Replication files for “Does Directed Innovation Mitigate Climate Damage? Evidence from US Agriculture”

by Jacob Moscona and Karthik Sastry
Last Edited: September 30, 2022

1. Summary

This folder contains the materials required to replicate all results in the paper “Does Directed Innovation Mitigate Climate Damage? Evidence from US Agriculture.”

Organization. The main directory has three folders. The “Data” folder includes all data necessary for the analysis, including both the raw data and cleaned crop-level and county-level data sets. The “Code” folder includes programs necessary to construct key variables from the raw data and to generate the results in the paper. The “Counterfactuals” folder includes all necessary code and data for the counterfactual analysis (Section 6). The “Results” folder contains generated results.

Software Requirements. The replication files were prepared and verified using Stata 16.1 and an installation of Python 3 (Version 3.10.7) with the following installed packages: Numpy 1.23.3, Pandas 1.4.4, Matplotlib 3.5.2, and Seaborn 0.11.2. Data files are provided .csv (comma-separated values) and .dta (Stata dataset) formats. The tabular results are saved as .xls/.xlsx (Microsoft Excel spreadsheet) and .txt (raw text), and the graphical results are saved as .pdf (Portable Document Format).

2. Replicating the Results

Empirical Results. The programs necessary for replicating all main empirical results in the paper are: *crop_level.do* and *county_level.do*. These are both stored in the “Code” folder. The file *crop_level.do* generates Tables 1-2 and Figures 2-5 of the main text, while *county_level.do* generates Table 3 and Figure 6. Before running either program, set the path to the working directory on Line 11 — this is the only line that needs to be changed before running either file. The results are saved in the “Results” folder.

Two data sets in .dta format are required to generate the main results. These data sets, both stored in the “Data” folder, are:

- *crop_level_data.dta* — This is a crop-by-decade level data set that is used to generate all crop-level results in the analysis (Tables 1-2 and Figures 2-5). All variables are labeled in the data file and all source materials are described in the main text.

- *county_level_data.dta* — This is a county-by-decade level data set that is used to generate all county-level results in the analysis (Table 3 and Figure 6). All variables are labeled in the data file and all source materials are described in the main text.

Counterfactual Results. The program *plot_counterfactuals.py* generates Figures 7 and 8 of the paper. Before running this program, set the path variable on line 25. The program *plot_pathway_comparison.py* generates Figure 9 of the paper. Before running this program, set the path variable on line 16.

To replicate the underlying calculations for these figures, we provide three files. First, *do_counterfactuals_insample.py* calculates the main counterfactual scenarios for in-sample (i.e., historical) climate damage and mitigation. It takes as input our raw county-level data (see the section below on "Generating Main Temperature Variables"), the coefficients from our main estimation (which are printed in the code), and the estimated fixed effects (stored in an excel spreadsheet). It outputs summary statistics as plain text. By editing the options within this file, one can generate the results in alternative model specifications. Second, *do_counterfactuals_future.py* calculates the main counterfactuals for out-of-sample climate damage and mitigation. It takes analogous inputs and has analogous outputs. Third, *do_counterfactuals_bootstrap.py* re-does these calculations using bootstrap draws of model coefficients and fixed effects. Along with the inputs specified above, it also uses coefficient draws saved as .xls (Excel Spreadsheet) files in the Data/bootstrap_results/ directory. It produces files of the form *bootstrap_out_{options}.csv*, which summarize the model calculations for each labeled bootstrap draw. We also include these generated files as well as an option in the code to load them directly without re-doing the calculation. Finally, the script prints as text the bootstrap standard errors for each statistic. These statistics are also recorded in "CI_summary.xlsx."

3. Generating Main Temperature Variables (Crop and County Extreme Exposure)

Crop-level Calculation. The script *make_crop_shocks.py* generates the crop-level shocks. It takes as input data files with county-level total exposure to degree days, averaged over growing seasons (*dday_database_raw.csv),* and variants that incorporate model projections for future temperatures); county-level planted area (*county_areas_raw.csv*); and information on crop-specific temperature tolerance, collected from the EcoCrop database (*crop_database_raw.csv*). Before running the file, set the path on line 21.

