THE CENTER FOR INDIAN COUNTRY DEVELOPMENT

WORKING PAPER SERIES

NO. 1904

The Value of HMDA Coverage of Home Lending in Rural Areas and Indian Country

Former CICD Data Analyst
Kevin Johnson

Former Vice President and Advisor to the CICD
Dr. Richard M. Todd

July 31, 2019

The Value of HMDA Coverage of Home Lending in Rural Areas and Indian Country

Kevin Johnson and Richard M. Todd[1]

July 2019

The Home Mortgage Disclosure Act (HMDA) was enacted by Congress in 1975 and was implemented by the Federal Reserve Board's Regulation C.[2] This regulation applies to a wide array of financial institutions involved in home lending—such as most banks, savings associations, credit unions, and other mortgage lending institutions—and requires them to annually report data on the applications they receive from borrowers seeking home loan financing (including for mortgage refinancing or home improvement, as well as for home purchase). Due to the breadth and detail of HMDA's coverage, the HMDA data are viewed as "the most comprehensive source of publicly available information on the U.S. mortgage market." Bhutta, Laufer, and Ringo (2017). However, HMDA's coverage is not complete, as many small and nonmetropolitan lenders are exempt from reporting under HMDA. This has led to concerns that the HMDA data are not reliable for monitoring the patterns and trends in rural home lending. The existing evidence on this issue is limited and mixed. We provide new evidence and conclude that, despite significant coverage gaps in some areas, HMDA is a useful and important source of information about rural home lending, including for most of the American Indian reservations in rural areas.

Background on HMDA's home lending data

The data collected annually under HMDA provide detailed information about the home loans originated each year and are also the primary source of information on home loan applications, including the sex, race, and ethnicity of loan applicants and other information in applications that do not result in a loan origination. According to Bhutta, Laufer, and Ringo,

> The HMDA data [provide] unique details on…the disposition of each application for mortgage credit; the type, purpose, and characteristics of each home mortgage that lenders originate or purchase during the calendar year; the census-tract designations of the properties related to those loans; loan pricing information; personal demographic and other information about loan applicants, including their race or ethnicity and income; and information about loan sales.

However, as also noted by Bhutta, Laufer, and Ringo,

> … the HMDA data do not provide universal coverage of residential mortgage lending in the United States. There are two main reasons HMDA coverage is not universal. First, not all lenders are required to report data. Among deposit-taking institutions like banks, the smallest institutions as well as institutions without any

[1] Respectively, former intern and retired advisor for the Center for Indian Country Development at the Federal Reserve Bank of Minneapolis. For their advice and assistance, we thank Tedd Beulow, Neil Bhutta, Kent Weyrauch, and Michael Williams, but we retain responsibility for any errors.

[2] In 2011, the rule-writing authority of Regulation C was transferred to the Consumer Financial Protection Bureau(CFPB).

> branches in a metropolitan statistical area (MSA) do not have to report data. Among institutions that take no deposits, nonprofits, smaller institutions, and those that operate entirely outside of an MSA also do not have to report data.[3]
>
> Second, not all types of mortgage originations are reported. In particular, lenders do not report mortgages that are not for the purpose of purchasing a residential property, refinancing an outstanding mortgage, or making home improvements. Thus, a mortgage taken out solely to finance education expenses, for example, would not be reported. In addition, home equity lines of credit (HELOCs), regardless of their purpose, are not required to be reported under current rules[4]… According to estimates based on the consumer credit records maintained by Equifax, one of the three nationwide consumer credit-reporting agencies, the number of first-lien home-purchase and refinance loans in the HMDA data is approximately 90 percent of the number reported in consumer credit files.[5]

Mixed evidence on HMDA's coverage of rural lending

In light of HMDA's urban focus and incomplete coverage of home lending, concerns have been expressed about the quality of HMDA data in rural areas. HMDA's rules on which lenders must report data were not designed to produce spatially representative statistics but rather to focus attention on large lenders in urban areas, where the historical problem of mortgage lenders "redlining" (discriminating against) certain neighborhoods was prominent. Accordingly, small mortgage lenders and those with limited activity in metropolitan areas are exempt from reporting on their home loan applications. If these lenders account for a large share of home loans in many rural areas, then the HMDA data may not accurately represent mortgage lending trends or patterns in those areas. The Housing Assistance Council (2010) articulated the logic of this concern by noting that "of the 989 FDIC-insured lending institutions with assets totaling less than the HMDA filing threshold in 2009, 70 percent were headquartered in rural counties." HMDA data experts from the staff of the Board of Governors of the Federal Reserve System (Avery, Brevoort, and Canner 2007, p. 357) reported that "rural areas are often dropped from the analysis of HMDA [because] reporting exceptions lead to significant distortions in the coverage of rural areas in HMDA."

---

[3] Bhutta, Laufer, and Ringo clarify that, under the rules in force in 2017, depositories with less than $44 million in assets and nondepositories that had less than $10 million in assets and originated fewer than 100 home-purchase and refinance loans in the previous year were not required to report. For additional details, see Federal Financial Institutions Examination Council (2017), "A Guide to HMDA Reporting: Getting It Right!" web page, https://www.ffiec.gov/hmda/guide.htm.

[4] These limits have changed. Bhutta, Laufer, and Ringo note that "Beginning on January 1, 2018, covered loans under the HMDA rule (Regulation C) generally will include closed-end mortgage loans and open-end lines of credit secured by a dwelling. For more information, see CFPB (2017), "Home Mortgage Disclosure Act Rule Implementation: Resources to Help Industry Understand, Implement, and Comply with the Home Mortgage Disclosure Act and Regulation C," web page, https://www.consumerfinance.gov/policy-compliance/guidance/implementationguidance/hmda-implementation."

[5] Bhutta, Laufer, and Ringo state that the "dollar volume of first-lien home-purchase and refinance loans for one- to four-family properties reported under HMDA is about 94 percent of the dollar volume of first-lien home-purchase and refinance originations estimated by Equifax for 2016."

However, there have been few attempts to directly measure HMDA's coverage of rural lending. Scheessele (1998) developed an indirect measure based on FHA loan data, inferring that "HMDA…does not adequately measure mortgage market activity in non-metropolitan areas" from the fact that "loans in non-metropolitan areas account for 7.7 percent of FHA loans in 1996 which is approximately the percentage of FHA loans that were unreported to HMDA."

Williams (2015) took a more direct approach, "analyzing the relationship between loan origination levels in HMDA data and loan origination levels in a proprietary dataset known as the Federal Reserve Bank of New York/Equifax Consumer Credit Panel (CCP)" that is based on a 5 percent random sample of credit history files. Williams finds (p. 31) that "large parts of rural America remain underrepresented" but also estimates (his Figure 3) that HMDA's coverage of rural mortgage originations averaged around 80 percent in the early 2000s before rising to 90 percent or more in 2010–2012. Although lower than HMDA's level of coverage in urban areas, the average rural coverage rates that Williams estimates suggest that HMDA can provide useful information on nonmetropolitan home lending activity.

Recognizing that HMDA coverage may vary across rural areas, Williams also conducts a geospatial analysis that estimates coverage county-by-county. After spatially averaging the results (to mitigate the effects of sampling variation stemming from the 5 percent sample of credit files in the CCP), Williams concludes that "HMDA coverage is clearly lower than predicted [by the CCP data] in the central part of the United States, from western Montana through the Dakotas and down through the less populated parts of Texas. Additional pockets dot inland midwest, southeast, and Appalachian pockets of the country." Overall, then, Williams finds that HMDA coverage in rural America is (a) lower than the coverage rates typical of urban areas, (b) on average relatively high (80 percent or more) and rising (to 90 percent or more after 2010), but (c) variable within rural areas, with coverage dipping below the overall rural average in areas such as the Great Plains, western Corn Belt, and Appalachia.

New Approaches

We take a new look at the question of the coverage of HMDA data in rural areas by benchmarking HMDA's loan data against three other sources of information on loans originated. We begin with two relevant but partial comparisons, in both cases matching a specific category of government-backed mortgages reported under HMDA against comprehensive federal executive agency data on the total number of these loans. Our first comparison uses state-level data on single-family home loans guaranteed by the Rural Housing Service (RHS) of the U.S. Department of Agriculture (USDA). The second comparison uses county-level data on mortgages insured by the Federal Housing Administration (FHA). Finally, we follow Williams in using the credit history data in the CCP, but with a different approach.

RHS-guaranteed loans provide a relevant benchmark for HMDA's rural coverage, because they are only available in rural areas and are overwhelmingly used for the purpose of purchasing a home.[6] We find that overall, and in most states, the number of RHS loans reported in the HMDA data are a large proportion—typically 80 to 90 percent—of USDA's figures on total RHS loans.

[6] According to a communication from USDA staff and the presentation at www.hudexchange.info/course-content/housing-counseling-webinar-overview-of-usda-housing-programs/Housing-Counseling-Webinar-Overview-

The data on FHA-insured loans are by county. When we add up the figures on FHA-insured loans for all rural counties in each state, we again find that the HMDA data represent 80 percent or more of total loans in almost all states. When we look at individual counties, we naturally see a wider range of outcomes, including some with HMDA coverage ratios well below 50 percent. However, we show that very low HMDA coverage is the exception, not the rule, even in rural areas and Indian reservations.

For a broader evaluation, we expand on Williams' (2015) idea of comparing HMDA data on rural loan originations to estimates of originations derived from CCP credit histories. We again find that, overall, HMDA's coverage in rural areas is near to or above 80 percent, and in some areas above 90 percent. However, like Williams, we also conclude that coverage falls below these levels in some other regions.

Using RHS loan originations to assess HMDA coverage

RHS manages the Section 502 Single Family Housing Program (SFHP) that provides "direct loans or loan guarantees to help low- and moderate-income rural Americans buy safe, affordable housing in rural areas."[7] Direct loans (i.e., in which RHS is the lender) made under this program are not included in the HMDA data.[8] The SFHP's guaranteed loans are reportable under HMDA, depending on whether the lender originating or purchasing them is required to report. These loans can only be used by owner-occupants to "[p]urchase new or existing homes and refinance existing Rural Development guaranteed or direct loans"[9] and are the only RHS loans (apart from the not-HMDA-reported SFHP direct loans) for purchases or refinancings by owner-occupants. Based on these characteristics, we compare RHS and HMDA data nationally and by state. For RHS activity, we use USDA data on annual total loan numbers and amounts for SFHP guaranteed loans. We compare these figures to HMDA's annual figures on loans whose "type" is classified as "FSA/RHS," meaning loans guaranteed by RHS or the Farm Service Agency (FSA).[10] We limit this comparison to loans to owner-occupants for first-lien home purchase or mortgage refinance.

