Replication instructions for "Fixed Effects and Beyond
Bias reduction, Groups, Shrinkage, and Factors in Panel Data"

Authors: Stéphane Bonhomme and Angela Denis

May 31, 2025

Overview

The code in this package replicates the analysis using weather data (downloaded from Wolfram Schlenker´s website) and corn yields data (from Burke and Emerick (2019)) using R-studio and Stata.

Data Availability and Provenance Statements

Daily Weather Data for Contiguous United States (Schlenker 2020): corresponds to daily weather records across measuring stations in the US. The data is publicly available at Wolfram Schlenker´ website and distributed freely for research purposes. Accessed on August 2nd, 2021. The current version of the data is stored at Zenodo. These files are big (and the updated version does not include all files initially downloaded; file number 1 in Table 1); hence we included in this replication package the summary data sets at the county-year level that we use for analysis.

The data from Burke and Emerick (2019) is free to use and redistribute as specified in this Creative Common license.

Dataset list

See Table 1 for the list of datasets used.

Requirements

Software requirements: we used stata 18, R 4.1.1, Rstudio 2021.09.0.

Packages in R: haven, data.table, fixest, ggplot2, ggpattern, Hmisc, mvtnorm, quantreg, RColorBrewer, xtable.

Table 1: Dataset list

N	Data file (in folder Data_ori)	Provided	Source
1	Schlenker/metaData/cropArea.dta	No	
2	Schlenker/metaData/linkGridnumberFIPS.dta	No	Wolfram Schlenker´ website
3	Schlenker/rawDataByYear_v2020 (yearly files)	No	
4	Schlenker/Schlenker_data_1950_2019.dta	Yes	Constructed using 1, 2 and 3 and files
5	Schlenker/Weather_data_1950_2019_1C.dta	Yes	"Code/1a. CODE Weather data.do" and "Code/1b. Weather data main.do"
6	Burke and Emerick (2016)/us_panel.dta	No	Burke and Emerick (2019)
7	Data_us_map.RDtata	Yes	Baseline data to plot maps of US counties

Hardware requirements: most of the code was run on a 6-cores Intel-based computer with windows 11, and 32G RAM. Portions of the code where run on Acropolis Server at the University of Chicago that allows up to 100G per user.

Controlled randomness: see Table 2.

Table 2: Controlled randomness

File	Description	Lines
Code/4a. CODE Jackknife version 2.R	Splitting the sample for bias correction (within functions)	44, 83, 228
Code/4b. Jackknife version 2.R	Splitting the sample for bias correction	33
Code/5a. CODE grouped FE.R	Selecting initial values for GFE (within function)	157
Code/5b. Grouped FE.R	Selecting initial values for GFE	120
Code/8a. CODE grouped FE by period.R	Selecting initial values for GFE by period (within function)	282
Code/8b. Grouped FE by period.R	Selecting initial values for GFE by period	62
Code/9. Summary table and figures.R	Simulating data for introduction	552, 596
Code/12b. Inference in GFE and GFEpp.R	Selecting bootstrapped samples	187, 224
Code/12c. SE for time-invariant models.R	Selecting bootstrapped samples	35, 45, 201, 210

Runtime: the code is not demanding in terms of running time (see Roadmap file), except for the grouped fixed-effects and grouped-factor models, and the bootstrap for random-effects. The grouped fixed-effects and grouped-factor models were run sequentially and they took several days each, depending on the model and number of groups. In each case, we applied Lloyd's algorithm for 3,000 starting values. Hence, a user may reduce the running time by running the code in parallel for different starting values. Similarly, the bootstrap for random-effects was

run sequentially, so time could also be reduced by running it in parallel.

Instructions for replicators

As a preliminary step, run the file "0. Setup.do" to modify the overall path of this project on your computer and to create the folders.

