**Measures of Bias in Mortgage Lending**

**Summary & Technical Notes**

**Mortgage Lending Bias Mapper**

Last updated: 12/12/22

**Why measure mortgage lending bias?**

Research has shown that neighborhoods, including the built, social, and natural environments they contain, are influential determinants of health and health inequity.1 Among these, residential racial segregation has been recognized as a fundamental neighborhood determinant.2 Racial segregation results in fundamentally different exposures and experiences based on race, impacting health disparities. Further, evidence indicates that there are wide gaps in home ownership and home equity derived wealth by race and ethnicity in the United States.

An important driver of residential segregation, and the economic health of a neighborhood or an individual, is mortgage lending. Mortgage lending bias – the systematic denial of mortgage financing to specific neighborhoods or applicants – can thus have important implications for housing access, wealth accumulation, economic development, and segregation. By promoting inequitable economic investment across neighborhoods, mortgage lending bias can promote inequities in neighborhood quality, housing quality, and access to resources. By promoting inequitable economic investment across applicants of different racial or ethnic identities, mortgage lending bias can promote inequities in wealth accumulation, housing tenure, and socioeconomic status.

**How do we measure mortgage lending bias?**

Our team at the Medical College of Wisconsin created three primary mortgage lending bias metrics to help researchers and the public measure bias in access to mortgage lending across neighborhoods and individuals. These metrics can be used to identify bias, map spatial patterns of bias, and link bias with health outcomes or other socioeconomic metrics. We developed these measures using a combination of the disease mapping method adaptive spatial filtering (ASF) and logistic regression models predicting application denial. Methods are described in detail below. The three bias metrics are:

* **Location bias or redlining**: the odds ratio of denial of a mortgage application for a property in a local area, compared to properties across the metropolitan statistical area (MSA)
* **Racial bias**: the odds ratio of denial of a mortgage application for a Black/African American applicant, compared to a non-Hispanic white applicant, for a property in the local area.
* **Ethnic bias**: the odds ratio of denial of a mortgage application for a Hispanic applicant, compared to a non-Hispanic white applicant, for a property in the local area.

**How are metrics (for 2010-2017) calculated?**

Data source and variables: The data source is the Home Mortgage Disclosure Act (HMDA) database, which contains mandatorily reported data on mortgage lending practices.3 HMDA data includes the approval/denial status of an application; race, ethnicity, and sex of the primary applicant; loan amount; and income of the applicant(s). These are the key variables used for estimation of mortgage bias metrics. Only conventional loans for new purchases of owner-occupied homes between 2010-2017 with no missing information on these variables are included. Information on the census tract of the property for which a mortgage was requested enables analysis at a relatively small geographical unit. We aggregated data over several years to provide additional spatial specificity in calculated estimates. Mortgage lending bias metrics are available for 389 Metropolitan Statistical Areas delineated in 2017.

Geographic units: Metrics are calculated based on two geographic units – the census tract (CT), and the metropolitan statistical area (MSA). Census tracts are selected as this is the smallest geographic identifier available for HMDA data. It is also a well-regarded geographic unit as it is relatively small and was built by the US Census Bureau for the purpose of data collection and analysis. The MSA is another US Census Bureau unit; MSAs represent aggregations of US counties and approximate the commuting areas and housing markets of major US cities. Given the focus on housing in this work, we elected to use MSAs as both a reporting unit (for racial and ethnic bias) and a comparison unit for the calculation of location bias/redlining.

Estimation process. Mortgage lending bias metrics are estimated using logistic regression modeling in an adaptive spatial filtering (ASF) framework, as described in our 2017 paper in *Health & Place*.4 ASF is a process that uses a series of spatial filters, centered at each point of a grid placed over the study area, to pull in data from the surrounding geographic area to ensure a stable statistic calculated for the grid point.5 The radius of the filter is increased until the filter contains enough observations from geographic units (e.g. census tracts) surrounding the grid point to meet a specified threshold. Values of the statistic at all grid points are interpolated into a continuous surface. The grid utilized in these calculations consisted of all census block group centroid locations within a study area. A key step for implementing ASF is to set the “threshold” that will guide filter size; in other words, what is the minimum number of individuals needed in order to make a valid estimate? The threshold for each metric is described below.