---

of-USDA-Housing-Programs-Slides.pdf, over 90 percent of this USDA program's guaranteed loans are first-lien home-purchase loans for primary residences, the remainder being mainly refinances of the same type of guaranteed loan (with essentially no second liens and no home-improvement loans or other loan types). The RHS loans reported in HMDA in 2012–2016 support this characterization; only about 5 percent are for refinancing and less than 0.6 percent are reported (perhaps erroneously) as home-improvement loans.

[7] See www.rd.usda.gov/about-rd/agencies/rural-housing-service. For this program, rural areas are defined in section 520 of the Housing Act of 1949, as amended. Although once simpler and based on the population of nonmetropolitan areas, the definition has become somewhat more complicated due to special factors, such as the grandfathering of previously eligible areas that would otherwise have lost eligibility. Accordingly, USDA provides an online tool for checking which addresses are eligible; see eligibility.sc.egov.usda.gov/eligibility/welcomeAction.do?pageAction=sfp&NavKey=property@11.

[8] Personal communication from USDA staff.

[9] *USDA Rural Development 2016 Progress Report*, p. 70, at www.rd.usda.gov/files/USDARDProgressReport2016.pdf.

[10] The "FSA" portion of this HDMA loan type label may be an anachronism, since FSA has not offered significant volumes of home mortgages in recent years. As noted above, SFHP direct loans are not included in the HMDA data, and SFHP guaranteed loans are the only RHS loans for home purchases and refinancing by owner-occupants that are reported in the HMDA collection. Thus, it is appropriate to compare HMDA's figures for owner-occupants' home purchases plus refinancings in the FSA/RHS loan type category to USDA's figures on SFHP-guaranteed loans.

We obtain USDA's figures for SFHP-guaranteed loans from the *USDA Rural Development 2016 Progress Report*.[11] This document provides data by state for the fiscal years 2016 and 2015 and the fiscal-year period 2009–2014. We also calculate the totals for the fiscal-year period 2009–2016. We match these figures against the publicly available HMDA data for calendar years 2016, 2015, and the calendar-year periods 2009–2014 and 2009–2016. Because the mismatch between the fiscal-year period of the USDA data and the calendar-year period of the HMDA data is most significant for individual years and shorter time periods generally, we focus on the broadest averages, for 2009–2016.

We find that HMDA covered 83 percent of the number of SFHP-guaranteed loans and 87 percent of the dollar value of SFHP-guaranteed loans that USDA reported over 2009–2016.[12] These figures are in line with Williams' results for HMDA's general coverage of rural lending and not far below 90 percent, the general rule of thumb for the share of mortgage applications covered by HMDA. Overall, then, HMDA's coverage of lending activity under the SFHP appears sufficient to provide fairly reliable aggregate information about this program, which operates solely in rural areas.

Next, we assess HMDA's coverage of SFHP activity by state, to see if there are significant regional gaps in coverage. Figure 1 shows ratios based on total loan numbers by state.[13] These ratios are again not far below 0.90. Only five states—Nebraska, Wyoming, Illinois, Hawaii, and Vermont—have ratios below the range of 0.80 to 1.00 that is marked by the parallel red lines, and only in Nebraska and Wyoming do the ratios fall below 0.70, bottoming out near 0.60. These results are generally positive—HMDA's coverage of the SFHP's rural lending is near or above 80 percent in most states, where 80 percent is a level of coverage we regard as sufficient to capture the broad features and general trends in lending activity, even if not all details and nuances. Even in Nebraska and Wyoming, HMDA covers the majority of SFHP activity. However, the results for those two states demonstrate that certain areas can have significant HMDA coverage gaps even when coverage is above 80 percent overall and in most places.[14]

Although HMDA's generally adequate coverage of SFHP's rural loans is encouraging, it is not clear what this implies for rural lending generally. Based on HMDA data, SFHP-guaranteed loans were a small share (about 5 percent) of HMDA's total home loans in nonmetropolitan counties in 2009–2016, so the direct implications of the findings above are very limited. Furthermore, it is not clear that HMDA's coverage of SFHP lending is representative of HMDA's coverage of rural home lending generally. For example, SFHP-guaranteed loan originators must apply to USDA for pre-approval. It is possible that this administrative burden alone might discourage smaller lenders from applying and skew SFHP participation toward

[11] See www.rd.usda.gov/files/USDARDProgressReport2016.pdf.
[12] This is true whether or not we include data for Puerto Rico.
[13] Results based on loan amounts are very similar.
[14] The results for individual years are generally similar overall, although ratios corresponding to those in Figure 1 can take more extreme values in some states and years. For example, the ratio (HMDA/USDA) is especially high using 2016 data, with many states' values edging above 1.00. This may reflect timing differences between fiscal and calendar years. By contrast, the lowest state ratios in 2015 and 2016 were near or above the lowest values for 2009–2014. We also note that if SFHP data were available at the county level, we would expect significantly wider dispersion in the HMDA coverage ratios, especially in counties with low volumes. The pattern of wider dispersion in county ratios as compared to state ratios is clear in the FHA results below.

larger lenders that are more likely to be HMDA reporters, in turn pushing HMDA's coverage of RHS loans above its overall coverage of rural loans. Partly for that reason, we report next on a broader category of loans—those insured by the FHA—as well as, further below, on a more complicated but also more comprehensive assessment of HMDA's rural coverage.

Using FHA loan originations to assess HMDA coverage

The FHA, a federal agency within the U.S. Department of Housing and Urban Development (HUD), provides mortgage insurance on loans made by FHA-approved lenders throughout the U.S. and its territories. FHA insures mortgages on single-family and multifamily homes, including manufactured homes, nursing-care facilities, and assisted living facilities. The data we analyze are by county for 2013 and 2014, primarily from FHA's 203b program but also its 203k program.[15]

The data cover loans for single family homes (detached and attached) and also manufactured housing loans and loans on two-to four-unit properties. This represents essentially all FHA home-purchase loans reportable under HMDA, except for loans on multifamily structures (five or more units) and certain reverse mortgages. We exclude multifamily loans when we extract comparison data from HMDA.

Because of the possibility that some loans might be recorded in one calendar year in the FHA data but an adjacent year in the HMDA data, we again aggregate the years available. That is, we focus on the FHA and nearly corresponding HMDA data for the combined years 2013–2014.[16]

We mainly report results based on the number of FHA-insured home-purchase loans. We obtained similar results using the number of FHA-insured refinancings and the dollar value of FHA-insured home-purchase or refinance loans.

Across all counties in the 50 U.S. states and District of Columbia, we estimate that HMDA covered 95 percent of the number of FHA-insured loans and 96 percent of the dollar value of FHA-insured loans in 2013–2014. When we limit our calculations to only nonmetropolitan counties,[17] these ratios drop slightly, to 89 percent of the number of loans and 91 percent of their dollar value. These figures for the rural portions of states are again in line with Williams' results for HMDA's general coverage of rural lending and very near the general 90 percent rule of thumb for the share of mortgage applications covered by HMDA. As for SFHP loans, HMDA's coverage of FHA lending activity appears sufficient to provide fairly reliable information about FHA-insured mortgages in rural areas collectively.

Next, we assess HMDA's coverage of FHA-insured mortgages in the "rural" (technically, nonmetropolitan) portion of each state. We define the rural portion of a state by combining all of

[15] We thank Neil Bhutta of the staff of the Board of Governors of the Federal Reserve System for providing the data and key insights into its nature.
[16] Neil Bhutta aligned the FHA data he provided to us to calendar year periods, but some timing discrepancies may still exist.
[17] In our 2013–2014 data, 47 states have nonmetropolitan counties with HUD-recorded FHA 203b- or 203k-insured loans. Delaware, New Jersey, and Rhode Island have no such counties. The District of Columbia and Puerto Rico are also omitted from our analysis of FHA mortgages.

its nonmetropolitan counties, those with a USDA 2013 rural-urban continuum code (RUCC) of 4 or more.[18] The results are similar to those for states using the SFHP data. Figure 2 shows HMDA's coverage ratio for FHA-insured home-purchase mortgages in the nonmetropolitan portion of each state. With a few exceptions, these ratios are near to or in excess of 0.90. Only three states—Nevada (67 percent), Wyoming (76 percent), and Nebraska (79 percent)—have ratios lower than 0.80. (The parallel red lines again mark the range of 80 to 100 percent.) However, the fact that two of these states, Nebraska and Wyoming, were also among the lowest states for HMDA's coverage of SFHP loans suggests a potentially broad pattern of relatively low HMDA coverage rates in these two states.

Figure 2 also shows three states with HMDA coverage ratios above 1.00, which in principle should not be possible. In reality, discrepancies in data reporting can occur between HUD and HMDA, such as when a loan in December–January is recorded in one year in the FHA dataset we use and reported in another year in the HMDA data collection. This type of error could readily account for ratios slightly above 1.00 in some counties, or even well above 1.00 in counties with a very small volume of loans (e.g., one actual FHA-insured loan that is correctly reported in HMDA plus one loan erroneously reported in HMDA as FHA-insured).

However, when the number of FHA-insured loans is large, ratios well above 1.00 seem to require a more systematic explanation. Small loan numbers and a ratio of just 1.05 may leave nonmetropolitan Hawaii (174 loans) as a borderline case, but it is harder to attribute the cases of nonmetropolitan Oklahoma (3,355 loans and a ratio of 1.06) and nonmetropolitan Alaska (443 loans and a ratio of 1.43) to random errors. We return to this below.

Next, we examine HMDA's coverage of FHA loans in each of the 1,906 nonmetropolitan counties for which we have data on FHA-insured loans in 2013–2014. Figure 3 displays the

---

[18] According to USDA (www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx),

> The 2013 Rural-Urban Continuum Codes form a classification scheme that distinguishes metropolitan counties by the population size of their metro area, and nonmetropolitan counties by degree of urbanization and adjacency to a metro area. The official Office of Management and Budget (OMB) metro and nonmetro categories have been subdivided into three metro and six nonmetro categories. Each county in the U.S. is assigned one of the 9 codes. This scheme allows researchers to break county data into finer residential groups, beyond metro and nonmetro, particularly for the analysis of trends in nonmetro areas that are related to population density and metro influence.

