

Proxies

**Recommendation**

The Bureau should publicly announce the empirical methodology we use for assigning sex, race, and ethnicity in CFPB fair lending supervisory activities. One of the aims of this announcement should be to emphasize that proxying is a commonly accepted statistical technique.

**Background on Proxies**

The ECOA forbids creditors from inquiring about an applicant’s demographic information, with very limited exceptions.<sup>1</sup> Therefore, outside of mortgage lending (for which sex, race, and ethnicity data ~~is~~ are collected and reported pursuant to the Home Mortgage Disclosure Act), fair lending analyses must rely on proxies to assign race, ethnicity, or sex. Such proxies have been in use for many years, and they are an accepted means of obtaining otherwise unavailable information in discrimination matters.<sup>2</sup>

The only widely accepted method of proxying for sex relies on a name database from the Social Security Administration, which reports counts of individuals by sex and birth year for first names occurring at least five times in a birth year.<sup>3</sup> The proxy method assigns a probability that a particular applicant is female based on the documented distribution of the total U.S. population across sex categories (male or female) for a given first name.

Most prior analyses have assigned sex using a threshold, such that any person with a likelihood of being female over, say, 80% is assumed to be female. The Bureau’s approach instead weights each loan with the exact probability that the loan is male or female, so if the name on the loan has a 75% chance of being male, then that loan is assigned a male probability of 75%. In retaining the raw probabilities, rather than using a threshold, this method makes fewer assumptions and therefore more accurate predictions. However, because most names are highly sex-specific, the practical difference between the two methods is small.

For race and ethnicity, the most commonly used methods of proxying use the borrower’s surname, the borrower’s residence (geocoding), or both. In some cases, race and ethnicity have been assigned using available DMV information, including photographs.

**Comment [BES1]:** This section discusses the construction of proxies and touches on their use. We might want to be clear on what we intend to publish, which I thought was construction (and not use).

**Comment [BES2]:** Just as a general comment: folks regularly use “data,” which is the plural form of “datum,” to mean data set.

**Comment [BES3]:** “Exact” may mean more than expected. Maybe “value of the probability” or something.

**Comment [BES4]:** I suppose that by construction, the use of the probability directly is one less step. The threshold takes this probability and overlays the threshold rule to determine full or no assignment. We could highlight some of the potential concerns that this alleviates, which are discussed in “Power of Tests for a Dichotomous Independent Variable Measured with Error” (McCaffrey and Elliott 2008).

**Comment [BES5]:** Of what? True sex? Total population affected?

<sup>1</sup> 12 CFR § 1002.5(a), (b).

<sup>2</sup> See, e.g., *Interagency Fair Lending Examination Procedures*, at 12-13, available at <http://www.ffiec.gov/PDF/fairlend.pdf> (suggesting “the potential use of surrogates” in a comparative file review “in instances where no direct evidence of that characteristic is available” and providing examples of surname proxies for race/ethnicity and first name proxies for sex); *CFPB Supervision and Examination Manual*, at Procedures 19, available at [http://files.consumerfinance.gov/f/201210\\_cfpb\\_supervision-and-examination-manual-v2.pdf](http://files.consumerfinance.gov/f/201210_cfpb_supervision-and-examination-manual-v2.pdf); *Benavidez v. City of Irving*, 638 F. Supp. 2d 709, 717 (N.D. Tex. 2009) (“The Spanish surname may be used as a proxy for Hispanic ethnicity when self-identification is not practical.”); *U.S. v. Reyes*, 934 F. Supp. 553, 560-62 (S.D.N.Y. 1996) (accepting a proxy that assigns race based on geocoding, and noting that an expert calls this method “commonly used”). But see, e.g., *Rodriguez v. Bexar Cnty.*, 385 F.3d 853, 870, n.18 (5th Cir. 2004) (“The use of ‘Spanish-surname’ registration is novel and highly problematic.”).

<sup>3</sup> <http://www.ssa.gov/oact/babynames/limits.html>.

DRAFT – SENSITIVE – PRE-DECISIONAL – ATTORNEY WORK PRODUCT

Surname methodologies are most commonly used to proxy race/ethnicity for Hispanics and Asians, based on the full count of Hispanics and Asians using those surnames in the decennial census. Surname analysis alone does not tend to be as effective a proxy for African-Americans, because surnames for that population are not as readily distinguished from non-Hispanic White surnames.

