

Readme - Replication package for “Tax Base Fragmentation as a Dimension of Metropolitan Inequality”

Overview

This replication package includes code and intermediate results for the paper “Tax Base Fragmentation as a Dimension of Metropolitan Inequality” (Manduca, Highsmith, and Waggoner 2025). This paper is based on proprietary data from the real estate company CoreLogic (CoreLogic 2024). The replication package includes build code to construct the analysis dataset from independently accessed primary data. It also includes the analysis dataset itself, and code to conduct the analyses detailed in the paper.

Clean data by metropolitan area and jurisdiction are provided in the folder “clean_data”. We recommend using these data for further analyses.

Prior to running this replication package, the following setup procedures should be conducted:

- Create a directory for the replication package and copy all folders in the replication package to that directory.
- Change the path on line 3 of “source_calls.R” to the newly created directory
- Download the following publicly available datasets:
 - The Government Finance Database (Pierson, Hand, and Thompson 2015), available at <https://my.willamette.edu/site/mba/public-datasets>. All component datasets should be downloaded and saved to a folder “nonproprietary_data/TheGovernmentFinanceDatabase”
 - Census 2020 population data from NHGIS (Schroeder et al. 2025) for the full country, at the following spatial scales:
 - Census blocks
 - Census places
 - Census county subdivisions
 - School districts
 - CountiesAll population data should be saved in “nonproprietary_data/nhgis”
 - Census block shapefiles from NHGIS for all states, saved “nonproprietary_data/nhgis”

Build code

The following programs are used to construct the analysis dataset from the raw and proprietary CoreLogic data. They are included for transparency and to allow researchers with independent access to CoreLogic data to replicate the complete analysis. Note that prior to running these programs the CoreLogic data should be spatially joined to 2020 Census blocks. Programs marked with “*” require proprietary data not included in this replication package to run. These programs are all found in the “build_code” folder.

- `prep_block_fu_crosswalk.R` – Creates a crosswalk of Census blocks to fiscal units (legally independent local governments, as defined in the main text and Appendix 1).
 - Output: Crosswalk from 2020 Census blocks to legal jurisdictions (both general purpose local governments and school districts) is output as “intermediate_data/fiscal_units/fu_block_crosswalk.Rdata”.
- `prep_fu_shapefiles.R` – Creates shapefiles of fiscal units.
 - Output: “intermediate_data/fiscal_units/shapefiles” and “intermediate_data/fiscal_units/shapefiles_national”.
- `prep_eqrs.R` – Cleans equalization ratio data for New York and Pennsylvania.
- `*prep_cl_to_block.R` – Aggregates property parcel records to Census blocks and computes the number of properties, population, and total and per capita appraised property values by block.
- `*prep_percentiles.R` – Computes percentiles of property value for each CBSA
 - Output: “intermediate_data/cbsa_properties”
- `*prep_cl_to_fus.R` – Aggregates block-level data on appraised value to fiscal units, computing the total and per capita appraised property value for each fiscal unit.
 - Output: “intermediate_data/fu_prop”. These files form the basis for most of the analysis.

Analysis code

The following programs are used to conduct the analyses and create the visualizations presented in the paper. These programs are found in the “analysis_code” folder. They use the non-proprietary output found in the “intermediate_data” folder and should run completely. It is recommended to run them completely, in order, by running the “source_calls.R” program found in the main folder of the replication package.