The RUCC documentation page (www.ers.usda.gov/data-products/rural-urban-continuum-codes/documentation/) further clarifies that

> metro counties are divided into three categories according to the total population size of the metro area of which they are part: 1 million people or more, 250,000 to 1 million people, and below 250,000. Nonmetro counties are classified along two dimensions. First, they are divided into three urban-size categories (an urban population of 19,999 or more, 2,500 to 20,000, and less than 2,500) based on the total urban population in the county … Second, nonmetro counties in the three urban-size categories are sub-divided by whether or not the county is adjacent to one or more metro areas. A nonmetro county is defined as adjacent if it physically adjoins one or more metro areas, and has at least 2% of its employed labor force commuting to central metro counties. Nonmetro counties that do not meet these criteria are classed as nonadjacent.

See Table 2, below, for the total number of counties and residents (as of 2010) in each RUCC group.

HMDA coverage ratios for these counties, grouping them by their RUCC codes from 4 (nonmetropolitan county with urban population of 20,000 or more and adjacent to a metro area) to 9 (nonmetropolitan county that is completely rural or less than 2,500 in urban population, and not adjacent to a metropolitan area). The horizontal axis in each panel shows the number of FHA-insured loans recorded, while the vertical axis shows the HMDA coverage ratio for these loans. Each dot represents a county, and the horizontal red lines mark the range of 80 to 100 percent coverage. Some counties have identical coverage ratios—including 11 RUCC 9 counties (out of 368) with ratios of 0.00—so overlapping points are "jittered," or shifted slightly in the horizontal direction, to make the presence of these multiple observations more visible. As a result, the number of FHA loans that appears to be associated with a given county's point may be somewhat inaccurate for counties that share the same HMDA coverage ratio.

We highlight four conclusions in light of Figure 3:

1. Our estimated HMDA coverage for FHA-insured loans is relatively high, 80 percent or more, in the vast majority (76 percent) of nonmetropolitan counties in 2013–2014.

2. A nontrivial minority of nonmetropolitan counties had much lower coverage for FHA-insured loans in this period. Forty counties, or about 2.1 percent of all nonmetropolitan counties for which we have data, had coverage of less than 50 percent. Another 68 counties had coverage less than 60 percent. And 341 more nonmetropolitan counties, or 18 percent of nonmetropolitan counties in our data, had coverage of at least 60 percent but less than 80 percent. In all, 449 nonmetropolitan counties, or 24 percent, had HMDA coverage below 80 percent.

3. To a small degree, the distribution of coverage ratios shifts downward as rurality and remoteness increase (i.e., as RUCC values increase). The share of counties with HMDA coverage below 80 percent rises to 29 percent for counties with RUCC equal to 9.

4. In another nontrivial minority of nonmetropolitan counties, HMDA coverage seems unaccountably high. Specifically, 178 counties, or about 9 percent of the total, had coverage equal to 105 percent or more. Although the volume of FHA-insured loans was very small in most of these 178 counties, many of them reported dozens or even hundreds of loans. As in Figure 2, we find examples of significant loan volume and HMDA coverage well in excess of 100 percent.

One possible explanation for the areas with high loan numbers and coverage ratios above 1.00 is that some lenders incorrectly report HUD 184 mortgages as FHA-insured home loans in their HMDA filings.[19] HUD 184 mortgages are designed to promote homeownership for American

[19] We thank Neil Bhutta for suggesting this explanation. He also noted a second possible explanation involving reverse mortgages. Reverse mortgages are not identified in HMDA and thus cannot be directly excluded from our HMDA data. This would tend to bias upward our HMDA coverage ratios for FHA-insured loans. The total volume of FHA-insured reverse mortgages in 2013–2014 was large relative to the number of FHA-insured loans in the data we use. However, the many reverse mortgages issued as home-equity lines of credit were subject to voluntary rather than mandatory HMDA reporting in 2013–2014. When reported, they would often be classified as junior lien loans, which we exclude from our HMDA data. In addition some reverse-mortgage lenders were not required to report. The fact that our overall estimate of HMDA's coverage of FHA-insured loans was 95 percent—not far above the

Indians and Alaska Natives and their communities. They are reportable in HMDA (subject to the usual criteria). Technically, they are not insured by the FHA and should be reported as "conventional" rather than "FHA-insured" loans in HMDA filings. However, they carry a credit guaranty from FHA's parent, HUD, which might lead some HUD 184 lenders to incorrectly report these loans as "FHA-insured."

We find evidence consistent with possible misreporting of HUD 184 mortgages as FHA-insured loans in HMDA. The two state nonmetropolitan areas with the most excessive FHA coverage ratios were in Alaska and Oklahoma, the same two states that, by far, dominate in the volume of HUD 184 loans.[20] At the county level, Figure 4 shows a clear tendency for high coverage ratios of FHA loans to occur in nonmetropolitan counties where a high percentage of the FHA loans reported in HMDA were made to American Indian and Alaska Native borrowers. The correlation coefficient between these two variables is 0.50 and highly statistically significant.[21] Further research on the extent to which HUD 184 mortgages are misreported as FHA-insured loans in HMDA seems warranted.

On net, we regard the results illustrated in Figure 3 as generally supporting the relevance of HMDA data in nonmetropolitan counties, despite the very low coverage ratios in a minority of counties. If we assume that even a 60 percent coverage ratio would provide useful information about the mortgage market, then only 108 out of the 1,906 nonmetropolitan counties we analyze, or 5.7 percent, fall short of this threshold.

Further analysis suggests that the HMDA coverage ratio for FHA-insured loans is also generally adequate in tribal communities. The most significant exceptions we identified, at least for FHA-insure loans, are the 18 counties summarized in Table 1. These nonmetropolitan counties were selected because they combine a HMDA coverage ratio less than 60 percent, an American Indian/Alaska Native population percentage greater than 2 percent, and proximity to a tribal community we were able to identify. They are spread far and wide, including 4 counties in Nevada; 2 each in North Carolina, North Dakota, and Washington; and 1 each in eight more states. The HMDA coverage ratio for FHA insured loans is below 50 percent in 8 of these counties, with a low of 24 percent in Pershing County, Nevada. However, the vast majority of American Indian reservations and Alaska Native homelands are not in these counties, including the majority of the largest tribal communities. The exceptional cases and the possibility that erroneous reporting of HUD 184 loans as "FHA-insured" inflates coverage ratios indicate that caution is appropriate when using HMDA data in Indian Country. Nonetheless, by the best measure we have now, HMDA's coverage of FHA-insured loans in 2013–2014 was at least 60 percent, and often above 80 percent, in the counties that overlap most reservations and homelands. And in the next section, we reach a similar conclusion about HMDA coverage in Indian Country using a method that is not affected by possible misclassification of HUD 184 mortgages.

---

estimate of 91.7 percent for HMDA's coverage of home loans nationally reported by Bhutta, Laufer, and Ringo—suggests that reverse mortgages had at most a small upward effect on our overall coverage estimates. But we cannot rule out more significant effects in some counties or perhaps in the nonmetropolitan areas of some states.

[20] Kunesh (2017).

[21] We see a similar but somewhat weaker (correlation 0.27) tendency toward excessive coverage ratios if we instead rank counties by their share of American Indian and Alaska Native residents.

Using CCP data on loan originations to assess HMDA coverage

The previous two sections examine two types of government-backed home loans that together make up less than half the mortgage market (and in some years far less than half). For a more comprehensive assessment of HMDA's coverage of rural home loans, we turn to the credit history data in the CCP. The CCP is based on a nationwide 5 percent sample of almost all of Equifax's credit files.[22] Each file contains data on the consumer's debts, from which we infer when a consumer takes out a new home mortgage.[23] The files also provide an assessment of the census block in which the consumer resides. HMDA data are reported by census tracts, which are amalgams of census blocks. We can thus use the CCP to compile a credit-history-based sample of mortgage originations by census tract and compare the results to the HMDA data for rural census tracts.

Because the CCP and HMDA provide different types of information about each mortgage, we must compare these two sources at a fairly high level of aggregation by loan, property, and borrower characteristics. The CCP distinguishes first liens from second liens but lacks HMDA's more detailed information on borrowers (owner-occupant or not), property type (single-family, manufactured home, or multifamily), and loan purpose (purchase, refinance, or home improvement). HMDA provides only annual data, so we combine four quarters of CCP data into an annual total and then multiply by 20 to allow for the 5 percent sampling rate of the CCP. We divide that CCP-based estimate of total first-lien mortgage originations into the corresponding total reported in HMDA to arrive at an estimated HMDA coverage ratio.

Sampling variation in our CCP-based estimates of an area's total mortgage originations complicates assessment of these HMDA coverage ratios.[24] Consider, for example, a lightly populated area (like a rural census tract or even many rural counties) with 1,000 adults. On average, about 50 of these adults will have credit files in the primary CCP sample we use. Furthermore, the CCP and other sources imply that only about 3 percent of adults originate a mortgage in a given year. So, for this lightly populated area, we might expect the CCP files to include, on average, between 1 and 2 mortgage originations in a typical year. To estimate total mortgage originations in the area, we would then multiple the number reported in the CCP by 20,

[22] Because the sample is drawn, anonymously, based on Social Security Numbers (SSNs), the small fraction of files lacking the consumer's SSN are systematically omitted.

[23] The CCP provides data on mortgages individually, so it is simple to tell when a new mortgage appears in a consumer's credit file. To avoid double counting of joint mortgages, which appear in the credit file of both borrowers in the Equifax records underlying the CCP (and potentially in the CCP too, if both individuals are randomly selected to be in the CCP), we count each file with a new joint mortgage as having half of a new mortgage.

[24] It is important to note that HMDA coverage rates also may be affected by a difference in how loans are spatially coded in HMDA and the CCP (Williams 2015). The CCP locates loans by the census tract of the borrower's billing address, whereas HMDA locates loans by the census tract of the property that secures the loan. For example, a mortgage financing a second home not in the tract of the borrower's primary residence would be coded in the latter tract in the CCP but in the tract of the second home in HMDA. These discrepancies may not be randomly distributed across space, so that, for example, HMDA coverage may appear high in an area with many second homes and correspondingly low in an area where the primary residences of the owners of these second homes are clustered. We make no explicit adjustment for this potential source of error in our measures of HMDA coverage.

given the 1 in 20 sampling rate of the CCP.[25] Note that if, by chance, the CCP includes 1 origination for this area in year 1 but 2 mortgage originations in year 2, our estimate of total originations would be 20 in year 1 and 40 in year 2. However, in any given year, observing 1 versus 2 originations for the area in the CCP could reflect random variation in how a relatively stable number of total originations in the area are distributed among the 5 percent of area credit files in the CCP. Since we use the total origination figures as the benchmark for HMDA's coverage of originations in the area, a random shift from 1 to 2 originations in the CCP in a given year makes a big difference in the HMDA coverage ratio (cutting it in half).

