by reading the code.

☐ Graphics should be output as .eps files when possible.

☐ Tables should be output as .csv or .tex files when possible.

☐ In-text numerical citations that are not drawn directly from tables and figures should be computed and recorded in a separate file, such as a dynamic document format like .stmd using markstat in Stata or .rmd using R.