Author: Kien Hoang-Le

University of Nottingham

Created date: 20 July 2023

Description: Documents record all the variables which are used in my dissertation, and can be found in the data file *reg\_panel.dta.*

--------------------------------------------------------

The panel data includes 3 main categories of variables: (1) mortgage variables from the HMDA dataset, (2) toxic releases variables from the TRI dataset and (3) census and macroeconomic variables on the county-level from multiple sources.

The following table provides details:

|  |  |
| --- | --- |
| Variables from Toxic Release Inventory dataset | |
| **Variables name** | **Sources and description** |
| Frs\_id | Facility ID made by TRI |
| State, state\_code, county, fips, year, year\_fips | Geographical Identifiers |
| Latitude, longitude | Geographical Identifiers |
| Carcinogen | Binary variable, indicating the type of chemical substance observed at a specific datapoint is carcinogen or not |
| Classification | 3-level variables. Indicates if the chemical is classified as a dioxin or dioxin-like compound, a Persistent Bioaccumulative and Toxic chemical, or a general EPCRA Section 313 chemical.  Values: {TRI, PBT, DIOXIN} where:   * TRI = General EPCRA Section 313 Chemical * PBT = Persistent Bioaccumulative and Toxic * DIOXIN = Dioxin or Dioxin-like compound |
| Fugitive\_air | An estimate of the total quantity of the toxic chemical released as fugitive air emissions at the reporting facility. |
| Stack\_air | An estimate of the total quantity of the toxic chemical released as fugitive air emissions at the reporting facility. |
| Onsite\_release\_total | Total quantity of the toxic chemical released to air, water and land on-site at the facility |
| Total\_releases | Total onsite release and total offsite release |

**Note on TRI dataset:**

* The family of variables “releases” make up the main treatment variable

|  |  |
| --- | --- |
| Variables from HMDA dataset | |
| **Variables name** | **Sources and description** |
| Lei | Lender’s ID |
| Rate\_spread | The difference between the covered loan’s annual percentage rate (APR) and the average prime offer rate (APOR) for a comparable transaction as of the date the interest rate is set |
| US30\_spread | The difference between the covered loan’s annual percentage rate (APR) and average 30-Year US treasury bond yield of that year |
| Applicant\_race1 | Race of the first applicant or borrower. Aggregated based on the HMDA data   * 2: Asian * 3: Black and African American * 5: White * 14: Other |
| Applicant\_age | Age of the first applicant or borrower:   * 7 age bins, from under 25 to over 74 |
| loan\_amount,  dec\_loan\_amount | The amount of the covered loan, or the amount applied for.  Divided into decile bins |
| loan\_to\_value\_ratio,dec\_loan\_to\_value | The ratio of the total amount of debt secured by the property to the value of the property relied on in making the credit decision.  Divided into decile bins |
| property\_value,  dec\_property\_value, | The value of the property securing the covered loan or, in the case of an application, proposed to secure the covered loan, relied on in making the credit decision.  Divided into decile bins |
| income,  dec\_income, | The gross annual income, in thousands of dollars, relied on in making the credit decision, or if a credit decision was not made, the gross annual income relied on in processing the application. Divided into decile bins |
| Loan\_to\_income,  Dec\_loan\_to\_income | Loan amount/ Income  Divided into decile bins |
| purpose | Purchase new home or refinancing   * 1 buy new * 2 refinance |
| **Newly created** | |
| overconf\_limit | 2 if the loan amount is higher than the FHFA conforming loan limit.  1 if the loan amount is smaller than FHFA conforming loan limit |
| bank | 2 if the mortgage is initiated by bank lenders  1 if the mortgage is initiated by non-bank lenders |
| type\_of\_purchaser |  |

**Note on HMDA and mortgage data:**

* Since all loan in the HMDA dataset is conforming loan, they need to follow the FHFA requirement of conforming loan limit values – which is a ceiling loan amount for mortgages to be acquired by Fannie Mae and Freddie Mac. The limit is around 600k – 700k for most places, and ~1000k for high cost areas.

Source: <https://www.fhfa.gov/Media/PublicAffairs/Pages/FHFA-Announces-Conforming-Loan-Limits-for-2023.aspx>

**With this info,** it should be noted that for loan\_to\_value\_ratio might be low for very high property value.

|  |  |
| --- | --- |
| Variables from US Census data Source: **US Census Bureau** and **U.S Department of Labour, Bureau of Labour Statistics**  <https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html>  <https://www.bls.gov/lau/tables.htm#cntyaa>  <https://www.bls.gov/cew/downloadable-data-files.htm>  Tract level data from USDA: <https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes.aspx> | |
| **Variables name** | **Sources and description** |
| pop\_cou | 2020 Census total population of the County |
| aland\_cou,  aland\_cou\_sqkm | 2020 land area of the County (square meters) |
| houden\_cou | 2020 housing unit density of the County (square miles) |
| Cnty\_unemp\_rate | County’s unemployment rate in each year |
| Cnty\_avg\_income | County’s average income in each year |
| Property\_urb\_ru | Define whether a census tract is urban or rural |