We deal with this issue in two ways. In one approach, we avoid the issue by examining rural areas with populations large enough to mitigate the effects of sampling variation by averaging across a large number of files. For large rural areas, the CCP-based estimates imply that HMDA coverage is near to or above 80 percent of loan originations. In the second approach, we examine smaller spatial areas, focusing mainly on individual counties, including many with small populations. In this approach, we use simulation analysis to explicitly allow for the effects of sampling variation in the CCP. The simulations do not fully explain away the tendency for some rural counties to have very low CCP-based HMDA coverage rate estimates. However, we argue that the simulations account for a large portion of this tendency and conclude that evidence of systematically very low HMDA coverage rates is limited to a minority of rural counties, so that the HMDA data provide useful insights into mortgage markets in most of rural America. This conclusion applies also to rural areas on and near most American Indian reservations.

In the first—or large-populations-only—approach, we compute HMDA coverage ratios for the entire U.S., for all metropolitan areas combined and all nonmetropolitan areas combined, and for groups of counties organized into the 9 categories of the RUCCs. For each of those areas in any given period, we count the number of originations in the CCP and multiply by 20 to obtain a CCP-based estimate of total first-lien mortgage originations. For the same areas and time period, we add up the first-lien mortgage originations reported in HMDA and divide this HMDA total by its corresponding CCP-based total.

The resulting HMDA coverage ratios for the areas described above are shown in Table 2 for 2015. The national and all-metropolitan-county ratios are both about 93 percent, in line with the conventional rule of thumb that HMDA covers about 90 percent of all mortgages. For all nonmetropolitan counties, the ratio falls just a bit below 90 percent, to 88 percent. Looking at the details for nonmetropolitan county groups (RUCCs 4 to 9), we see that the coverage ratio is above 90 percent in nonmetropolitan counties adjacent to metropolitan areas (RUCCs 4, 6, and 8). In nonmetropolitan counties not adjacent to metropolitan areas (RUCCs 5, 7, and 9), coverage is lower, as expected for remote nonmetropolitan areas, but remains near to or a bit above 80 percent. On that basis, it appears that HMDA covers lending activity adequately across rural areas collectively, in the sense that it should reflect the broad patterns and general trends prevailing in rural areas as a whole, including even in the group of remote and decidedly non-urban counties categorized by RUCC 9.

[25] In so doing, we assume that the approximately 10 percent of adults who lack a credit file never get a new mortgage. This assumption may not be perfectly true, but we regard it as approximately correct and maintain it throughout the analysis here.

Can we extend our analysis of HMDA coverage to smaller county groups or even individual counties or census tracts? For individual areas with small populations, the answer is probably "no," due to the sampling variation inherent in the CCP benchmark. This is demonstrated in Figures 5 and 6 below, which are based on actual HMDA coverage ratios for 2015, computed as described above for each U.S. county. At this level of detail, HMDA coverage ratios fall as low as 0.00. Note, however, that these HMDA ratios also reach very high levels, including infinitely high in counties where loan activity is reported in HMDA but no mortgage originations appear in the CCP.[26] Sampling variation in the number of originations reported in the CCP and the need to inflate CCP figures by a factor 20, illustrated above in the example of the county with 1,000 people, presumably explains many of these extremely low or high ratios. If so, how can we assess whether this wide range of actual HMDA coverage ratios is primarily the result of CCP sampling variation and therefore not inconsistent with fundamentally good coverage of rural lending activity in HMDA?

The approach we take is to assume a fairly high level of HMDA coverage everywhere and then simulate the effects of CCP sampling variation on CCP-based estimates of HMDA coverage ratios for individual counties. If these simulations closely replicate the distribution of the actual CCP-based estimates of HMDA coverage ratios shown in Figure 5, we will conclude that the underlying assumption of fairly high HMDA coverage everywhere, including rural areas, is not contradicted by the wide range of estimated coverage ratios in Figure 5.

Our simulations are based on the following maintained assumptions:

1. The American Community Survey provides an exact measure of the number of adults (people aged 18 or more) in each county,[27] so that County(i) has N(i) adults.
2. Uniformly across all counties, the probability that an adult has an Equifax credit file is 0.90.[28]
3. The probability that an Equifax credit file is in the CCP is 0.05.
4. The probability that an Equifax credit file reports a mortgage origination in any given year is 0.029 (the share of CCP files that report an origination in 2015 for the U.S. as a whole).[29] Adults who do not have an Equifax credit file never originate a mortgage.
5. The share of total mortgage originations in a county reported in HMDA equals the value shown in the "County Areas" column of Table 2, based on the county's RUCC value.

With these assumed probabilities, we conduct multiple random simulations, each of which produces an estimated HMDA coverage ratio based on a CCP benchmark for each County(i), as follows:

[26] For convenience, these infinite values are displayed above the horizontal value 5, which is larger than all actual finite coverage ratios.
[27] A potential enhancement to our method would be to take account of the margin of error that the ACS provides for its estimates of adults per county.
[28] A potential enhancement would be to model this ratio as a function of county characteristics, such as the age and income of the population or the county's RUCC.
[29] We also experimented with an RUCC-specific fraction given by the share of CCP files in all counties of a given RUCC that report an origination in year 2015. The results from this variation were similar to those shown here and are not shown.

1. Simulate the number of adults with a credit file: Compute a vector, with one row for each adult in County(i), whose value in any row is either 1 (for "has an Equifax credit file") or 0 (for "does not have an Equifax credit file"), based on N(i) independent Bernoulli[30] experiments each with probability 0.90 (from assumption 2 above).
2. Simulate the number of adults in the CCP: Compute a vector, with one row for each adult in County(i), whose value in any row is either 1 (for "included in the CCP") or 0 (for "not included in the CCP"), by choosing the value to be 0 whenever that adult's value in the previous vector is 0 (i.e., adults with no Equifax file are not in the CCP) and otherwise choosing based on independent Bernouli experiments each with probability 0.05 (from assumption 3 above).
3. Simulate the number of total mortgage originations: Compute a vector, with one row for each adult in County(i), whose value in any row is either 1 (for "originated a mortgage") or 0 (for "did not originate a mortgage"), by choosing the value to be 0 whenever that adult's value in the first vector is 0 (i.e., adults with no Equifax file do not originate mortgages) and otherwise choosing based on independent Bernouli experiments each with probability 0.029 (from assumption 4 above).
4. Simulate the number of mortgage originations in the CCP: Create a vector, with one row for each adult in County(i), whose value in any row is either 1 (for "originated a mortgage in the CCP") or 0 (for "did not originate a mortgage in the CCP"), by choosing the value to be 1 whenever that adult's value in both of the previous two vector is 1 (i.e., the adult is in the CCP and the adult originated a mortgage) or 0 otherwise (i.e, either the adult is not in the CCP, or the adult did not originate a mortgage, or both).
5. Calculate a CCP-based estimate of total mortgage originations. Sum the row values of the previous vector to find the total number of mortgage originations in the CCP for County(i) and multiply by 20, given that the CCP includes on average one out of every 20 files.
6. Calculate the total number of mortgages reported in HMDA. Sum the row values of the vector in step 3 to find the simulated "true" number of mortgage originations in County(i) and then multiply this total by the fraction shown in the "County Areas" column of Table 2 for County(i)'s RUCC value.
7. Compute the simulated CCP-based HMDA coverage ratio. For each County(i), divide the result from step 6 by the result from step 5.

Each simulation involves hundreds of millions of independent Bernoulli experiments at multiple stages. However, we can fairly readily execute 100 simulations.[31] Figure 7 provides one illustration of the results. It is a bar graph that shows two distributions of county-level HMDA coverage ratios. The red distribution reproduces the distribution of actual 2015 county HMDA coverage ratios already shown in Figure 5. The blue distribution is based on the median value within each bin across the 100 simulated distributions, and purple shows where the two distributions overlap. The simulation-based distribution is not as dispersed as the actual distribution, with smaller bars at "infinity" (shown as 5) and on the far left side of the distribution and a higher bar in the middle, near 0.9. A formal statistical test strongly rejects the

[30] A Bernoulli experiment with probability p can be thought of as picking either 1 or 0 based on the random spin of a dial whose probability of pointing to 1 is p and to 0 is (1-p).
[31] We experimented with increasing the number of simulations to 250 but found no material change in the results.

hypothesis that the two distributions are the same.[32] That is, the simulations fail to closely replicate the actual distribution of 2015 HMDA coverage ratios for counties, thereby rejecting the hypothesis underlying our simulation—that HMDA's coverage ratios are relatively high everywhere.

Notwithstanding this formal rejection, Figure 7 visually suggests that the simulations capture much of the wide dispersion of HMDA coverage rates. The main failing of the simulations seems to be an inability to account for the far left of the distribution, with hundreds too few simulated county values below about 0.55 and a somewhat more than offsetting excess of simulated values between about 0.85 and 1.05. However, the two distributions appear to match relatively closely between 0.55 and 0.85 and are not too far apart above 1.35. Despite some remaining differences in the right sides of the two distributions, which may in part reflect systematic effects related to second homes,[33] the remainder of this paper will focus on discrepancies on the left sides of the two distributions.

To assess whether rural counties drive the discrepancy between the actual and simulated distributions in the left side of the distribution, we reorganize the simulation output on a county-by-county basis. We first identify the 724 counties whose actual HMDA coverage ratio for 2015 falls in, or is lower than, the lowest 10 percent of the simulated values for the same county. The simulations imply that the actual values for these counties are unlikely to be accounted for by the sampling variation in the CCP. Nonetheless, in some of these counties, the actual coverage ratio is not very low in an absolute sense. In particular, 292 of these 724 counties have actual coverage ratios equal to or greater than 0.67, leaving 432 counties whose actual coverage ratios are both absolutely low (below 0.67) and probably not accounted for by CCP sampling variation. This represents about 22 percent of the 1,985 counties with a nonmetropolitan RUCC code (codes 4 to 9).