The files are organized in 4 folders:

- Code: code to run. Files with names including "CODE" contain auxiliary functions and need not to be run directly.
- Data_ori: data used for estimation. This folder includes that summary weather data at the county-year level.
- Data_new: folder where generated data are saved.
- Results: folder where generated tables and figures are saved.

The code should be run in order (files are numbered), except for files named "CODE" that define functions being called from the other files. Alternatively, a file can be run if the files that generate its inputs were already run. See roadmap file. Note that files 1a and 1b are the ones used for generating the summary weather data, and are included as a reference, but need not to be run.

List of results

Results	Description	Source code
Figure 1	Examples of time-invariant heterogeneity	Code/9. Summary table and figures.R
Figure 2	Examples of time-varying heterogeneity	Code/9. Summary table and figures.R
Figure 3	Temperature and de-trended yields	Code/2. Descriptives.R
Table 1	Regression estimates with constant coefficients	Code/3. Preliminary regressions.R
Figure 4	Fixed-effects estimates with heterogeneous coefficients	Code/9. Summary table and figures.R
Figure 5	Jackknife results – marginal density	Code/9. Summary table and figures.R
Table 2	Distribution of marginal effects of temperature across counties and years	Code/12c. SE for time-invariant models.R
Figure 6	Grouped fixed-effects estimates	Code/9. Summary table and figures.R
Figure 7	Random-effects estimates	Code/9. Summary table and figures.R
Figure 8	Fixed-effects estimates by subperiod	Code/9. Summary table and figures.R
Figure 9	Grouped-factor estimates by subperiod	Code/9. Summary table and figures.R
Table A1	Descriptive statistics	Code/1c. Data.do
Figure A1	Descriptive figures	Code/2. Descriptives.R
Figure A2	Maps of averages per county	Code/2. Descriptives.R
Table B2	Yield regressions - Driscoll and Kraay standard errors	Code/3. Preliminary regressions.R
Figure B3	Fixed-effects results - Density of marginal effects un-weighted	Code/9. Summary table and figures.R
Figure B4	Fixed-effects results in a regression of log yields	Code/10. FE regression log yields.R
Table B3	Various measures of temperature	Code/11. Non linear temperature.R
Table B4	Distribution of effects of increasing daily temperature by 1C	Code/11. Non linear temperature.R
Figure B5	Density of effects of increasing daily temperature by 1C	Code/11. Non linear temperature.R
Figure B6	Maps of marginal effects of increasing daily temperature by 1C	Code/11. Non linear temperature.R
Figure B7	Maps of coefficients in the nonlinear specification	Code/11. Non linear temperature.R
Figure B8	Sensitivity to trimming of Jackknife results	Code/4b. Jackknife version 2.R
Table B5	GFE - Distribution of marginal effects	Code/9. Summary table and figures.R
Figure B9	GFE - Maps of marginal effects per county	Code/9. Summary table and figures.R
Table B6	Random-effects estimates – robustness	Code/6b. RE.R
Table B7	Distribution of marginal effects - fixed-effects-per-period estimates	Code/7. FE by period.R
Table B8	Distribution of marginal effects - grouped-factor estimates	Code/9. Summary table and figures.R
Table B9	Distribution of marginal effects with standard errors	Code/12b. Inference in GFE and GFEpp.R
Figure B10	Maps of uncertainty in grouped fixed-effects estimates	Code/12b. Inference in GFE and GFEpp.R
Figure B11	Maps of uncertainty in grouped fixed-factor estimates	Code/12b. Inference in GFE and GFEpp.R

References

Shlenker, W. (2020): "Daily Weather Data for Contiguous United States (1950-2019)". Version March 2020. Accessed on August 2nd, 2022. http://www.columbia.edu/ ws2162/links.html

Burke, M., and K. Emerick (2019). "Replication data for: Adaptation to Climate Change: Evidence from US Agriculture". American Economic Association, 2016. Inter-university Consortium for Political and Social Research. https://doi.org/10.3886/E114567V1-128141