*To estimate the racial bias measure*, only data within a single spatial filter is used to estimate a logistic regression model for a grid point. The logistic regression model predicts approval vs. denial of a mortgage application and includes the following variables as covariates: race and ethnicity of the primary applicant (Black and non-Hispanic white applicants only), sex of the primary applicant, and the loan amount to income ratio. The racial bias measure is essentially the exponentiated coefficient of the race/ethnicity variable, which can be interpreted as the odds ratio of denial of a mortgage application if the primary applicant is Black, compared to Non-Hispanic White. *A parallel approach is taken to estimate the ethnic bias, measure comparing Hispanic to non-Hispanic White applicants*. The threshold (minimum number of applicants) that guides each filter size is determined as follows. Because the estimation model compares race/ethnicity (of two types) and application status (approved or denied), the threshold is defined based on a 2x2 table of these two variables. At least two individuals are required for each cell of the 2x2 table. Specifically for the racial bias index, the data in the filter must include at least two approved Black, two approved white, two denied Black, and two denied white applicants. The same threshold criterion applies when calculating the ethnic bias index.

*To estimate the location bias (or redlining) measure*, we estimate a model at each grid point *including data for the whole MSA*, instead of limiting calculations to data within a single filter. Each model includes a variable indicating whether the property for which the mortgage is being requested is inside the local spatial filter or not. That variable is then interpreted as the odds ratio of denial of a mortgage application if the application is for a local property, compared to properties in the rest of the MSA. In contrast to the racial and ethnic bias indices, the model constructed for location bias does not include the race or ethnicity variable because the interest is in bias based on the location of the property, regardless of the race or ethnicity of the applicant. The threshold criterion is four denied and four approved mortgages, which is chosen to keep the number of applications (eight in total) similar to the number used for the race and ethnicity-based bias measures.

*Quality checks and adjustments*: The model used for estimation of all three metrics initially contains the primary applicant’s sex as an adjustment variable. However, in developing the metrics, we found that imposing sample size requirements in the four cells by sex (which would be necessary for models to be calculated) leads in some cases to unreasonably large geographic filter sizes. We therefore adopted a strategy to ensure stability of model estimates. Specifically, we omit sex from the model if there are indications that model estimates are unstable: (a) the coefficient estimate for sex is greater than 10 or less than -10, or (b) the standard error for sex is greater than 50. After issues regarding the sex variable are resolved, it is possible that the model estimated for a specific grid point will still be unstable (i.e. the standard error for the variable of interest—race, ethnicity or location—is greater than 5). If this occurs, we do not include the model result for this grid point in further analyses. This is rare. In calculating the 2010-2017 metrics, this only occurred 314 times for racial bias estimation and 87 times for ethnic bias estimation (with a total of 194,111 block group centroids estimated). Finally, to further guard against unstable or extreme values, we applied a rule to exclude extreme outliers when estimating the racial and ethnic bias measures. For each MSA, IQR is calculated as the difference between the 75th and 25th percentiles obtained from the estimates for all the block group centroids of that MSA, and any grid point coefficient estimates that are more than 5 IQR below the 25th percentile or more than 5 IQR above the 75th percentile are excluded (45 cases for racial and 34 for ethnic bias measure).

**Data Limitations & Future Directions**

There are several limitations of these metrics; our team is working on each of these issues as future directions.

