Predators and Prey: Instructional Memo

K. Steven Brown, Jackelyn Hwang, and Michael Hankinson

Harvard University
Government 1016

May 9, 2011

Overview

The purpose of this memo is provide scholars with steps necessary to replicate our study of segregation's effect on subprime lending. All of our analyses come directly from the following data, either in original form or through some modifications, as explained below:

1) The replication dataset, as provided by authors Jacob Rugh and Douglas Massey. However, we only include an extract of their data. The extract contains only variables used in our final analysis, excluding a number of pay-for-access variables in their original data. These data are at the MSA/CBSA geographic level. All of the variables included in the extract are either publicly available to anyone (via the Census website and the Federal Housing Finance Agency) or publicly available to anyone affiliated with a major research library (e.g. the loan data from the Home Mortgage Disclosure Act website)

2) Demographic data for each tract included in the largest 100 MSAs - a total of 39,401 tracts. These demographic data were collected from the 2000 Census SF3 files and include the total population and racial and ethnic composition for each tract.

3) 2006 Subprime loan data as provided by the Home Mortgage Disclosure Act `http:`

//www.ffiec.gov/hmda/. Data from the last three years is publicly available on the website, but data prior to that (including ours) must be requested. We acquired our data from CDs available through the Harvard University Library.

We include 3 data files - 2 pre-analysis/raw files and 1 mid-analysis/modified file. The pre-analysis files are the extract for (1) "rugh massey extract.csv" and a merged datafile of (2) and (3) "tract data.csv" The "tract data.csv" file excludes tracts for which loan data was not reported (less than 5 percent of tracts). The third data file "cbsa data (publicview)" is a version of the extract file merged with variables from the tracts file and was used for the matching analysis (explained in paper and in attached code).

Data Creation and Modification

Our analyses had four steps, each outlined below.

1)

We created a new black and Hispanic dissimilarity index using the demographic data (2) and the following formula. We included this dissimilarity index into the tractsdata.csv file.

$$BHD = \frac{1}{2}\sum_{i}^{N} |\frac{(b_i + h_i)}{B + H} - \frac{w_i}{W}|$$

where b_i, h_i, and w_i are the black, Hispanic, and white populations, respectively, in a census tract i and B, H, and W are the black, Hispanic, and white populations, respectively, for the entire MSA.

2) - Spatial Methodology

The purpose of this section is to guide researchers through the steps necessary to compose our outcome variable, the CBSA Predation Index. Regarding software, this memo follows steps applicable to ArcMap 10.

- Begin by locating the appropriate set of tracts, as grouped by CBSA within our data set. For the purposes of our analysis, we spliced the entire data set into its 100 individual CBSAs. A bold researcher may try to run all 100 CBSAs at once within ArcMap, but let it be done at his/her own risk. For the purpose of this replication, we will assume that your splice your data into 100 individual .csv files.

- For your first CBSA, open the .csv and create a new column of subprime lending rate by dividing the "subprimecount" by "loanscount".

- Create a second new column to be used as a modified dummy variable for minority tract. This dummy variable will code 1 for a black-Hispanic population greater than 50 percent, a 0 for a black-Hispanic population less than 40 percent, and a range between 0 and 1 for a black-Hispanic population between 40 and 50 percent. To create this dummy, we used the following Excel code, based on "blhisptra"'s location in column U:

  $= IF(U2 >= 50, 1, IF(U2 <= 40, 0, (U2 - 40)/10))$.

- Save the newly updated workbook as an Excel 97-2003 Workbook, as these files seem to function better within ArcMap 10.

- Open a Blank Map within ArcMap and use the "Add Data" feature to import the newly updated CBSA file. Similarly, import the 2000 U.S. Census tract boundaries shapefile.

