

Overview

The code in this replication package allows the user to replicate both of the empirical applications in the paper “On Binscatter.”

To load all necessary Stata packages to run the replication code, please first run the Stata file `CCFF_2024_AER--configStata.do`

To replicate the first empirical application, based on Akcigit, Grigsby, Nicholas, and Stantcheva (2022), please run the State file `CCFF_2024_AER--AGNS.do`.

To replicate the second empirical application, based on Moretti (2021), please run the State file `CCFF_2024_AER--M.do`.

Data Availability and Provenance Statements

Statement about Rights

- ✓ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Summary of Availability

- ✓ All data are publicly available.

Details on each Data Source

The data sources are the replication files for the original papers. For the first empirical application, based on Akcigit, Grigsby, Nicholas, and Stantcheva (2022), the data were obtained from:

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/SR410I>

For convenience, the necessary data are saved in a single `.dta` file.

Datafile(s): `CCFF_2024_AER--AGNS.dta`

For the second empirical application, based on Moretti (2021), the data were obtained from:

<https://www.openicpsr.org/openicpsr/project/140662/version/V1/view>

For convenience, the necessary data are saved in two `.dta` files.

Datafile(s): `CCFF_2024_AER--M1.dta`
`CCFF_2024_AER--M2.dta`

Computational requirements

Software Requirements

- Stata 15 or higher¹
 - `binsreg` package (see <https://github.com/nppackages/binsreg/>)
 - `reghdfe` package (version number greater than or equal to 5.9.0 [released: June 3, 2020])
 - `ftools` package (required for use of `reghdfe` package)

Memory and Runtime Requirements

- The code was last run on a Mac Studio with Apple M1 Max chip with 64GB memory, 2 TB disk storage, and macOS Sonoma 14.2.1.

Summary Approximate time needed to reproduce the analyses on a standard (2024) desktop machine:

- ☐ <10 minutes
- ☐ 10-60 minutes
- ☒ 1-2 hours
- ☐ 2-8 hours
- ☐ 8-24 hours
- ☐ 1-3 days
- ☐ 3-14 days
- ☐ > 14 days
- ☐ Not feasible to run on a desktop machine, as described below.

Description of programs/code

- `CCFF_2024_AER--configStata.do` loads all necessary Stata packages to run the replication code
- `CCFF_2024_AER--AGNS.do` uses the data set `CCFF_2024_AER--AGNS.dta` to produce all of the underlying charts for Figures 1–5 and Figure SA-1.
- `CCFF_2024_AER--M.do` uses the data sets `CCFF_2024_AER--M1.dta` and `CCFF_2024_AER--M2.dta` to produce all of the underlying charts for Figure 6.

¹For replication code in R and Python, please visit https://github.com/nppackages-replication/CCFF_2024_AER.

Instructions to Replicators

- Move all replication files to the desired directory.
- Run `CCFF_2024_AER--configStata.do`
- In `CCFF_2024_AER--AGNS.do` and `CCFF_2024_AER--M.do`, set this directory as the global variable “main”.
- In the same directory, create the folders “graphs” and “data/temp”.
- Run `CCFF_2024_AER--AGNS.do`.
- Run `CCFF_2024_AER--M.do`.

List of tables and programs

The provided code reproduces all figures in the paper.

- `CCFF_2024_AER--AGNS.do` produces all of the underlying charts for Figures 1–5 and Figure SA-1.
- `CCFF_2024_AER--M.do` produces all of the underlying charts for Figure 6.

References

Akcigit, U., J. Grigsby, T. Nicholas, and S. Stantcheva, “Taxation and Innovation in the Twentieth Century.” *Quarterly Journal of Economics*, Volume 137, Issue 1, February 2022, 329–385, <https://doi.org/10.1093/qje/qjab022>.

Moretti, E., “The Effect of High-Tech Clusters on the Productivity of Top Inventors.” *American Economic Review*, Volume 111, Issue 10, October 2021, 3328–3375, <https://www.aeaweb.org/articles?id=10.1257/aer.20191277>.
