1. *Stata code*
   This data library contains three Stata programs. *data.do* prepares the data; it defines key variables and sets up the sample for analysis. *match.do* is a routine called from data.do to generate a synthetic control sample based on a simple propensity-score algorithm. *HB4050_tables_and_figures_AEJPolicy.do* generates all of the tables and figures in the paper.

2. *Data files (publicly available)*
   We use publicly available tract-level demographic and employment status data from the 2000 Census and zip-level tax return data from the 2005 IRS Statistics of Income data. These data are used for construction of Table 1, Panel A. The Stata file *table 1A.do* contains details on each of the data sources and instructions on obtaining them. The datasets themselves (both in Stata and Excel formats) are included in this archive.

3. *Data file (contractually limited)*
   The loan-level data used in the paper is subject to contractual limitations by the data vendor (CoreLogic). Furthermore, the paper uses the non-public version of mortgage applications data collected under the Home Mortgage Disclosure Act (HMDA). The non-public version of the Home Mortgage Disclosure Act data is maintained by the Federal Reserve Board and is available for use by employees of the Federal Reserve System. The non-public HMDA data include precise dates of mortgage applications and originations, which is necessary for our sample design. A public version of the HMDA micro data is available. The public version of the data reports the origination year, but not the exact origination date. The public version of the data can be accessed here: https://www.consumerfinance.gov/data-research/hmda/explore

   The data available in this archive contains only zip-level summary statistics (mean and standard deviation) for key variables described in Table 1.C in the paper. It is presented both in a formatted Excel version and as a comma-delimited file. The Excel file also contains a data dictionary. Note that construction details for each variable are documented in *data.do*.

4. *Access to loan-level data*
   As mentioned above, the data used for this paper are subject to vendor contractual limitations and supervisory data restrictions. CoreLogic has partnered with a number of academic institutions in providing data for qualified academic research. Additional details are available at https://www.corelogic.com/solutions/university-data-portal.aspx

5. *Data glossary*
   As noted above, the Excel file contains a mapping between variable names used in tables and Stata code. The code itself documents construction of variables used in the analysis from raw data.

6. *Other Data Appendix files*
   - *Appendix Figures 2-3 Adjusted Default Calculations.xlsx* contains the code, all of the point estimates, and the resulting charts (Figures 2 and 3). These charts display actual and

model-predicted default rates for treatment and control groups measured over different horizons and the corresponding confidence bounds.

- The Appendix also contains a map of Chicago area zip codes that were selected for the HB4050 pilot, as well as of similar zip codes we chose to serve as the geographic control group.