- `analysis_tfq.R` – Calculates the tax base fragmentation quotient by metropolitan area

- Output: all files in “output/tfq”. Clean metro-level data with the tax base fragmentation quotient are output to
“output/clean_data/metro_data_[metro_type]_[jurisdiction type].csv”
- analysis_hilow.R – Identifies municipal tax havens and fiscally impoverished jurisdictions.
 - Output: all files in “output/hilow”. Clean jurisdiction-level data including fiscal capacity ratio and tax base per capita is output to
“output/clean_data/jurisdiction_data_[metro type]_[jurisdiction type].csv”
 - This program creates the data used in Tables 2 and 3 in the main text and Appendix Tables A7.1 and A7.2.
- analysis_econ_segregation.R – Computes rank-order information theory economic segregation metric of property value across census tracts by CBSA.
 - Output: all files in “output/econseg”.
- analysis_assessors.R – Calculates a rough total of the number of assessing jurisdictions based on the classification by state in Almy (2000).
- analysis_fisc_cbsa.R – Calculates the percentage of all state and local revenue that is from property taxes in each CBSA.
 - Output: “output/fisc/fisc_cbsa.csv”
- analysis_fisc_regression.R – Runs regressions of revenue by source on tax base
 - Output: “output/fisc/fisc_reg_[*].csv”
 - This program creates the data used in Table 4 of the main text and Appendix Table A8.1.
- maps_nation_tfq.R – Creates national maps of the tax base fragmentation quotient
 - Output: “output/nation_maps”
 - This program creates Figure 5 of the main text and Appendix Figures A3.1.1, A3.2.1, A3.3.1, A3.4.1, and A3.5.1.
- maps_nation_dq.R – Creates national maps of data quality
 - Output: “output/nation_maps/dq”
 - This program creates Appendix Figure A1.1.
- maps_cbsas_fcr_all.R – Maps the fiscal capacity ratio by metro
 - Output: “output/metro_maps”
 - These figures are not used in the article but may be useful for reference.
- *maps_cbsa_fcr_properties_figures.R – Maps the fiscal capacity ratio and the top and bottom 1% of properties for Detroit, Honolulu, and Des Moines. This program relies on proprietary block-level property data not included in this replication package.
 - This program creates Figures 2, 3, and 4 in the main text.

- `graphs_sc_proptax_reliance.R` – Makes scatterplots of TFQ versus reliance on property taxes by CBSA
 - Output: “output/sc_proptax_reliance”
 - This program creates Appendix Figure A4.1.
- `graphs_prev_metrics.R` – Creates scatterplots of TFQ versus economic segregation and jurisdictional fragmentation. Also creates Table 1 in the main text, summarizing the top and bottom 10 CBSAs by TFQ.
 - Output: All files in “output/prev_metrics”. Clean metro-level summary data (including the tax base fragmentation quotient and measures of economic segregation and jurisdictional fragmentation) is output to “output/clean_data/cbsa_summary_gen_purpose_govts.csv”. This data is only available for our baseline specification.
 - This program creates the data used in Table 1 of the main text and Appendix Table A6.1. It also creates Appendix Figures A5.1 and A5.2.
- `graphs_sc_robustness.R` – Creates scatterplots of baseline versus alternate specifications for the Supplementary Appendix.
 - Output: All files in “output/sc_robustness”
 - This program creates Appendix Figures A3.1.2, A3.2.2, and A3.3.2.

References

- Almy, Richard R. 2000. “Property Tax Policies and Administrative Practices in Canada and the United States: Executive Summary.” *Assessment Journal* 7(4):41–57.
- CoreLogic. 2024. “Parcel Data.” <https://www.corelogic.com/360-property-data/parcel-data/>.
- Manduca, Robert, Brian Highsmith, and Jacob Waggoner. 2025. “Tax Base Fragmentation as a Dimension of Metropolitan Inequality.” *Socio-Economic Review* mwaf055. doi:10.1093/ser/mwaf055.
- Pierson, Kawika, Michael L. Hand, and Fred Thompson. 2015. “The Government Finance Database: A Common Resource for Quantitative Research in Public Financial Analysis” edited by F. Emmert-Streib. *PLOS ONE* 10(6):e0130119. doi:10.1371/journal.pone.0130119.
- Schroeder, Jonathan, David Van Riper, Steven Manson, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. 2025. *National Historical Geographic Information System: Version 20.0*: Version 20.0. Minneapolis, MN: IPUMS. doi:10.18128/D050.V20.0.