One example of surname proxying was the fair lending analysis conducted by the OCC at [REDACTED], which identified disparities in dealer markup solely for Hispanic borrowers.<sup>4</sup> Other examples of surname proxying are the recent fair lending settlements by the DOJ in which discriminatory disparities were found in unsecured lending focusing exclusively on Hispanic borrowers.<sup>5</sup>

Comment [EBW6]: Check with Marta

A second type of proxy used in fair lending analysis uses the demographics of the census tract in which a borrower's residence is located, and assigns a race or ethnicity to the borrower based on the predominant demographics in that census tract. If, for example, the population of a census tract is 80% or more African-American, then loans in that tract will be assigned as African-American. The same assignment process would apply to census tracts whose population is 80% or more of other racial or ethnic groups, and then the lending outcomes of borrowers of different groups would be compared.

One example of geographical proxying was [HOW WAS THE FDIC'S PROXYING CONDUCTED? LET'S ADD IT IN WHERE APPROPRIATE]. [The geocoding methodology has also been used in impartial jury cases to determine the racial composition of the jury pool.<sup>6</sup>

Comment [EBW7]: Need an example here. Check with Marta/OR/FDIC

Comment [BES8]: FDIC did analysis on Ally. We have the code. It isn't clear what the data inputs included. I suspect threshold based. It is my understanding that the FDIC has a standard way of doing this. We should reach out to Karen.

A third method of assigning race or ethnicity is not really a proxy at all; it uses information gathered from state DMV records, either through actual data fields collected by the state or by a visual assessment based on the borrower's driver's license photograph. One issue with this method of assigning race or ethnicity is that DMV data is nonpublic and lenders would therefore be unable to adopt this method in their internal analyses.

Comment [BES9]: A minor note: there would still be the need to match records. Also, match rates may not be 100% (in a non-random way). For instance, if DMV data are available for only a subset of states, then we could have loan level observations with no matches.

One example of using DMV data to proxy for race is a DOJ case from the 1990s...[GIVE THE DOJ CASES FROM THE 1990S AS AN EXAMPLE]

Comment [EBW10]: Check with Marta

<sup>4</sup> See the August 26, 2011, letter from [REDACTED] counsel) to Patrice Alexander Ficklin, describing "significant enhancements to its fair lending modeling," including "updated proxies based on the Census 2000 list of surnames to identify Hispanic borrowers."

<sup>5</sup> In June 2011, DOJ reached a settlement with [REDACTED] to resolve allegations that the bank had violated the Equal Credit Opportunity Act ("ECOA") by charging higher prices on unsecured consumer loans made to Hispanic borrowers, which required [REDACTED] to pay approximately \$100,000 in restitution. [REDACTED] did not maintain written loan pricing guidelines for its unsecured consumer loans; instead, the bank's loan officers were granted broad discretion in handling all aspects of the unsecured consumer loan transaction. DOJ alleged that this policy had a disparate impact on Hispanic borrowers. A similar settlement was reached with [REDACTED] in February 2013 for \$700,000.

<sup>6</sup> See, e.g., *United States v. Reyes*, 934 F. Supp. 553, 560-62 (S.D.N.Y. 1996) (citing an expert saying that "geocoding is 'commonly used'" and deciding that "[o]nly the geocoded data from the Jury Wheel study will be considered").

Over the last decade, another method of proxying race and ethnicity has been developed that integrates the surname and geographical approaches. This method was developed by researchers at RAND,<sup>7</sup> and it combines the respective probabilities generated by the surname and geographical proxies. Published research on this proxy methodology demonstrates that it consistently outperforms other proxies when verified against actual reported race and ethnicity data.<sup>8</sup> In addition, our Office of Research has conducted internal analyses comparing this proxy's performance to other proxy methods and has verified its superior performance using HMDA data containing self-reported race and ethnicity, as well as surname and address. For these reasons, the Bureau has been using the integrated proxy in all of our nonmortgage analyses.

**Comment [BES11]:** We should check that this is the conclusion of the article. I think they motivate this approach based on comparisons of overall performance.

**Comment [BES12]:** This is rather strong...but if it reflects your opinion...then keep it. We have not published for internal consumption.

