Replication Code for Leveraging the Disagreement on Climate Change: Theory and Evidence

Laura Bakkensen Toàn Phan Tsz-Nga Wong

January 30, 2025

The replication material contains two main STATA .do files: main.do and createData.do. Please note that the do files require propriety data from CoreLogic, First Street Foundation, Gallup, and confidential data from the Government Sponsored Enterprises (GSEs).[1] Due to the propriety and confidential nature, **the data files cannot be shared**.[2]

main.do This file replicates the empirical tables and figures in "Leveraging the Disagreement on Climate Change: Theory and Evidence."

Output (as csv/xlsx and png files in folder output):

- Part 1: Main regression Tables 2-5
- Part 2: Appendix Tables A1-A14
- Part 3: GSE Table 7 and Figure 4.

Input: corelogic_slr.dta, gse_fs.dta, and gse_fs_time_series.dta (not included).

createData.do This wrapper file (located in folder data_creation_code) calls several subprograms to create the data sets that are used as input for main.do.

Output: data\final\corelogic_slr.dta, gse_fs.dta, and gse_fs_time_series.dta (and various intermediate data files in folder data\intermediate).

Input: data files from data\raw (not included).

Subprograms:

[1]The CoreLogic data and GSE data were retrieved from the Federal Reserve's data warehouse (RADAR) in Aug 2020 and Feb 2024, respectively. The First Street data was retrieved from the First Street API that was available in 2021 — we thank Brett Lingle for kindly sharing the API. The Gallup data was purchased directly from Gallup in Jan 2020. Researchers can purchase the propriety data directly from CoreLogic, First Street Foundation, and Gallup. Researchers with approved access can retrieve the confidential GSE data from the Federal Reserve's RADAR data warehouse.

[2]Hence, the data folder in the replication material will be empty.

1. coreLogicData.do

   This file imports and formats data from CoreLogic, which had separate tax roll and deed information of single-family residential properties in the following East Coast states: CT, DE, FL, GA, MA, MD, ME, NC, NH, NJ, NY, PA, RI, SC, and VA.

   The data was directly queried from the Federal Reserve Data Warehouse (RADAR) in August 2020.[3]

2. slrShapeFiles.do

   This file imports processed NOAA sea level rise (SLR) shape files for each state and separately saves the data and coordinates for each foot of SLR (0ft, 1ft, ..., 6ft) as .dta files.

   The SLR data were converted from raw .gdb files/folders to shape files using the Python code called noaaDataToShp.py. It's necessary to have ArcGIS Pro installed to run this script.

   The raw NOAA sea level rise vectors come from `https://coast.noaa.gov/slrdata/Sea_Level_Rise_Vectors/index.html`, originally retrieved in August 2020 for each state listed above. We downloaded the 'state'_slr_data_dist.zip files for each state and unzipped them.

3. distFromCoast.do

   Each property's distance from the nearest coastline is calculated using its precise coordinates (given in the CoreLogic data) and NOAA's Continuously Updated Shoreline Product (`https://coast.noaa.gov/digitalcoast/data/cusp.html`). The following steps were executed using ArcGIS Pro:

   (a) Load .csv layer of property lat/long

   (b) Plot lat/long using x/y coordinates

   (c) Export data to create shape file with ObjectID

   (d) Run "Generate Near Table (Analysis)"

      i. Input Feature: Property points

      ii. Near Feature: Gulf_East_Coast CUSP shoreline (Continuously Updated Shoreline Product)

      iii. Set Output table name and location

[3]This code only works on earlier vintage of data from CoreLogic. As of 2024, CoreLogic has combined the tax roll and deed information into Property Basic.

iv. Use "Geodesic" method

v. Use "meters" as search radius unit (but do not set a search radius)

(e) Join property points shape file with newly created distance near table

i. Join based on FID (for properties) and IN_FID (for distance measures)

(f) Manually examine random set of points to be sure they processed correctly

(g) Save joined file as .txt

The .do file then takes this data by state, appends all states together, and saves distancefull.dta with the intermediate distance data.