It produces two output files. The first, by default called *crop_shocks.csv*, lists crops (identified by Census of Agriculture name, and a unique numerical ID consistent in our code); their total area planted (in log units), as well as exposure variables of the following form: {type}_{threshold}_{decade}. {type} is either gdd (growing degree days) or days (total number of days). {threshold} is either maxOpt or minOpt, corresponding to the maximum and minimum of the EcoCrop range. {decade} is the studied decade (with the convention that, e.g., the 1950s are 1950-1959, inclusive of endpoints). Finally, gdd30_{decade} are growing degree days in excess of 30 degrees C. All variables are sums over the growing period (April to October, inclusive) and sums over the decade. Our "baseline" temperature-exposure variables are of the form gdd_maxOpt_{decade}.

The second output file of *make_crop_shocks.py* is an intermediate input into the county-level calculation, which gives crop-specific exposure at the county level. By default, it is called *crop_x_county_shocks.csv*.

The file supports the following options. First, instead of the default area shares from the 1959 Census of Agriculture, one can use the 1955 or 2012 Census of Agriculture. Second, one can use projected temperatures based on an ensemble forecast model of Rasmussen, Meinshausen, and Kopp (2016, *J. of Applied Meteorology and Climatology*) (see also "Original Data Sources" below). Our methodology for incorporating these forecasts is described in the paper. The files *dday_database{pathway}_{decade}.csv* summarize projected days and degree-days, in excess of specific thresholds, at the county level for the 2050s and 2090s and for pathways RCP 4.5 ("45"), 6.0 ("60"), and 8.5 ("85"). Third, one can do the calculation for the entire US, or the subset of the US east or west of the 100th meridian longitude.

County-level Calculation. The script *make_county_shocks.py* makes the county-level extreme exposure and innovation exposure variables. It takes as input the crop-level and county-by-crop level data files produced by *make_crop_shocks.py*, as well as the aforementioned raw data on planted areas. Before running the file, make sure that you have generated these intermediate files, and set the path on line 24.

It produces as output one file, called by default *county_shocks.csv*. Its columns are of the form {variable}_{type}_{decade}. The {variable} is either "gdd" (growing degree days) or "days" (days). The {type} is either "own," indicating the local average shock ("Local EE"), "loo," indicating the leave-one-out aggregate crop-mix shock ("Innovation Exposure"), or "lso," indicating a leave-own-state-out variant of the same. The {decade} is the decade over which the temperature variables are averaged. Our "baseline" local agricultural productivity shocks are of the form gdd_own_{decade}.

The file supports the following options, which are similar to the ones for the crop-level calculation: using the 1955 or 2012 Census of Agriculture weights and using the forecasted temperature variables.

4. Original Data Sources

Here, we briefly review the original sources for our climate, variety, and crop physiology data.

Fine-Grained Weather Dataset. We accessed these data in May 2020 from the website of Wolfram Schlenker at this page: http://www.columbia.edu/~ws2162/links.html. Our study uses the "March 2020 version." These data are based on the PRISM Climate Data, with some differences in methodology indicated on Wolfram Schlenker's website. To calculate county-level variables (i.e., the raw data in *dday_database_raw.csv*), we used the suggested method of averaging over cropland grid cells, using (a re-writing of) the provided example Stata code.

Ensemble Climate-Change Forecasts. We accessed these data in May 2020 from the Zenodo database entry managed by DJ Rasmussen and Robert Kopp, which is available

online at this permanent link: https://doi.org/10.5281/zenodo.582327.

Historical Census of Agriculture. For historical land values and planted areas, we used the data posted by Michael Haines, Price Fishback, and Paul Rhode on the ICPSR, available at this permanent link: https://www.icpsr.umich.edu/web/ICPSR/studies/35206. For total planted areas and agriculture land values from the 2017 Census of Agriculture, we downloaded data from the USDA National Agricultural Statistics Service (USDA-NASS).