The Office of Research continues to refine this method and test it against the relevant alternatives to ensure that it continues to outperform other methods. For example, as with sex, our method refines the race/ethnicity proxy, by weighting each loan in proportion to the exact probabilities created by the proxy, rather than using a threshold (e.g., 80%) for assignment. So, for example, if the proxy predicts that a borrower is 85% Hispanic, 10% African American, and 5% non-Hispanic White, then a loan to that borrower will be assigned as 85/10/5, rather than simply being assigned as Hispanic.<sup>9</sup>

**Comment [BES13]:** This is more akin to what Siskin is doing. We can help with language here.

#### Costs and Benefits of Publishing Our Methodology

Several institutions and industry groups have asked the Bureau to publish the methodology it uses for proxying in its auto lending analyses. Their stated rationale is that a published proxy methodology will enable them to conduct self-analyses on their own portfolios, allowing them to identify and address issues before a CFPB exam or investigation commences.

**Comment [BES14]:** Is there a history of regulators publishing their approaches to fair lending analysis (outside of a research context)? Would it be novel to do so? This would be useful information.

#### *Benefits*

The primary benefit of publishing our methodology is that it fulfills our desire to be transparent with financial institutions. Transparency fulfills many goals. It will relieve some uncertainty about how we assign race and ethnicity. It encourages dialogue about and further enhancements of our methodology. It may also encourage institutions to conduct self-analyses. As some lenders roll out dealer monitoring programs, it is possible that even some auto dealers may use our proxy method to conduct analyses of their own portfolios. Thus, publishing our methodology could reduce discrimination through voluntary industry action. Moreover, if our method improves upon existing methods, then publishing it will also encourage adoption by institutions that currently proxy using other methods, thereby improving the overall quality of industry fair lending self-assessments.

<sup>7</sup> Elliott et. al., "Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities," *Health Services and Outcomes Research Methodology*, Sept. 2009.

<sup>8</sup> *Id.*

<sup>9</sup> [REDACTED]'s PARR response accepts not only the Bayesian Improved Surname Geocoding methodology, but it also uses proportional weights rather than a threshold. See [REDACTED] PARR Response (Jan. 17, 2013), at 6. Ally's PARR response argues that proxying for race "introduces error" and suggests that our methodology does not "acknowledge this potential for error." See Ally PARR Response (Feb. 11, 2013), at 18-19.

DRAFT – SENSITIVE – PRE-DECISIONAL – ATTORNEY WORK PRODUCT

These benefits may be mitigated by certain factors. We have already stated publicly that our proxy methodology uses both surname and geography, and we have even pointed some entities to the RAND study in particular. Thus, publishing even greater detail on our methodology may only marginally increase transparency or the likelihood that lenders or dealers use it to self-analyze. Moreover, as described above, various proxy methodologies have existed for decades. In that time, some institutions have failed to adopt those publicly available methods for their nonmortgage compliance programs, so those entities may not use our method now.

**Comment [BES15]:** Again, I don't know that this is technically a RAND study. It was published in a journal.

*Risks*

There are three primary risks to publishing our proxy methodology. First, a public statement risks locking us into one particular methodology, notwithstanding that proxying remains an evolving area of economics and we expect that conversations with other federal agencies and experts may further inform our thinking. In other words, our announcement could be used in the future to show that we did not use our publicly espoused methodology. We may be able to manage this risk with careful drafting by emphasizing that many methods are available, that we are not espousing any one method, and that we often use different methods to evaluate the robustness of our results.

A second risk is that publishing our proxy methodology risks conflicts with our sister regulatory institutions, which may use different proxy methods. Publishing one method implicitly suggests that other methods are inferior, and some fair lending enforcement agencies have been using these other proxy methods for decades. Conflicts of this sort could be especially problematic when we are engaged in joint investigations with other agencies, as is the current case with auto finance.

**Comment [BES16]:** That are likely not published or otherwise publically communicated.

The third risk is that publicizing our methodology opens it up to attack. The more detail we provide about our proxy method, the more susceptible it is to criticism (reasoned and unreasoned alike). However, because the Bureau is dedicated to data-driven decision-making, and because of the important improvements we have made, we should be proud of our methodology and willing to both share it and defend it. If presented in the right way, we can use these values to allow further evolution of our method based on feedback.

## Tolerance

The Bureau and indirect auto lenders have a common interest in ensuring that our fair lending compliance efforts are focused on the areas of highest risk. Both have recognized that small disparities, while statistically significant at the 95% confidence level, may be deemed “materially insignificant,” and therefore within an acceptable tolerance.<sup>10</sup> The concept of material insignificance recognizes at least two factors. First, statistical models may never be perfect, and small disparities may be reflective of noise in the data (or failure to account for all relevant explanatory factors) rather than actual discrimination. Second, even if small disparities may properly be described as discrimination, the resources necessary to eliminate them may outweigh the benefit to consumers.

