The Gender Diversity Gaps in Mathematics

Replication Package

Code Author: Francine Montecinos.

Raw data were not uploaded to Editorial Manager because the replication package exceeds 500 MB. Download the full package (code, data, outputs) here: Link to Replication Package

Contents

- `code/` (Stata, R, Julia)
- `data.zip` (unzips to `data/src`, `data/tmp`, `data/dta`, `data/proc`)
- `figures/` (exported PDFs)
- `tables/` (LaTeX tables)
- `README.txt` and this `README.md`

Setup

1. Unzip `replication_package.zip`, then unzip `data.zip` inside it so you have `data/src`, `data/tmp`, `data/dta`, and `data/proc` alongside `code/`, `figures/`, and `tables/`.
2. Software and packages
   - Stata 17+ (used with Stata 19.5) and SSC: `ftools`, `gtools`, `reghdfe`, `estout`, `outreg2`, `psmatch2`, `oaxaca`, `rsource`, `julia`.
   - R 4.5.1 with `dplyr`, `ggplot2`, `tidyverse`, `lme4`, `merTools`, `twang`, `kableExtra`, `survey`, `parallel`, `haven`.
   - Julia 1.x with `DataFrames`, `StatsModels`, `GLM`, `StatsBase`, `Random`, `NearestNeighbors`, `Statistics`, `CategoricalArrays`.
3. In Stata, open the project root and ensure `global DIR` in `code/stata/0_makefile.do` points to your unzipped folder (pre-set for `C:/Users/aoimo/Dropbox/PROJECT_Gender_Diversity_Gaps/`). Then run:

   ```
   do code/stata/0_makefile.do
   ```

   This runs `1_data_construction.do`, pauses for the R step, then continues through all scripts.
4. When prompted, run the R script manually:

   ```
   Rscript code/R/01_ps_gbm.R
   ```

   Return to Stata to finish `1a_final_dataset.do`, `2_descriptives.do`, `3_results.do`, `4_mechanisms.do`, `5_robustness.do`, and `6_selection.do`.

Outputs

- Figures: `Figure_1.pdf`, `Figure_2.pdf`, `Figure_3a.pdf`, `Figure_3b.pdf`, plus appendix `Figure_A2a.pdf`, `Figure_A2b.pdf`, `Figure_B1a.pdf`, `Figure_B1b.pdf`.
- Tables: `Table_1.tex`, `Table_2.tex`, appendix `Table_A1`–`Table_A10`, and `Table_B1` in `tables/`.

Runtime and environment

- OS: Windows 11 Pro; Hardware: AMD Ryzen 7 7730U, 40 GB RAM, 2 TB SSD.
- Stata 19.5 (compatible with Stata 17+), R 4.5.1, Julia 1.x.
- End-to-end runtime: ~1 hour on the above hardware.