Additionally, we downloaded the shapefiles of NOAA's US medium shoreline (`https://shoreline.noaa.gov/med-res.html`) and select properties (based on their coordinates) within 1km of the shoreline.

4. countyControls.do

This file cleans and saves .dta files for the following county-level controls:

(a) Yale_Data_2014.dta; Yale_Buyer_2014.dta

These files contain county-level data from the Yale Climate Opinion Survey of 2014. The code saves versions for both the property county and the buyer's home county. The data are available for download from: `https://climatecommunication.yale.edu/visualizations-data/ycom-us/`.

(b) electionEC.dta

This file contains the shares of people that voted Democrat vs Republican, sourced from the MIT Election Lab. We sourced the county presidential election returns from 2000-2020: `https://electionlab.mit.edu/data`.

(c) countyincome.dta; buyer_countyincome.dta;

countypopulation.dta; buyer_countypopulation.dta

These files contain average personal income and county population for both the property county and buyer origin county. They use data from the Bureau of Economic Analysis' (BEA) Regional Economic Accounts; specifically, the table "CAINC1: Personal Income, Population, Per Capita Personal Income". This has annual data per county from 1969-2016. Link: `https://apps.bea.gov/histdatacore/HistFileDetails.html?HistCateID=5&FileGroupID=294`.

(d) buyer_countyeducation.dta

This file contains county-by-five-year average estimates for educational attainment from Table B15001 of the U.S. Census' American Community Survey. The following link can be used to download each year in the format we did; this is necessary for the cleaning code to work properly: `https://data.census.gov/table/ACSDT5Y2010.B15001?q=b15001&g=010XX00US$0500000&moe=false&tp=true`

(e) SEER_population.dta; county_2005_2016.dta

The SEER_population file contains annual county-level population files from the National Cancer Institute's Surveillance, Epidemiology, and End Results Program to calculate the share of each county that is female, nonwhite, age 65 and older, and age 5 and younger. The data can be found using the following link: `https://seer.cancer.gov/popdata/download.html#single`

The data are then combined with the election and education variables to make county_2005_2016.dta, a dataset with all the shares together.

(f) unemp_2001_2016.dta

This file contains annual county unemployment rates from the Bureau of Labor Statistics. The data can be accessed using the following link: `https://www.bls.gov/lau/tables.htm`. Scroll down to Annual Average Data → County tables.

(g) tests_2009_2016.dta

This file contains annual test scores data from the Stanford Education Data Archive (SEDA). SEDA provides average academic achievement for grades 3-8 at the county level, as measured by standardized tests in reading and math. The data can be accessed here: `https://edopportunity.org/get-the-data/`

(h) crime_2001_2016.dta

This file contains the yearly total number of arrests at the county level from the Uniform Crime Reporting (UCR) Program Data. The file, called "Arrests by Age, Sex, and Race, 1974-2021", was accessed through OpenICPSR using Jacob Kaplan's concatenated files. Link: `https://www.openicpsr.org/openicpsr/project/102263/version/V15/view?path=/openicpsr/102263/fcr:versions/V15/ucr_arrests_monthly_all_crimes_race_sex_1974_2020_dta.zip&type=file`

(i) build_2001_2016.dta

This file contains the yearly number of new housing units authorized by building permits in each county, calculated from the Building Permits Survey from the Census Bureau. We use the co<YYYY>a.txt files from `https://www2.census.gov/econ/bps/County/`.

(j) flood_2001_2016.dta

This file is created from NOAA's Storm Events Database to calculate the number of flood events each year, lagged by one year to control for the previous year's flood events. We use the "StormEvents_details-ftp" files from `https://www.ncei.noaa.gov/pub/data/swdi/stormevents/csvfiles/`.

The .do file also creates a crosswalk from zip code to county (in 2014, the year we have for buyer beliefs), in addition to a crosswalk between county FIPS codes and their full names (crosswalk from MDR Education `https://mdreducation.com/pdfs/US_FIPS_Codes.xls`).