Industry has asked whether the Bureau will publicly state its tolerances — that is, its standard for material insignificance. We have a number of concerns about doing so, and recommend against it. If the CFPB chooses to publish its tolerances, it should do so with language designed to avoid unintended impact on future enforcement activity.

### Factors in Setting Tolerances

Tolerances may vary based on a number of relevant factors, including the product type, the nature of the activity under consideration, and the nature of the data itself. Because of these distinctions, any announcement would need to recognize that a tolerance in auto dealer markup would not necessarily apply to other practices or products.

Loan products vary greatly, and features such as the loan term or loan amount may dramatically impact the magnitude of consumer harm represented by a particular number of basis points of disparity. For example, a mortgage loan is usually larger and is held for a longer term than an auto loan, and so similar disparities in APR may create many more dollars of harm in mortgage lending than auto lending.

Tolerances may also differ based on the activity under consideration. Fair lending analyses typically examine multiple aspects of the credit transaction, including underwriting decisions (denials), pricing (both APR and fees), steering, redlining, and more. Reasonable tolerances will likely vary for each aspect. For example, APR tolerances are typically lower than fee tolerances because discriminatory APR pricing has a greater impact over time. Put differently, a large upfront fee disparity might appear small if rolled into the APR, which amortizes that difference

<sup>10</sup> Some courts recognize the distinction between statistical significance and a higher level of significance which might be deemed “practical”, “substantive”, or “material” significance. See, e.g., *Waisome v. Port Auth. of New York & New Jersey*, 948 F.2d 1370, 1376 (2d Cir. 1991) (finding that “though the disparity was found to be statistically significant, it was of limited magnitude” and citing sources explaining the difference between statistical and practical significance). But see *Stagi v. Nat’l R.R. Passenger Corp.*, 391 F. App’x 133, 140 (3d Cir. 2010) (“As “practical” significance has not been adopted by our Court, and no other Court of Appeals requires a showing of practical significance, we decline to require such a showing as part of a plaintiff’s prima facie case.”). Statistical significance is mathematical and normally corresponds to the 95% confidence level. Material significance is subjective and depends upon the factfinder’s judgment of whether a given amount statistically proven difference matters for practical purposes.

**Comment [BES17]:** For supervisory or enforcement purposes? Seem to coningle the two.

**Comment [EBW18]:** I realize “practically insignificant” is the term used by practitioners, but it has a troubling double meaning. “Substantively insignificant” could also work.

**Comment [BES19]:** Is the only ask for a public statement? I thought that there was a desire to identify this for internal decision making purposes as well.

**Comment [BJK20]:** Given this point, might there be an argument for adopting a dollar threshold to address pricing disparities?

DRAFT – SENSITIVE – PRE-DECISIONAL – ATTORNEY WORK PRODUCT

over the life of the loan. Moreover, tolerances also may be expressed in different ways. For example, underwriting decisions are often reflected in odds ratios (e.g., African American borrowers were denied at 1.8 times the rate of non-Hispanic White borrowers), while redlining may be reflected by comparing one institution’s rate of lending in minority areas with that of its peers, which can be reflected in both absolute differences (e.g., the share of Lender A’s originations that occur in minority neighborhoods is 10 percentage points lower than its peers’) and relative differences (e.g., Lender A is only 1/5 as likely as its peers to make loans in minority neighborhoods).

Tolerances also depend on the nature of the data itself. In auto lending, for example, dealers and lenders do not collect information on race or gender, so we proxy for those characteristics. Our proxy methodology, although state-of-the-art, necessarily introduces greater uncertainty about the precision of our results. Although the effect is minor, it may increase our tolerances in auto lending relative to mortgage lending, where race and ethnicity are reported for each loan. Additional data considerations might be specific to a particular lender. For instance, if a lender had serious data integrity issues that affected our confidence in our analyses, we might have a higher tolerance compared to a lender with clean data. In short, our tolerances will depend on the circumstances of each case, or at the very least on a multitude of factors, which makes it difficult if not impossible to specify a single threshold for all cases.

Comment [BES21]: Do you mean this in a statistical sense or just that folks are skeptical?