1. It is well-known that loan approval/denial could be affected by an applicant’s credit scores, educational level, and employment status. However, these important individual characteristics are not available in the HMDA database for 2010-2017. Therefore, the current estimates of racial, ethnic, and location bias should be interpreted with caution. Our plan is to continue to improve the estimates by leveraging the most recent HMDA datasets and incorporating loan-to-value ratio, interest rate, total loan cost, applicants’ age, and other information.
2. We have only calculated metrics using information about the primary applicant. Incorporating data on the secondary applicant may provide improved precision in these metrics and is an active area of interest.
3. We have only calculated racial and ethnic bias metrics for two comparisons (Black to non-Hispanic white, and Hispanic to non-Hispanic white). In many places, these groups are the only racial and ethnic groups with sufficient numbers to calculate these geospatial metrics. However, we are working with partners in California and Hawaii to expand these calculations.
4. While racial and ethnic bias can be estimated for all tracts across the United States, location bias/redlining is based on a direct comparison of the local spatial filter to the MSA containing the property. This was done deliberately, as the intent was to create a measure of contemporary redlining, which is historically understood as an urban phenomenon. However, there are clear needs for expansion of these metrics and ideas to include rural areas and American Indian reservation land.
5. The location bias/redlining metric is calculated without including the race or ethnicity of the applicant in the model. This is intentional, as it the goal is to measure bias against the location, regardless of the race or ethnicity of the applicant. *However*, while the calculation of redlining does not include this information, the measure itself is definitely about race and ethnicity. Historically, “redlining” is understood as the systematic denial of mortgage funding to specific neighborhoods, and this systematic denial was based in large part on the race and/or ethnicity of the neighborhood’s residents. Further, our “applicant race agnostic” calculations reveal that across the country, neighborhoods with high location bias/redlining also have higher percentages of residents of color.

**Additional notes**

Estimates for racial bias are not available for census tracts in the Grants Pass, OR, Lewiston, ID-WA, and Bismarck, ND MSAs, as insufficient data was available to support model estimation. Interactive maps and charts were rendered using R Shiny6 and Leaflet.7

**Suggested citation for use of metrics**

Kirsten M. M. Beyer, Yuhong Zhou, Shana Terai Lara, Emily McGinley, Sima Namin, Bethany Canales and Purushottam W. Laud.Mortgage Lending Bias Metrics 2010-2017. Mortgage Lending Bias Mapper at the Medical College of Wisconsin. December 2022. Available at: www.mcw.edu/bcrp.

***Let us know what you find!***

If you elect to use the measures, we would be happy to receive any updates on their use and your findings. Please send any updates to Dr. Kirsten Beyer at kbeyer@mcw.edu.

***References***

1. Gomez, S.L., et al., The impact of neighborhood social and built environment factors across the cancer continuum: Current research, methodological considerations, and future directions. Cancer, 2015. 121(14): p. 2314-30.
2. Williams, D.R. and C. Collins, Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health. Public Health Reports, 2001. 116(5): p. 404-416.
3. Federal Financial Institutions Examination Council (FFIEC) [Internet]. The Home Mortgage Disclosure Act. Available from: <https://ffiec.cfpb.gov/> (last time accessed on December 1, 2022).
4. Beyer, K, Y Zhou, K Matthews, A Bemanian, P Laud, A Bemanian, and A B Nattinger. 2016. New Spatially Continuous Indices of Redlining and Racial Bias in Mortgage Lending: Links to Survival after Breast Cancer Diagnosis and Implications for Health Disparities Research. *Health and Place* 40: p. 34-43.
5. Beyer, Kirsten M. M., C Tiwari, G Rushton. 2012. Five Essential Properties of Disease Maps. *Annals of the Association of American Geographers* 102 (5): p. 1067-1075.
6. Chang W, Cheng J, Allaire J, Sievert C, Schloerke B, Xie Y, Allen J, McPherson J, Dipert A, Borges B (2022). \_shiny: Web Application Framework for R\_. R package version 1.7.2, <https://CRAN.R-project.org/package=shiny>.
7. Cheng J, Karambelkar B, Xie Y (2022). \_leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library\_. R package version 2.1.1, <https://CRAN.R-project.org/package=leaflet>.