Figure 8 maps these 432 counties, with the more intense orange and red colors showing counties with the lowest actual coverage values. Apart from Alaska, almost all of the counties shown are east of the Rocky Mountains, scattering mostly across the Great Plains, the Corn Belt, the Cotton Belt, western Appalachia, and counties near the central and eastern Canadian border. Many of these counties are isolated, with no similar neighboring counties, which may imply a very local gap in HMDA coverage perhaps linked to a few small local lenders taking a large share of the county's mortgage market. However, there are also clusters of these counties that suggest wider regional gaps in HMDA coverage, such as in the Iowa-Missouri-Illinois border area, the Iowa-Minnesota-South Dakota border area, eastern Kentucky, the Upper Peninsula of Michigan, central Georgia, and elsewhere.

Figure 8 also displays the boundaries of federally recognized American Indian reservations. We find no tendency for the highlighted counties to disproportionately overlap reservations. If anything, the dearth of highlighted counties in Pacific Coast and Intermountain regions leaves reservations somewhat lightly affected. But exceptions are easily found, too, such as the Crow, Fort Peck, Cheyenne River, Crow Creek, and Red Lake reservations.

[32] A two-side Kolmogorov-Smirnov test rejects the hypothesis that the two distributions are the same (p-value 0.32).
[33] See the previous footnote on spatial coding issues in the CCP.

Overall, we regard our CCP-based comparisons as also generally supporting the relevance of HMDA data in rural areas. However, we grant that caution is warranted, based on inability to account for the significant shortfall in HDMA's CCP-based coverage ratio in a nontrivial minority of nonmetropolitan counties across the country, on some American Indian reservations, and in some broader clusters of counties.

Concluding Remarks

The fact that many small and nonmetropolitan lenders are exempt from reporting their home loans under HMDA raises the possibility that the HMDA data are not reliable for monitoring the patterns and trends in rural home lending. We use federal agency and credit bureau data to generate new evidence that HMDA provides adequate-to-good coverage of home lending in rural areas taken as a whole and also in the majority of rural counties. That is, we find little evidence of a widespread or pervasive pattern of underreporting of rural loans in HMDA. On that basis, we conclude that HMDA is a useful and important source of information about home lending in rural America as a whole and also in most of its individual counties and American Indian reservations. However, we also present evidence that HMDA's coverage of home lending is below 60 percent—sometimes by a large margin—in a minority of rural counties and Indian reservations, so that caution is appropriate when using HMDA data to monitor home lending patterns in smaller rural subsections of the U.S.

References

Avery, Robert B., Kenneth P. Brevoort and Glenn B. Canner. 2007. "Opportunities and issues in using HMDA data." *The Journal of Real Estate Research*, 29(4): 352–379.

Bhutta, Neil, Steven Laufer, and Daniel R. Ringo. 2017. "Residential Mortgage Lending in 2016: Evidence from the Home Mortgage Disclosure Act Data." Board of Governors of the Federal Reserve System, *Federal Reserve Bulletin*, 103:6 (November).

Housing Assistance Council. 2010. "Improving HMDA: A Need to Better Understand Rural Mortgage Markets." Research Note.

Kunesh, Patrice H. 2017. "Creating Sustainable Homelands through Homeownership on Trust Lands." *Rural Voices*, 21:3 (Fall).

Scheessele, Randall M. 1998. "HMDA Coverage of the Mortgage Market." U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Housing Finance Working Paper Series (July).

Williams, Michael. 2015. "HMDA Coverage of the Mortgage Market." Retrieved from the University of Minnesota Digital Conservancy, http://hdl.handle.net/11299/172689.

Table 1: Rural Counties with a HMDA Coverage Ratio for FHA-Insured Home-Purchase Loans Less than 60 Percent, an American Indian/Alaska Native Population Percentage of at Least 2 Percent, and Nearby Tribal Communities

State	County	Population (2010)	HMDA Coverage Ratio	Comments
NV	Pershing	6,753	24%	Near Lovelock and Pyramid Lake Reservations, among others
MN	Clearwater	8,695	33%	Overlaps White Earth and Red Lake reservations
NV	Eureka	1,987	33%	Near South Fork and Battle Mountain reservations
WA	Ferry	7,551	40%	Overlaps the Colville Reservation
NC	Columbus	58,098	46%	Traditional area of the Waccamaw Siouan Indian Community
NC	Warren	20,972	47%	Traditional area of the Haliwa-Saponi Community
WA	Klickitat	20,318	47%	Overlaps the Yakima Reservation
NV	Humboldt	16,528	48%	Overlaps or near the Fort McDermitt, Summit Lake, Battle Mountain, and Duck Valley reservations and Woodfords Indian Community
SD	Day	5,710	50%	Overlaps the Lake Traverse Reservation
AK	Skagway	968	50%	Home of the Skagway Traditional Council and a traditional Tlingit area
ND	Mercer	8,424	51%	Overlaps the Fort Berthold Reservation
ND	Bottineau	6,429	54%	Adjacent to Rolette County, which includes the Turtle Mountain Reservation
NV	Lander	5,775	55%	Overlaps or near the Yomba, Battle Mountain, and other reservations
CA	Plumas	20,007	55%	Overlaps the Greenville Rancheria
OR	Harney	7,422	58%	Overlaps the Burns Paiute Indian Reservation
OK	Dewey	4,810	59%	Cheyenne-Arapaho Tribal Statistical Area
NM	Taos	32,937	59%	Overlaps or near the Taos, Picurus, and other Pueblo communities, as well as the city of Taos
LA	Sabine Parish	24,233	60%	Traditional area of the Choctaw Tribe of Ebarb (state recognized)

Note: Listed by HMDA coverage ratio, from smallest to largest. The ratio was calculated by the authors and is based on the number of loans.

Table 2: HMDA Coverage Ratios for Large Rural and Urban Areas Group by County RUCCs, 2015

County Code	Description of County Code	Number of Counties	2010 County Population	HMDA Coverage Ratios			
				Counties by RUCC	All Metro Counties	All NonMetro Counties	U.S.
RUCC 1	Counties in metro areas of 1 million population or more	432	168,523,961	.9133	.9331		.9278
RUCC 2	Counties in metro areas of 250,000 to 1 million population	379	65,609,956	.9487	.9331		.9278
RUCC 3	Counties in metro areas of fewer than 250,000 population	356	28,318,215	1.0034	.9331		.9278
RUCC 4	Urban population of 20,000 or more, adjacent to a metro area	214	13,538,322	.9235		.8826	.9278
RUCC 5	Urban population of 20,000 or more, not adjacent to a metro area	92	4,953,810	.8135		.8826	.9278
RUCC 6	Urban population of 2,500 to 19,999, adjacent to a metro area	593	14,784,976	.9040		.8826	.9278
RUCC 7	Urban population of 2,500 to 19,999, not adjacent to a metro area	433	8,248,674	.8271		.8826	.9278
RUCC 8	Completely rural or less than 2,500 urban population, adjacent to a metro area	220	2,157,448	.9504		.8826	.9278
RUCC 9	Completely rural or less than 2,500 urban population, not adjacent to a metro area	424	2,610,176	.7900		.8826	.9278

Figure 1: HMDA Coverage Ratios for RHS Loan Numbers, by State, 2009–2016

(Red lines mark the range from 80 to 100 percent.)

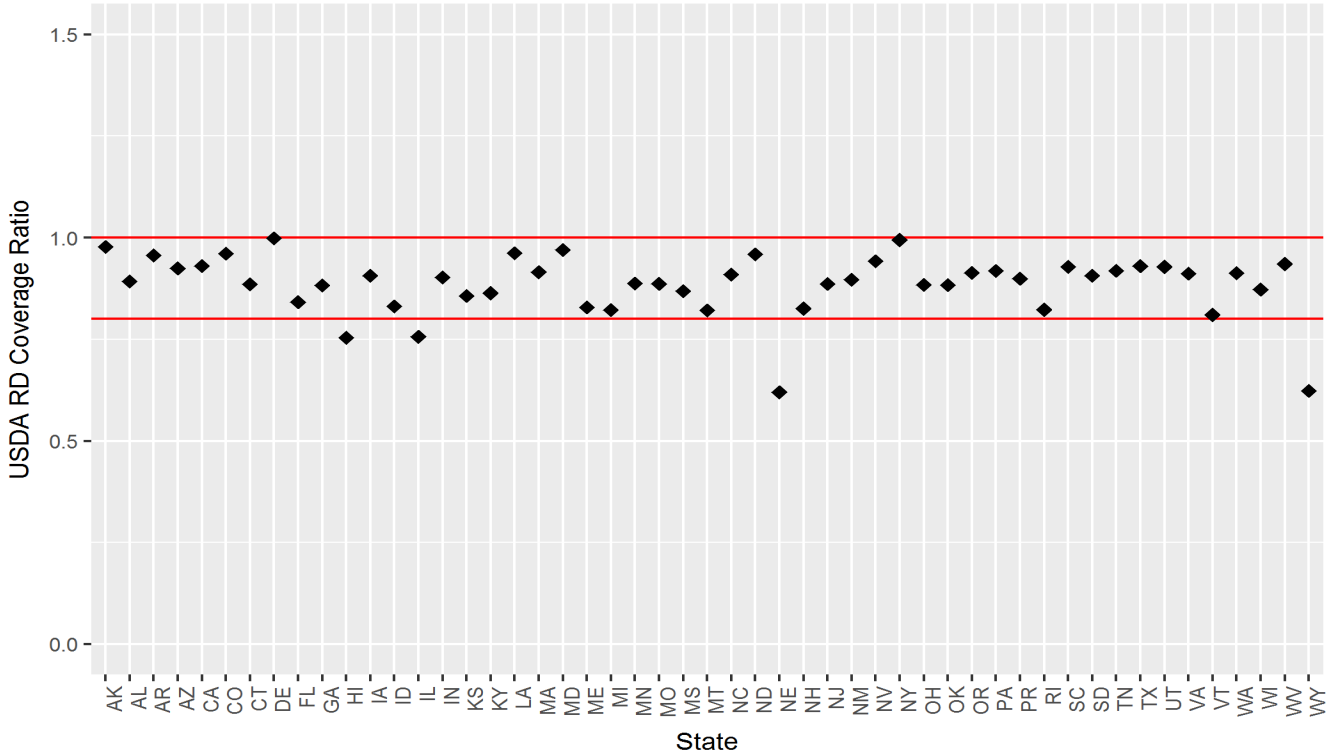


Figure 2: HMDA Coverage Ratios for FHA-Insured Loan Numbers, by Rural Areas of States, 2013–2014

(Red lines mark the range from 80 to 100 percent.)

