READ ME for replication data for: "Comment on 'Temperature and Decisions: Evidence from 207,000 Court Cases'." (previously distributed as "No, Judges Are Not Influenced by Temperature (or Other Weather): Comment")

Repository: https://doi.org/10.7910/DVN/3LOR3R

Author: Holger Spamann, hspamann@law.harvard.edu.
Date: July 2020, updated September 2021 (v.9 -- version of record checked by AEA Data Editor), updated again December 2024 (see note below) (see version history on Dataverse for full change log)

Content: This data repository contains code and data to replicate "Comment on 'Temperature and Decisions: Evidence from 207,000 Court Cases'." American Economic Journal: Applied Economics, 14 (4): 519–28; DOI: 10.1257/app.20200118 (previously posted to SSRN etc. as "No, Judges Are Not Influenced by Temperature (or Other Weather): Comment." The paper is a comment on Heyes and Saberian (2019a).

The code is complete from data download[1] through analysis, i.e., the code allows replicating the results of my paper starting only with the raw data available from government agencies online and from the data supplied by Heyes and Saberian (2019b) and Saberian (2022). However, agencies frequently make small changes to their databases, so a fresh download would likely generate minor discrepancies in the output. Moreover, some of the downloads would take a very long time (several hours). For this reason, the data folder contains the download of the government data that was actually used in the paper. In the case of weather and pollution, these data are not fully raw but rather already pre-processed to reduce file size (using the scripts found in this repository).

Release note on December 2024 update: As requested by the AEA Data Editor, this update modifies the analysis script (now analysis_code_final.do) and download instructions to use Heyes & Saberian's publicly posted erratum data (Saberian 2022) rather than slightly different erratum data that they had previously provided to me and the Data Editor privately. This results in five more observations, but no substantive change, in Table 1 Model 6 relative to my published comment. Now that all data are finalized, for simplicity's sake, I have also removed some earlier versions of scripts and replaced the single master script master_revised.do with datatransformclean_master.do and analysis_code_final.do, which do what their names suggest. All removed files and older versions of files are still available in prior versions of the repository; see there for a full change log.

Known Issues/Errors:
1) Five additional observations in Table 1 model 6: see December 2024 release note above.
2) Mislabeling of weather measurement time in Table 1 model 6: Table 1 in my published comment erroneously states that its model 6 uses weather measured at 6am-4pm *GMT*. In actuality, it is measured at *local time*.
3) Random variation in asylum data: On each run, the data transformation and cleaning script asylum_data_cleaning_and_checking_AEJ-RR.do will create a slightly different asylum_revised.dta: up to 115 observations (out of 1.58 million) can be different. The reason is that the script retains only the first of several application with the same ID but fails to deal with the fact that for 115 IDs, there are two applications tied for being the first. As a result, within that combination of ID and first filing date, which of the two applications is dropped is random. For details, see note at the top of the script. The fix would be to drop these 230 observations altogether, but I did not implement this here because it would differ from the code run to generate the results in the paper. As one would expect

[1] The download and unzipping of the data from Heyes and Saberian (2019b), Saberian (2022), and EOIR is not part of the scripts; it needs to be done manually as explained below.

given the small number of observations affected, the results are the same with all randomly different datasets that I have generated, except for the count of observations.

4) Unexplained discrepancy between 2021 and 2024 results for Table 1 model 10: The results for Model 10 of Table 1 produced by the analysis script now differ from those reported in my published paper, as set forth in the small table below. The differences do not affect the substantive conclusions of my comment. The source of the differences is unclear:

(a) On my hard drive, I have stored Stata estimates from August 24, 2020 that match those reported in the published version. I include these estimates in this repository in the analysis_output folder as est25_20200824.ster. The estimation command that produced those estimates[2]—automatically stored with the estimates by eststo—is the correct estimation command except that it uses `year` instead of `fyear` (fiscal year) as a control. I do not have a complete logfile (likely reason: since the full code then took a long time to run, I ran it piecemeal when adding model 10, which was a late addition).