- To isolate the specific CBSA, join the .xls document to the census shapefile. First, right click the icon for the tracts shapefile within the Table of Contents. Next, select "Join and Relates" -¿ "Join". Within the "Join Data" menu, under "Choose the field in this layer that the join will be based on:" select "STFID", as this field is Numeric. Under "Choose table", select the CBSA's .xls file. Finally, under "Choose the field in the table to base the join on:", select "geoid". To eliminate extraneous census tract, mark the "Keep only matching records" icon. Click "OK".

- To confirm the join, export the tracts shapefile by right clicking the shapefile icon, selecting "Data", then "Export Data". Save the output under a new name and be sure to save it as a shapefile. ArcMap will offer to add the exported data to the map as a layer. Select "Yes" and right click the new icon, followed by "Zoom To Layer".

- To identify the clustering of minority neighborhoods, first open the ArcToolbox, following the toolbox offerings from "Spatial Statistics Tools", to "Mapping Clusters", and finally to "Cluster and Outlier Analysis (Anselin Local Moran's I)".

- Within the cluster analysis menu, define your "Input Feature Class" as your recently exported CBSA shapefile. For "Input Field", select the label for your newly created dummy variable for minority tract. Under "Output Feature Class," define the new shapefile's name (minoritycluster). Leave the standardized settings of "INVERSEDISTANCE" and "EUCLIDEANDISTANCE". ArcMap should add the clustering shapefile as a new layer.

- For subprime lending rate, repeat the previous step, but for "Input Field", select your recently created column of subprime lending rate.

- Join the "subprimecluster" shapefile to the your first exported file of all the CBSA data. Use the same process as before, but select "FID" for both the layer field and the

table field. Be sure to "Keep all records".

- To create a shapefile of solely minority clusters, select all of the "HH" minority tracts, then use the "Export Data" command. To simplify, right click the "minoritycluster" shapefile and "Open Attribute Table". Scroll right and right click the heading for "COType IDW", selecting "Sort Descending". Using the Shift key, select both the first listing of "HH" and the last listing of "HH". Now that the data is selected, right click the "minoritycluster" shapefile again, and export the data as a new shapefile.

- Repeat the previous step, but use the base CBSA shapefile to which we joined the "subprimecluster" data. We joined the two files so that now our subprime data also contains population statistics, vital for creating the Predation Index.

- Now we should have two shapefiles of clusters, one for minority tracts, the other for subprime lending tracts. To find where they overlap, click the "Geoprocessing" drop-down menu and select "Intersect". Under "Input Features", select both of the new cluster shapefiles. Check to save your file under a memorable name. Click "OK". Your new shapefile of intersection should appear in the Table of Contents.

- To find the Predation Index, open MS Excel, or a similar data processing application. In Excel, open the appropriately compatible version of the intersect shapefile and use the summation function to find the total population of the intersection. Repeat the same steps for the exported shapefile of only clustered subprime tracts to find the total population of the subprime lending targeted areas.

- To calculate the Predation Index, divide the population of the intersection by the population of the entire subprime lending market.

- For future analysis, join the shapefile of only clustered minority tracts to the base CBSA shapefile, keeping all records. In Excel, create a new column granting a dummy

variable to whether the tract is part of a cluster. For instance, if column Z contains the COType, the new column's formula would read: $= IF(Z = "HH", 1, 0)$.

3)

We conducted a fixed effects analysis, examining differences in mean subprime lending rates by inclusion in a cluster of minority tracts, controlling for CBSA level controls.

- We used the merged data "tract data.csv".
- See code labeled "tract fixedeff.R".

4)

Our final step used CBSA level data for a matching analysis, using a datafile that merged the original Rugh and Massey file with the new dissimilarity index and an "overlap" variable created from the GIS analysis.

- See code labeled "msamatching final (publicview)".

Any further questions can be forwarded to Steven Brown (ksbrown@fas.harvard.edu), Michael Hankinson (mhankins@fas.harvard.edu), and Jackelyn Hwang (jihwang@fas.harvard.edu)