Data and Analysis Replication "Readme" for: Climate risk and beliefs in New York floodplains"

Matthew Gibson and Jamie T. Mullins

May 23, 2020

This "readme" file provides a brief description of the steps necessary to duplicate the analysis and output of the paper: "Climate risk and beliefs in New York Floodplains". All code and final datasets to complete the final analysis and duplicate the tables and figures in the paper and appendix are provided along with this document. A description of the methods and materials used in construction of both the central and secondary datasets is also provided herein.

Instructions for duplication:

1. Download Stata code for replication in file: "replication.do". This file will reproduce all tables and figures from the paper and appendix with the exception of the maps, which were created manually using ArcGIS.
2. Download datasets used for final analyses and output. Place datasets in a directory titled "Data". The relevant datasets are:
   (a) data_for_main_analyses.dta
   (b) ACS_11_16_Diffs.dta
   (c) Google_trends_master.dta
   (d) NYC_NFIP_data.xlsx
   (e) TC1_Assessments_2012_MKT_VALUES.dta
3. Add another directory titled "output_materials" in the folder with the "Data" directory.
4. Run the replication.do file in Stata (Stata Version 16.1 was used for for the original analysis), entering the directory in which the "Data" directory was added as the "working directory" when prompted.

Details of data set construction and implementation of analysis: The majority of the analysis in this paper is based on geocoded data and geographic overlaps and/or proximities. These calculations and relationships were all computed in ArcGIS. The resulting output was then read into Stata and the analysis proceeded from there.

1. For the main data set, the raw real estate transactions files are read into Stata, processed, and formatted. A data set for analysis and a list of addresses to be geocoded in ArcGIS are saved and exported respectively.

2. The list of addresses is read into ArcGIS, geocoded using the New York State GIS Program Office Geocoding Services (specifically the Street_and_Address_Composite locator) accessed through ArcGIS. Each geocoded property has the level of Sandy inundation or damage experienced mapped onto it and is then identified as within or outside of the 1% floodplain in each of the following flood maps:
   (a) Currently Effective Flood Insurance Risk Map (FIRM)
   (b) Advisory Base Flood Elevation (ABFE) Map
   (c) Preliminary Working Map
   (d) Preliminary FIRM
   (e) Estimated Future Floodplain extents for 2020 and 2050
3. In ArcGIS, each geocoded property is also assigned an indicator as to whether it is protected by each of a number of proposed infrastructure projects as estimated by authors based on publicly available project plans and projected areas of projection.
4. In ArcGIS, the distance between each property and the edge of each floodplain (and Sandy inundation zone) is calculated.
5. Geocoded information (latitude and longitude) and all the variables created in steps 2-4 above are then exported from ArcGIS in .csv format.
6. Data exported from ArcGIS based on properties' geographic location is read back into Stata, merged with the transaction data, and the data is prepared for use in the analysis.
7. Using the provided code: "replication.do", the prepared data set is analyzed and each of the analysis-based tables and figures in the paper and appendix is created and exported to the directory: "output_materials".
8. In addition to the main dataset, there are sub-analyses within "replication.do", that rely on other datasets to generate several tables and figures. The final data for each of these analyses is provided in the following files that need to be placed in a directory titled "Data" within the working directory:
   (a) ACS_11_16_Diffs.dta - Census data from the 2011 and 2016 American Community Survey 5-year samples along with 2010 Census Tracks which have had floodplain membership data mapped on to them in ArcGIS. Differences in a number of variables are taken across the sample years and compared between Census Tracts that overlap with the current floodplain and those that do not. The result is put into a table which is Table 5 in the paper.
   (b) Google_trends_master.dta - data from the Google Trends tool for the search term "floodplain" for New York City and the US as a whole. This data is detrended and deseasonalized, and plotted to become Figure A16 in the Appendix.
   (c) NYC_NFIP_data.xlsx - data on NFIP policy numbers for New York City over time which were scraped from historic versions (maintained by the Wayback Machine) of a FEMA website on current policies in force: `https://www.fema.gov/policy-claim-statistics-flood-insurance`.
9. For the analysis that ultimately goes into Figure 5, the code: "replication.do", merges on assessment data from 2012. This data is downloaded from the New York City Valuation/Assessment Archive, read in to Stata, processed, and saved.