(b) The AEA Data Editor confirmed in their report in September 2021 that the output of my analysis script (then analysis_code_revised.do) matches that reported in Table 1.

(c) However, the same analysis script does not now produce those results.

(d) Controlling for `year` instead of `fyear`, setting Stata's version to 16.1, using older versions of the analysis script, or using older versions of the asylum data (asylum.dta instead of asylum_revised.dta) did not resolve the issue. Nor do the reported results from model 10 match the current results from any unreported models.

	Temperature coeff. (se) [city-month c.i.] {city c.i.}	R^2	*N*	City-Months	Cities	*F*(all)	*F*(temp, clouds, rain)
reported	.29 (.15) [−0.01,0.60] {−0.10,0.74}	.31	561,132	661	56	0.93	1.62
script	.33 (.15) [0.05,0.62] {-0.08,0.87}	.35	487,431	660	55	1.01	1.90

QUICK START – To reproduce the paper's tables in Stata,

1. **Download and unzip the entire repository in its original format (not Tab) with the directory structure intact.**
2. **Open Stata and set its working directory to the top-level directory to which you have extracted the repository.**
   - If not already installed, `ssc install` user-written packages `estout`, `boottest`, and, for the raw data, `geodist`.
3. **Run analysis/analysis_code_final.do[3]**. This uses my final data. If you want to start from the raw data instead, first download and unzip Heyes and Saberian (2019b) to

[2] `regress grant temp6t4 press6t4 dew6t4 prcp6t4 wind6t4 skycover pm25 ozone carbon_monoxide i.(judge defensive nationality dow year city_month mid_pm25 mid_ozone mid_carbon_monoxide) if nati!="CH":nationality, cluster(city_month).`

[3] Stata script tested on versions 16.1 MP4 and 18 MP Parallel; it requires user-written packages `estout` and `boottest`.

Speed: This script used to take a long time to run (on the order of 10 hours). But with the newer version of `boottest`, it only takes about one hour on my 4-core i7-1185G7 @ 3.00GHz. If you want to speed things up more and do not care about precise bootstrap confidence intervals, you can reduce the number of bootstrap replications by one or even two orders of magnitudes (from 99,999 to 9,999 or even 999) in `analysis_code_final.do`'s "reg_and_tests" program. Note, however, that I get deviations up to .02 from run to run even with 99,999 replications (or at least recent runs with `boottest` have produced such deviations, possibly because of changes to the package from when I published the paper).

data/original_article/data[4] and run datatransformclean_master.do[5] before running analysis_code_final.do.

Other content: In addition to the Stata scripts described in Quick Start and scripts and data called by them, this repository contains scripts and helper files to download and pre-process the NOAA weather and AQS pollution data in data\weather\download_scripts_and_helperfiles. These are partially written in R (required packages: knitr, tidyverse, threadr, data.table – the latter three can be installed from the main script provided, Master_NOAA_EPA.Rmd [lines 12-13]).

Data sources: The raw data provided in this repository are excerpts from the following raw datasets, all of which are freely accessible online. If desired, you can replace the data files in the repository by re-downloading and preprocessing them as follows:

1. NOAA 2019: **weather data** downloaded through February/March 2019. Pre-processed files of that download are contained in the data/weather/semirawdata folder. The necessary scripts to reproduce this are provided in data/weather/download_scripts_and_helperfiles, which also contains instructions in README_for_NOAA_EPA_download_and_preprocessing.rtf. However, while working with the NOAA website, I noticed that data provided there slightly change over time, so you would probably get slightly different data if you re-downloaded the data from there today.
2. U.S. Department of Justice 2019: **EOIR asylum cases** (July 2019 release) downloaded from https://fileshare.eoir.justice.gov/FOIA-TRAC-Report.zip in August 2019 and archived at https://doi.org/10.3886/E120222V1. No further action necessary – the relevant files are already excerpted into this repository. If you do want to get them from the original source, replace the csv files in data/asylum/EOIR with the corresponding files from the original. You could also download the most recent release of the data directly from EOIR (https://fileshare.eoir.justice.gov/FOIA-TRAC-Report.zip) but then the ingestion scripts may not run without tweaking or even not at all because, in the past, EOIR has frequently made small changes to the data format and sometimes garbled the data.
3. U.S. Environmental Protection Agency 2019: **AQS pollution data**. (Same instructions as for 2., NOAA 2019).
4. matched.dta from **Heyes & Saberian** (2019b) and Saberian (2022) (in each case, folder Data/final).

In addition to these four main data sources, this repository contains several helper files. Finally, the analysis files pull in **USSC sentencing data** from Spamann (2017) (see his documentation if you want to access the original USSC files and/or understand how he processed them).

Detailed Contents

The description below follows the directory structure. Note that the data assembly code is stored with the respective data sources.

- **README-Dec2024update.pdf** – this file
- **analysis**
  - **analysis_code_final.do** – the final analysis code; produces the tables and statistics in the accepted version of the paper (updated December 2024)

[4] The file path should be, e.g., `data/original_article/data/Data/raw/asylum.dta`. Unzipping Heyes & Saberian (2019b) to this location will overwrite `data/original_article/data/Data/final/master.dta`, which is their final data set and which is already included in this repository.

[5] Stata script that calls other Stata scripts; tested on Stata 16.1 MP4; requires user-written package `geodist`.

- **analysis_output/** - a folder for writing the analysis results. Before you run any code, this folder will be empty except for …
    - … **est25_20200824.ster**: saved estimation results of Table 1 Model 10 from August 2020 that I cannot reproduce now as described in "Known Issues/Errors" above.

- **data**:
    - **datatransformclean_master.do** – a master script calling scripts described below
    - **EOIR Asylum**:
        - **asylum_revised.dta**: all asylum cases 1990-2019 regardless of how they were completed – used by analysis_code_final.do.
        - **asylum_data_cleaning_and_checking_AEJ-RR.do**: assembles asylum_revised.dta from raw data in EOIR folder
        - **EOIR:** a folder containing individual files from U.S. Department of Justice (2019) in csv-format:
            - tbl_Court_Appln.csv
            - A_TblCase.csv
            - B_TblProceeding.csv
            - eoir_case_data_code_key_05212019.pdf: a table explaining the meaning of some but not all data fields (from https://www.justice.gov/eoir/page/file/eoir-case-data-code-key/download)
            - Lookup: a subfolder containing some index files from the same dump that may help understand the meaning of the data.
        - **problematic_judge_codes.dta**: a helper file containing "test" and similar problematic judge codes in the EOIR asylum data hand-compiled by H. Spamann for checking data quality
    - **original_article**:
        - **data**: for now, mostly empty. However, to run my data assembly scripts reconstructing the final data from Heyes & Saberian's (2019a/b) raw data, you can and must download and unzip Heyes & Saberian (2019b) into this folder (which will, among other things, overwrite the matched.dta file provided here with the identical file from their online repository)
            - **Data/final/matched.dta**: final data set from Heyes & Saberian (2019a/b) – used by my analysis_code_final.do to create models 1 and 2 of my Table 1
        - **data_Errata**: for now, mostly empty. However, you may download here Saberian (2022) (which will overwrite the matched.dta file provided here with the identical file from their online repository)
            - **Data/final/matched.dta**: final data set for Heyes & Saberian (2022) from Saberian (2022) – used by my analysis_code_final.do to create models 1 and 2 of my Table 1
        - **reconstructed_data**:
            - **hourlyweather__inc_clean_and_vargen__and_GMTadj_nogaps_HS.do**: creates hourlyweather_vargen_GMTadj_nogaps_HS.dta[6]
            - **hourlyweather_vargen_GMTadj_nogaps_HS.dta**: a file of 6am-4pm local time weather averages created from raw weather data in Heyes & Saberian (2019b)
            - **matched_HS_improved_corrections.dta**: the equivalent of Heyes & Saberian (2019a/b)'s matched.dta (= final data set) but correcting