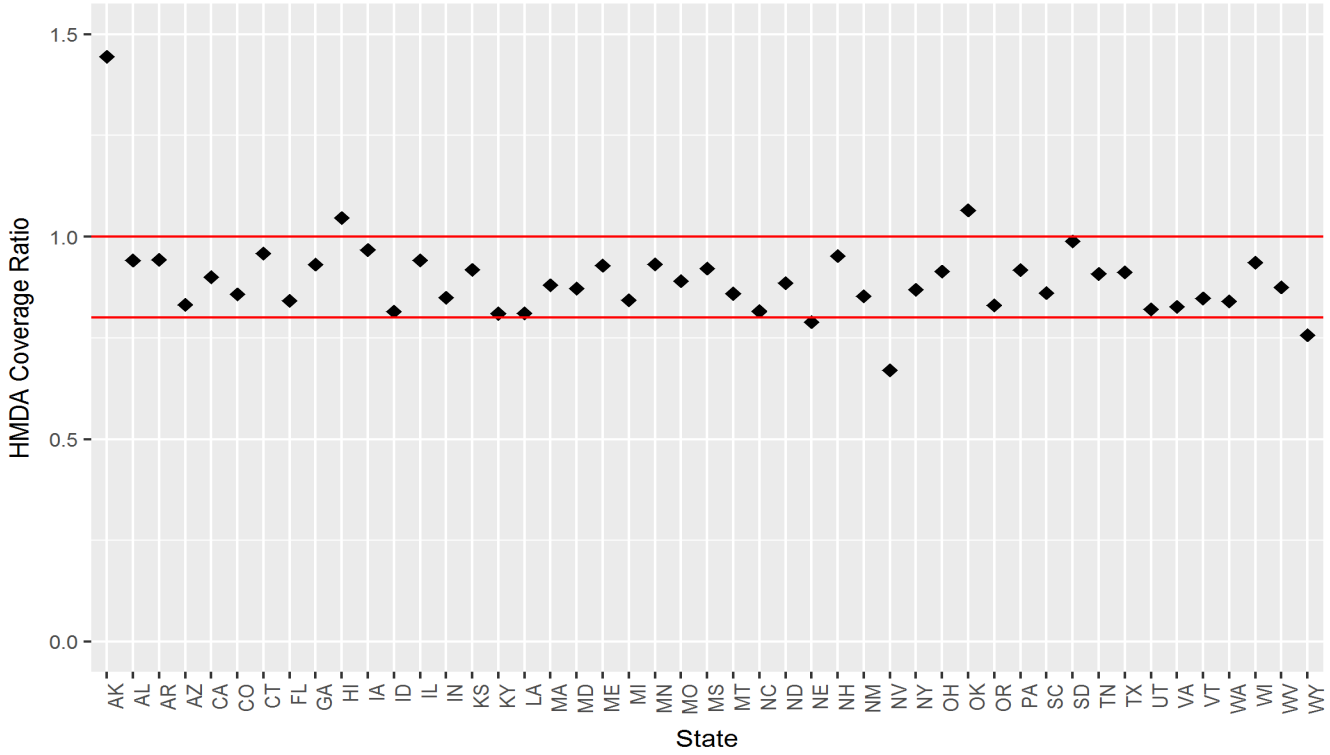


Figure 3: HMDA Coverage Ratios for FHA-Insured Loan Numbers, by Rural Counties Group by RUCCs, 2013–2014

(Most of the 28 counties with ratios greater than 1.5 are omitted, to focus on cases of low coverage)

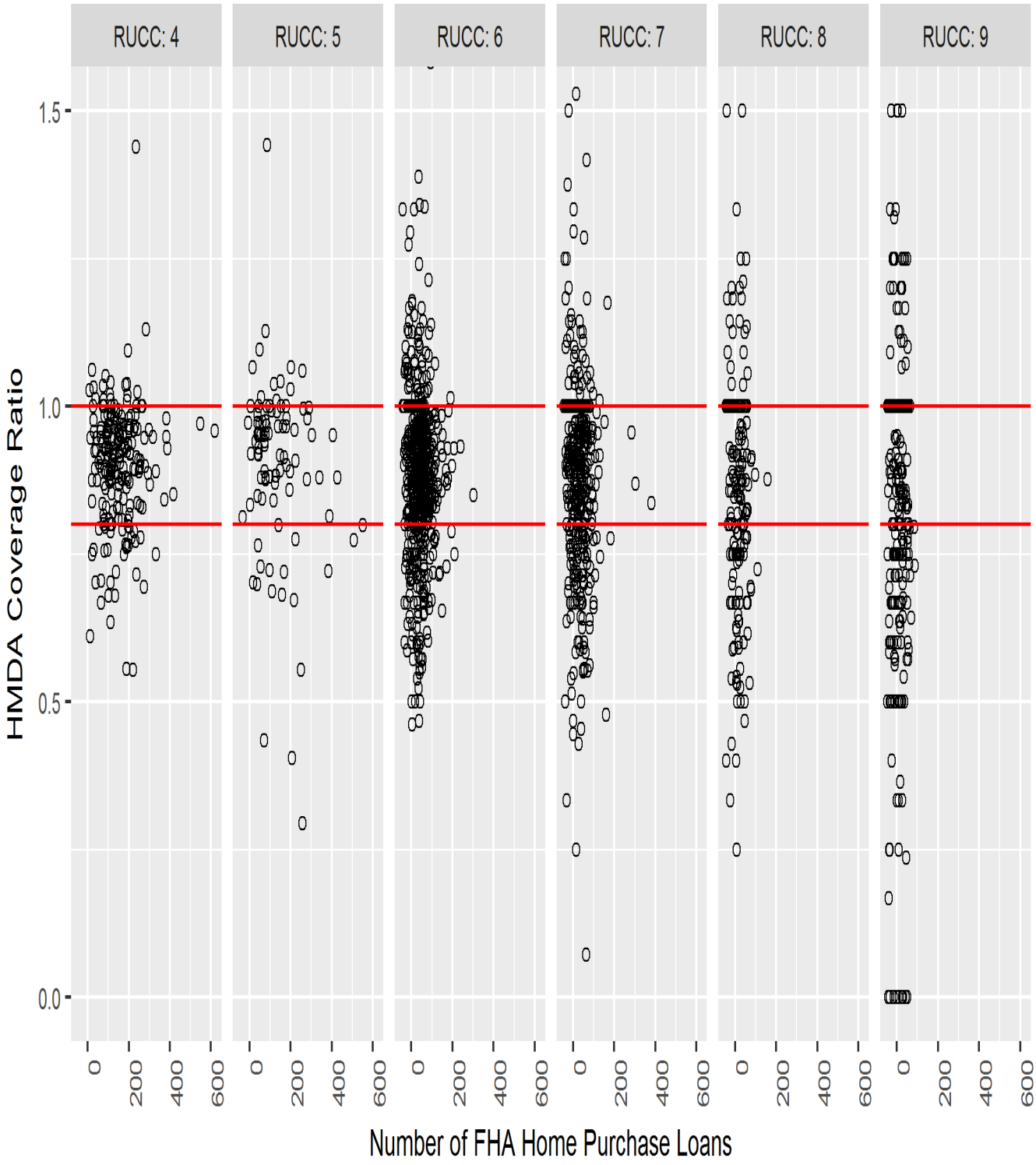


Figure 4: High Values of HMDA Coverage Ratios for FHA-Insured Loans Are Often Associated with a High Share of These Loans Reported as for American Indian and Alaska Native Borrowers
(Computed for Nonmetropolitan Counties in 2013–2014 with more than 10 FHA– insured loans recorded by HUD)