The USDA *Variety Name List*. These data were obtained by a FOIA request by the authors, and are collected by the US Department of Agriculture. The original data are dated June 26, 2018. The “ncrop” variable in *crop_level_data.dta* reports the total number of varieties released *since data were first recorded*, as of the end of the first year in that decade (i.e., “ncrop” increases over time, and its difference gives the number of new varieties in the studied decade; and “ncrop” in 1960 for corn is the number of corn varieties released during or before 1960).

List of Issued Plant Variety Protection (PVP) Certificates. These data were taken from the USDA Agricultural Marketing Service’s website. We accessed these data in September 2018 from the following link: https://apps.ams.usda.gov/CMS/.

Patent Counts. We tabulated these data based on searches in the PatSnap database, using the parameters described in our paper.

UN EcoCrop. We pulled data on the optimum temperature range for crops from the UN EcoCrop database on May 2020. A newer landing page for the website can be found here: https://gaez.fao.org/pages/ecocrop.

5. Glossary of Included Files

Below, we list all the included files and give a brief description of each. The wild card * is used to denote the set of all files with arbitrary characters in place of the *.

./Code

county_level.do: does county-level regression analysis (see Section 2)

crop_level.do: does crop-level regression analysis (see Section 2)

do_counterfactuals_*.py: does counterfactual analysis (see Section 2)

make_*_shocks.py: makes our productivity shock variables (see Section 3)

plot_counterfactuals.py: plots our counterfactual results for percent mitigation

plot_pathway_comparison.py: plots our counterfactual results comparing land value on different emissions pathways

./Data

CI_summary.xlsx: standard errors for counterfactuals (see Section 2)

bootstrap_out_*.csv: stored results for bootstrap of counterfactual statistics (see Section 2)

boostrap_results/.csv*: stored coefficient draws to bootstrap counterfactual estimates (see Section 2)

county_areas_raw.csv: original data for crop areas (see Section 3)

county_extra_for_counterfactual.csv: data for prices and modern total areas, loaded by the script which does the counterfactual

county_level_data.dta: clean data for county-level empirical analysis (see Section 2)

county_shocks.csv*: county-level productivity shocks (see Section 3)

crop_database_raw.csv: crop-level dataset with names and temperature cut-offs

crop_shocks.csv*: crop-level productivity shocks (see Section 3)

crop_x_county_shocks.csv*: crop-by-county-level productivity shocks (see Section 3)

dday_database.csv*: degree-day and days-above county-level data (see Section 3)

*gdd_panel_1950_2010*_fe.xls*: stored fixed effects from our main regression estimates, used in the counterfactuals

./Results

Figure_2.pdf*: components of Figure 2, “Changes in Extreme Exposure and Variety Releases Across Decades: Examples”

Figure 3.pdf*: components of Figure 3, “Extreme Exposure and Variety Development: Partial Correlation Plot (OLS)”

Figure_4.pdf: Figure 4, “Extreme Exposure and Variety Development: Panel Estimates”

Figure_5.pdf: Figure 5, “Temperature Distress and the Share of Climate-Related Patents”

Figure_6.pdf: Figure 6, “Marginal Effect of County-Level Extreme Exposure as a Function of Innovation Exposure”

Figure_7.pdf: Figure 7, “Historical Damage Mitigation Via Innovation”

Figure_8.pdf: Figure 8, “Projected Damage Mitigation via Innovation Over the 21st Century”

Figure_9.pdf: Figure 9, “Comparing Climate Scenarios, With and Without Innovation”

*Table_1.**: Table 1, “Temperature Distress Induces Crop Variety Development”

Table_2.:* Table 2, “Temperature Distress and Climate-Related Patenting”

*Table_3.**: Table 3, “Innovation and Resilience to Climate Damage”

(Note that Figure 1 is a description of theoretical results, and therefore not included.)