[6] Requires prior download and unzipping of Heyes & Saberian (2019b) as described in data/original_article/data's description.

court locations and time zone adjustments; created from raw data in Heyes & Saberian (2019b) and used by my analysis_code_final.do
        - **matched_HS_improved_corrections_stablestations.dta**: same, but using not the closest measurement station on any given day but for each city-year the station (within 20 miles) with the most complete coverage; used by my analysis_code_final.do
        - **organize_HS_improved_corrections.do**: creates matched_HS_improved_corrections.dta[7]
        - **organize_HS_improved_corrections_stablestations.do**: creates matched_HS_improved_corrections_stablestations.dta[8]
- **USSC**
    - **courts_crosswalk.dta**: a helper file linking the USSC "distr" code to the "location_code" used in the other files
    - Note: the USSC data themselves are stored and documented in another dataverse, https://doi.org/10.7910/DVN/TZRNKD (Spamann 2017). The analysis files automatically download the USSC data from there.
- **weather**
    - **NOAA_AQS.dta**: weather and pollution data used by my analysis_code_final.do
    - **NOAA_AQS_prep.do**: assembles NOAA_AQS.dta from the two dta files in the semirawdata folder
    - **download_scripts_and_helperfiles**: scripts to download and pre-process the NOAA weather and AQS pollution data; they produce the files in the semirawdata folder
        - **README_for_NOAA_EPA_download_and_preprocessing.rtf**: explains everything
        - …
    - **semirawdata** (assembled by the scripts and helper files in the download_scripts_and_helperfiles folder)
        - **AQS_daily_optimized_coverage.dta**: compilation of daily measurements from the NOAA's AQS data base
        - **NOAA_6a4p.dta**: 6am-4pm averages of hourly weather measurements from the NOAA's ISD Lite database.

References:

Heyes, Anthony, and Soodeh Saberian. 2019a. "Temperature and Decisions: Evidence from 207,000 Court Cases". American Economic Journal: Applied Economics 11(2): 238–265.

_____. 2019b. "Replication data for: Temperature and Decisions: Evidence from 207,000 Court Cases." Nashville, TN: American Economic Association [publisher], 2019. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-12. https://doi.org/10.3886/E113722V1.

_____. 2022. "Correction to 'Temperature and Decisions: Evidence from 207,000 Court Cases' and Reply to Spamann." American Economic Journal: Applied Economics, 14 (4): 529–33.

National Oceanic and Atmospheric Administration (NOAA), National Centers for Environmental Information. 2019. "Integrated Surface Data – Lite." ftp://ftp.ncdc.noaa.gov/pub/data/noaa/isd-lite/ (accessed February/March 2019).

[7] See note 6.

[8] See note 6.

Saberian, Soodeh. 2022. Data and Code for: Correction to “Temperature and Decisions: Evidence from 207,000 Court Cases” (by Heyes and Saberian AEJ: Applied Economics 11(2), 238-65, April 2019) and Reply to Comment by Spamann. Nashville, TN: American Economic Association [publisher], 2022. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2022-09-21. https://doi.org/10.3886/E127263V1.

Spamann, Holger, 2017, "Replication Data for: Are Sleepy Punishers Really Harsh Punishers?: Comment", https://doi.org/10.7910/DVN/TZRNKD.

U.S. Department of Justice, Executive Office for Immigration Review. “EOIR Case Data (July 2019)”. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2020-07-07. https://doi.org/10.3886/E120222V1.

U.S. Environmental Protection Agency. 2019. “Air Quality System (AQS).” https://www.epa.gov/aqs (accessed February/March 2019).