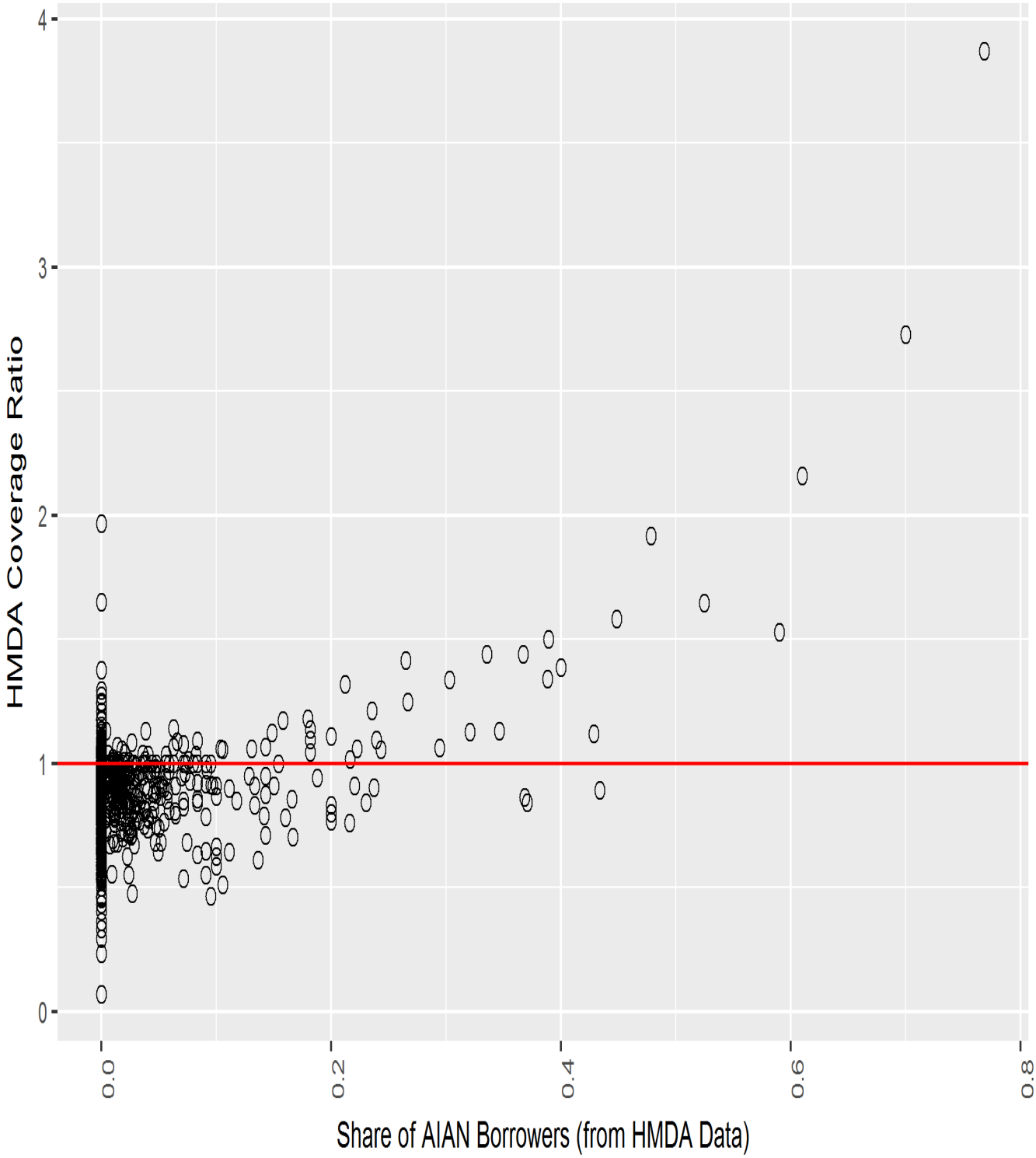


Figure 5: Range of Actual CCP-Based HMDA Coverage Ratios for U.S. Counties, 2015

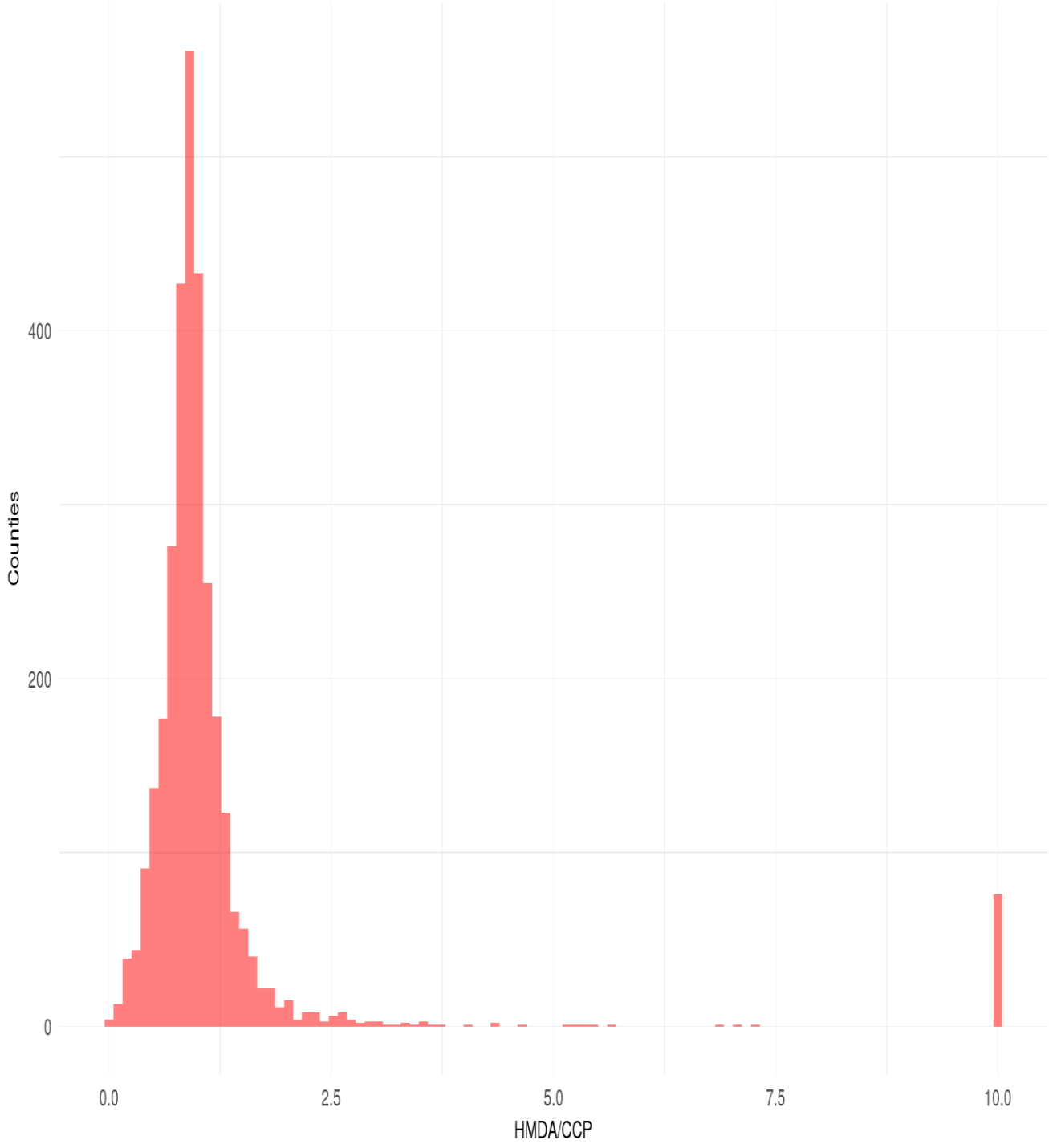


Figure 6: Map of Actual CCP-Based HMDA Coverage Ratios for U.S. Counties, 2015

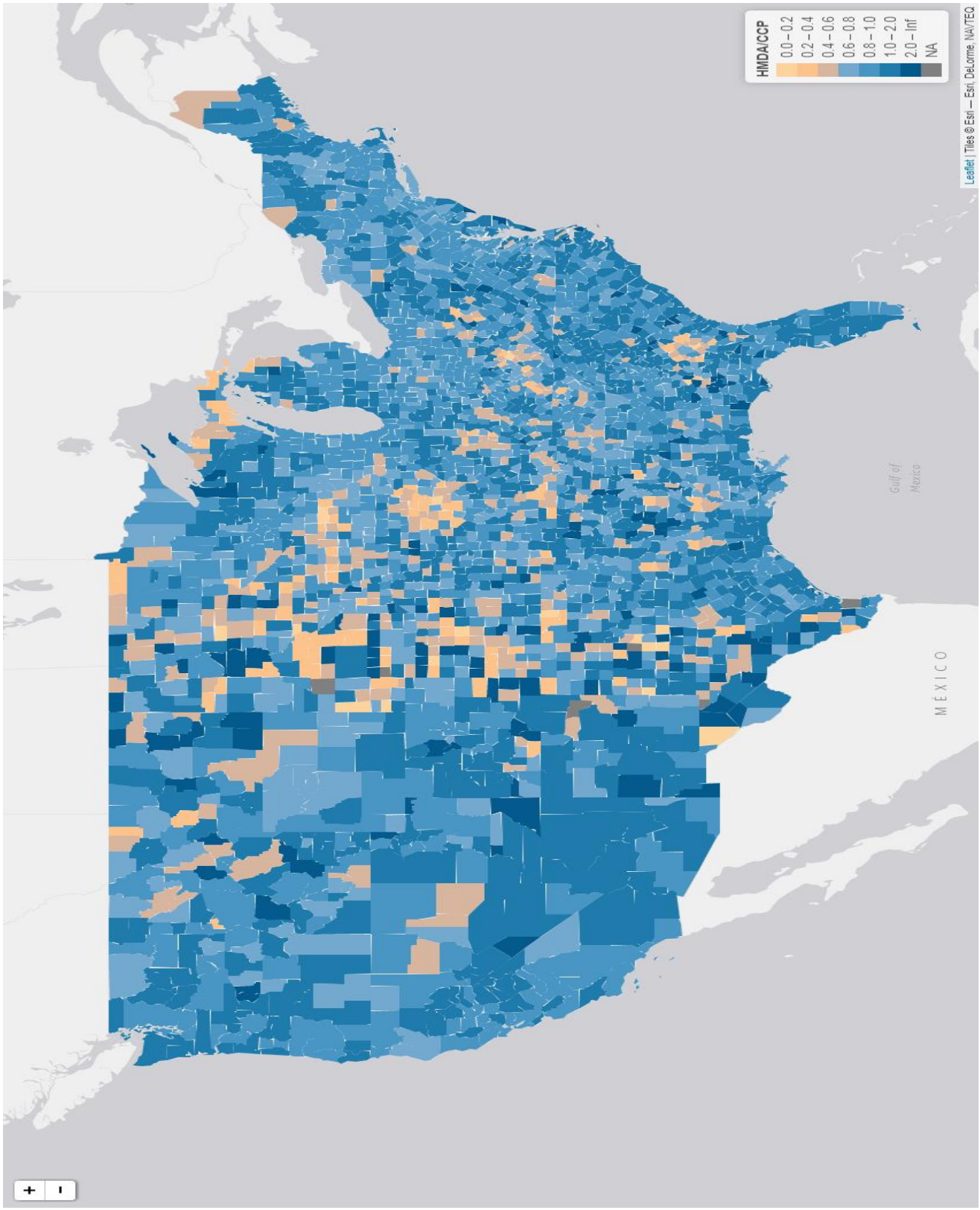


Figure 7: Distributions of Actual 2015 vs. Simulated CCP-Based HMDA Coverage Ratios
(Actual in red; median of simulated values per bin in blue; overlap in purple)