5. conformingLoans.do

   This .do file compiles county-by-year loan limit information from the Federal Housing Finance Agency (FHFA). We collect data from 2009-2016 here: `https://www.fhfa.gov/data/conforming-loan-limit`.

6. gallupAlternateBeliefs1.R; gallupAlternateBeliefs2.do

   These files clean and prepare data from Gallup's annual environment poll (part of the Gallup Poll Social Series; GPSS) to use as an alternative measure of climate beliefs. The proprietary Gallup data can be purchased directly from Gallup.

   The R file, which should be run first, combines the GPSS data with annual ZIP-level pollution level data (PM 2.5) from NASA, annual housing price data from the Census at the county level, and employment and wage data from the Quarterly Census of Employment and Wages (BLS) at the county level.

   The Stata file, which should be run second, cleans the merged dataset and adds it to the population and education data created above. It then follows the method of Howe et al. (2015) to impute time-varying county-level estimates for being "worried" about climate change, as not all years are present in the original Yale data.

7. firstStreetID.do

   To merge CoreLogic data with First Street data in order to retrieve each property's elevation,[4] this file matches First Street properties (identified with FSID) with CoreLogic properties (identified by prop_id_dw). This merging is done based on both property

[4] As of 2024, CoreLogic has added each property's elevation into their database (variable gr_el_used in the CoreLogic Climate Risk's database, which was not available at the time our data set was constructed). Hence, as of 2024, for those with access to CoreLogic Climate Risk data, it is no longer necessary to use First Street's elevation data.

coordinates (which are done in several iterations, as coordinates may differ slightly in the two data sets) and on street address.

8. createData1km.do

   This .do file combines many of the above intermediate datasets, raw First Street property-level data on bare earth elevation (along with some additional controls from First Street: historic flood events, climate adaptations, and sea depth probability). It also reshapes the data to the transaction level, creates bins for elevation and distance to the coast, and consolidates some mortgage variables. It saves intermediate.dta in the final data folder.

9. femaZone.do

   This .do file uses intermediate.dta to creates CSVs of property coordinates, which must then be intersected with FEMA's flood hazard layer in ArcGIS to get the flood zones. The ArcGIS output is then imported and appended into one file that can be merged to the main processed data. It creates a dummy variable that identifies if a property is in a FEMA flood zone. The output is prop_FEMA_zone.dta.

   To access the flood hazard layer, go to `https://hazards.fema.gov/arcgis/rest/services/public/NFHL/MapServer/28`.

10. createCoreLogicSLR.do

    This .do file combines many of the above input files into the dataset used for the final specification, corelogic_slr.dta. It cleans the data in the following ways:

    - Restricts to properties <= 1km from the coast
    - Drops properties with sale values below $50,000 and above $10,000,000
    - Flags mortgages without a standard 15 or 30-year term
    - Specifies high-cost counties
    - Adds conforming loan limits for years 2000-2008 (from the replication material of LaCour-Little et al. (2022), `https://www.dropbox.com/scl/fi/vnymkzrtxyutih2sx539p/loanlimitshistory07.pdf?rlkey=4k5jwxg1wj124r2vqxwrbkokm&e=1&dl=0`, retrieved in Dec 2022).
    - Other small variable changes

    Output to final data folder: corelogic_slr.dta

11. createGSEfs.do

    This file imports raw property-level First Street data with each property's Flood Factor (proprietary measure of flood risk) and aggregates it to the 3-digit ZIP code level, saving the average flood factor for each of these. It then merges this with GSE data on conforming loans using the 3-digit ZIP code, later cleaning and reshaping this to be a time series at the loan-quarter level, storing whether the loan is defaulted or not.

    Output to final data folder: gse_fs.dta; gse_fs_time_series.dta

References

Howe, P. D., M. Mildenberger, J. R. Marlon, and A. Leiserowitz (2015). Geographic variation in opinions on climate change at state and local scales in the USA. *Nature climate change 5*(6), 596–603.