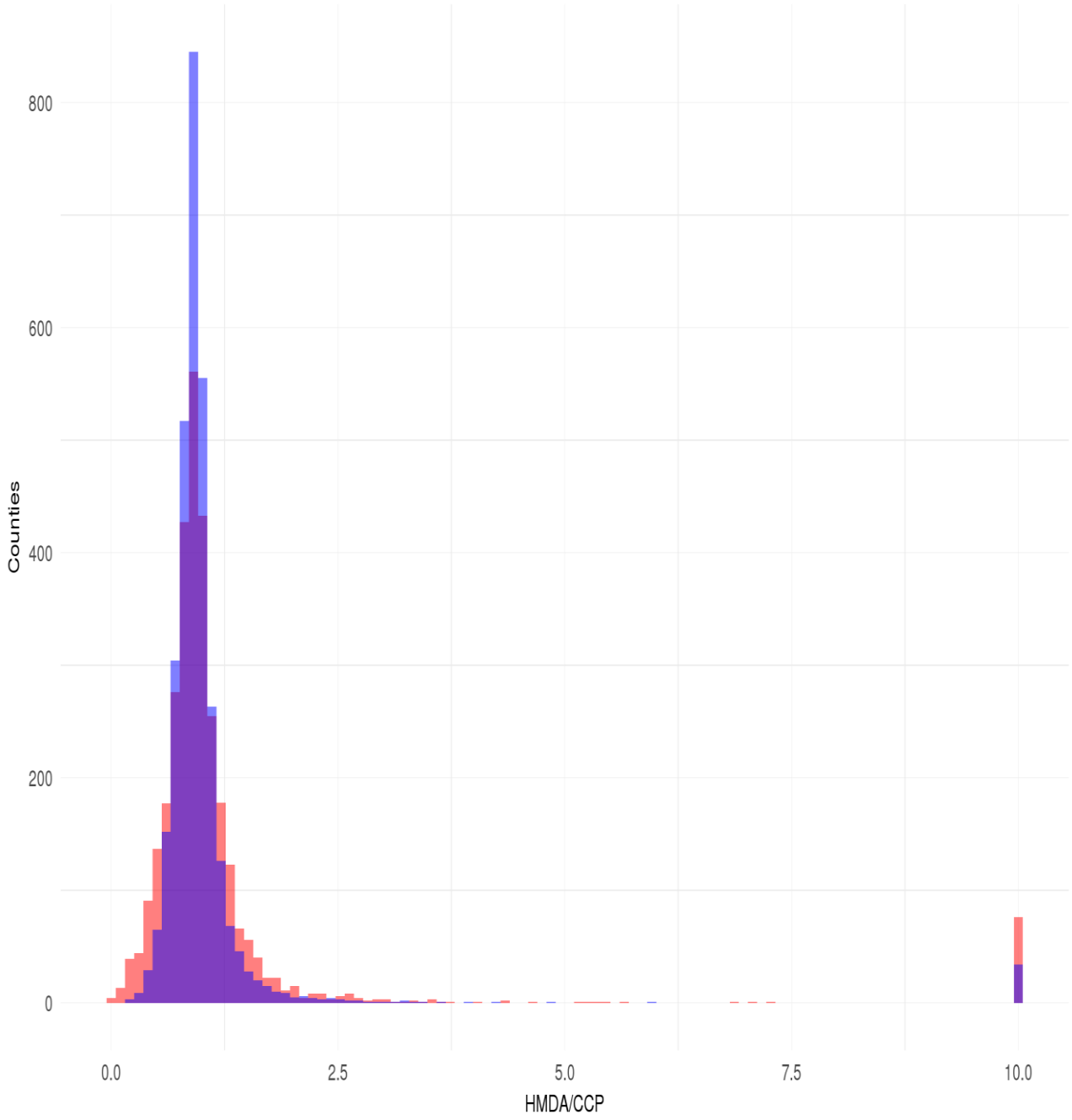
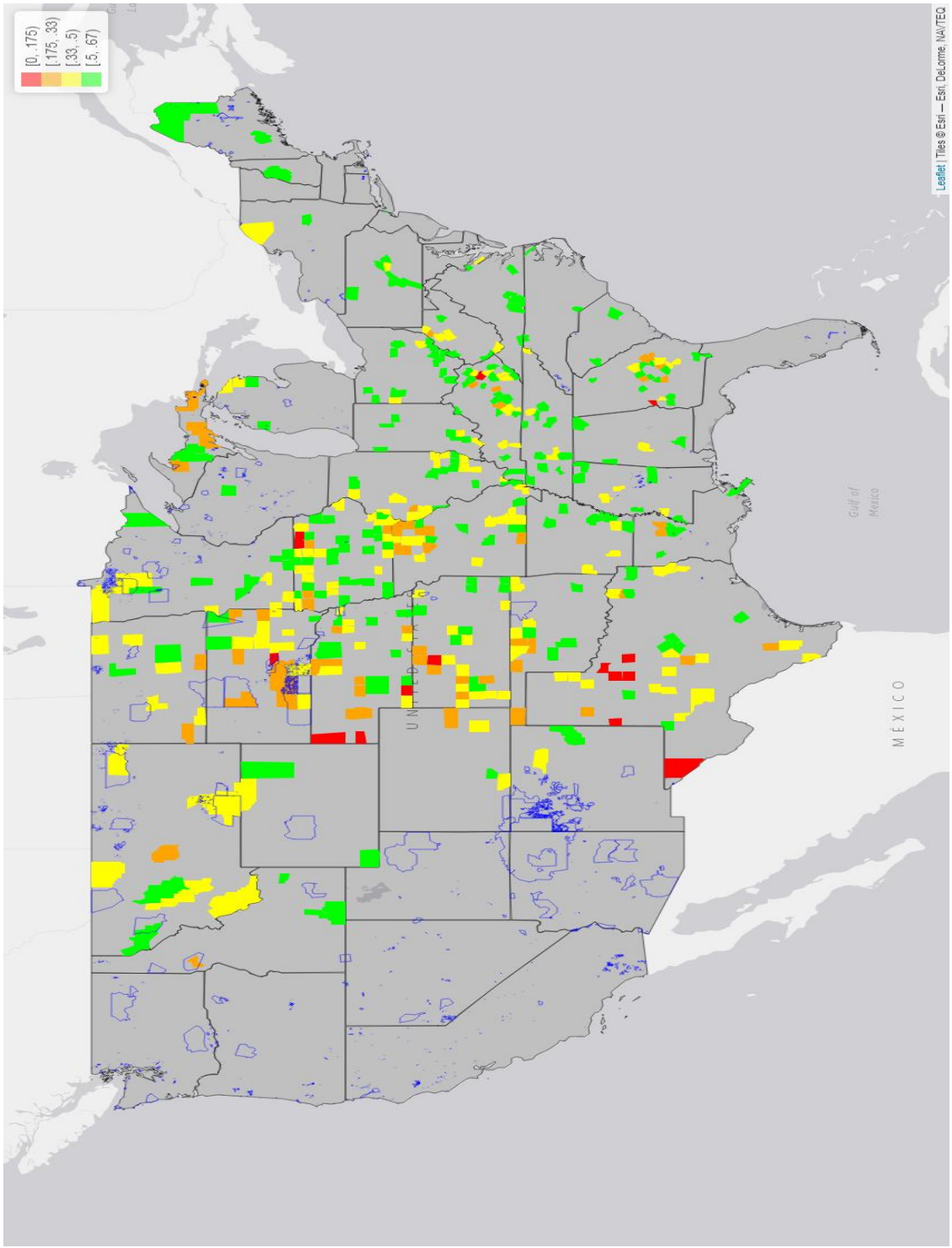
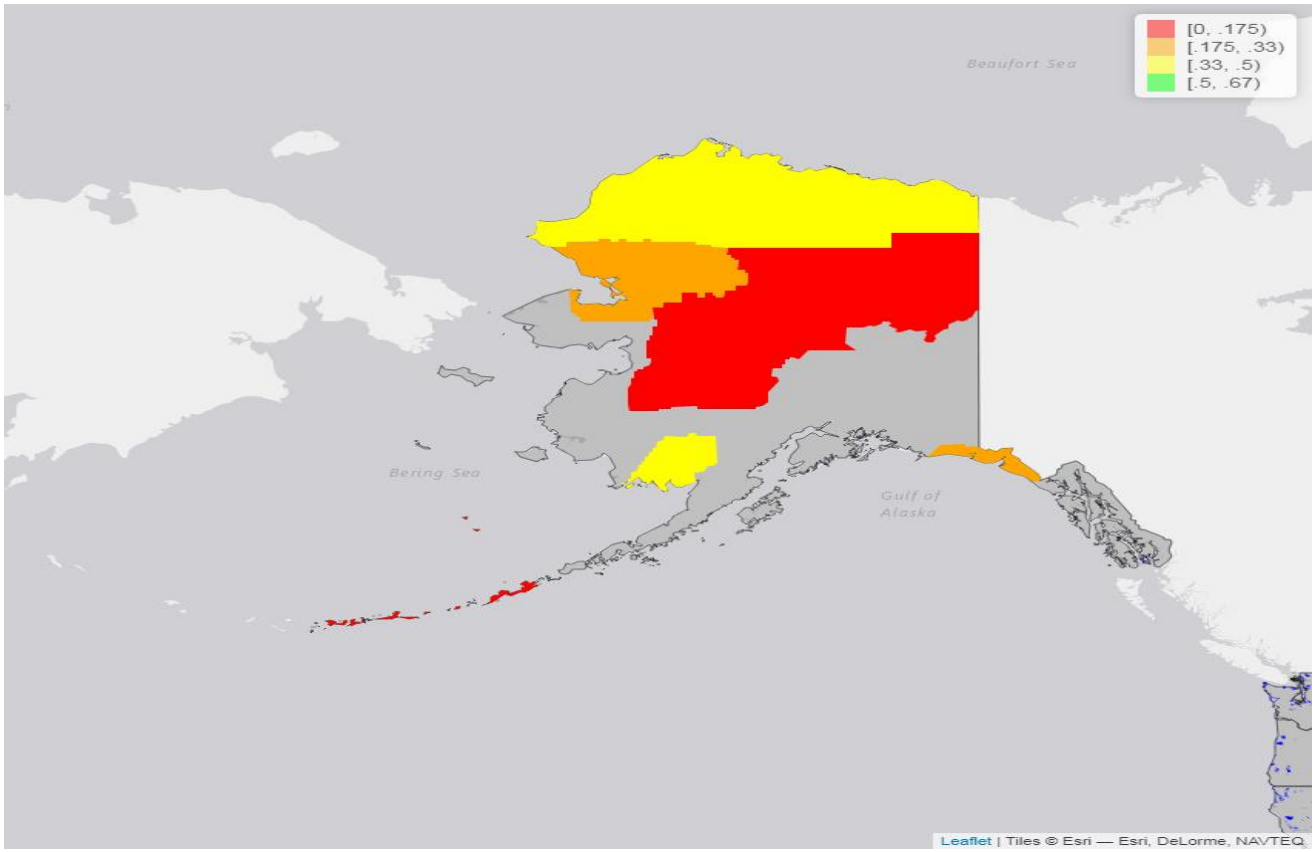


Figure 8: Map of Counties with Actual 2015 HMDA Coverage Ratios That Are Both Unlikely (Per Simulations) and Low

A: 48 Contiguous States (Boundaries of federally recognized American Indian reservations in blue)



B: Alaska (Boundaries of federally recognized American Indian reservations in blue)