Comment [BES22]: Um, I think that we would ask for better data.

**Peer Agencies**

The Bureau’s peer agencies (DOJ, FTC, OCC, FRB, FDIC, NCUA, HUD) make case-by-case assessments of whether to pursue supervisory or enforcement activity in response to statistical disparities. We are not aware of any agency publicly announcing its tolerances; rather, over time they tend to develop a reputation for leniency or stringency.

Comment [BES23]: I make the same point again: have our sister agencies published methodological choices (e.g., use of proxies)?

Through an examination of enforcement actions from our peer agencies, one can glean unstated tolerances for certain types of cases. These numbers should be viewed as informal enforcement tolerances; supervisory tolerances are almost certainly lower, but are not public.

For instance, over the last few decades, the lowest disparities used in a DOJ case were pricing disparities reflected through APR ranging from 5-14 basis points (██████████). The lowest pricing disparities reflected through overage/YSP involved retail disparities of 13-28 basis points, but that case also had much higher disparities (up to 107 basis points) in wholesale pricing (██████████). Another mortgage pricing case had overage/YSP disparities of 19-26 basis points in retail pricing and 16-66 basis points in wholesale pricing (██████████). The lower end of overage/YSP pricing disparities in two other cases (██████████) was 20 basis points. We can infer from these cases that DOJ deems mortgage pricing disparities of 5 basis points to be actionable when reflected through APR and deems mortgage pricing disparities of 13 basis points to be actionable when reflected through overage/YSP. However, the facts and circumstances of each case may well have influenced DOJ’s decision to pursue each case.

**The Rationale for Non-Publication**

The approach of our peer agencies, which have declined to publish tolerances, has several advantages. First, and quite simply, it allows these agencies to assess the facts and circumstances of each case and make a determination with the benefit of tangible details. For instance, an entity might have relatively low disparities, but the regulator might have evidence from a whistleblower to accompany the statistical evidence. Another hypothetical is a lender that specializes in high-value lending, such that small disparities still yield thousands of dollars in consumer harm for each loan. Yet another example is when an APR model fits the data extremely well—meaning that we believe we have accounted for most if not all factors that might legitimately explain the disparities—a lower tolerance may be appropriate. In short, a rule of thumb does not suit all fingers, and a public commitment to a certain tolerance makes it difficult to make exceptions where warranted.

Another disadvantage to publicizing a tolerance is that it may impair our future enforcement actions. A public announcement could be used by a defendant to weaken a case with disparities close to or below our publicized tolerance. This risk is particularly acute given that tolerances in APR disparities are lower than tolerances in fee disparities, but because such distinctions require a nuanced understanding of the methods involved, the higher tolerances could easily be used to undermine the lower.<sup>11</sup> Worse yet, if we publicize our tolerances, that information could be used to weaken the cases of our sister regulators as well. Even though we have no authority to decide tolerances for other regulators, a court might reasonably inquire why such tolerances would differ between federal regulators enforcing the same fair lending laws.

Another risk of publishing tolerances is that it appears to signal that the CFPB allows discrimination, just not too much. This creates headline risk for the Bureau because the public, and the consumers with whom we hope to build trust, may not fully comprehend the methodological reasons behind tolerating disparities that are statistically, but not materially, significant.

Conversely, industry will undoubtedly criticize our tolerances for being unreasonably low. News reports have already sought to minimize the potential harm in dealer markup by spreading the harm over dozens of monthly payments. If we come out with a tolerance of 5 basis points, we should expect a common refrain to be that we are worried about disparities amounting to less than “a dollar a month.”<sup>12</sup> This kind of news coverage not only risks diminishing our efforts, it also impairs our trust with American consumers and their political representatives.

Because of the foregoing risks, our recommendation is that we not publish a tolerance for disparities in the context of dealer markup (or, indeed, any other context).

### Recommended Conditions on Publication

If the Bureau chooses to announce a tolerance for markup disparities, we would recommend that the tolerance be framed as a compliance management system tolerance, and not a guide for the

**Comment [BES24]:** That could be remedied in a communication that would simply state that a threshold would not be the only criterion on which potential fair lending violations would be assessed. These are weak arguments.

**Comment [BES25]:** The same concern applies to the publication of a proxy methodology. If the methodology is publicly discredited, we might expect a similar effect on our ability to rely on this technique in enforcement matters.

**Comment [BES26]:** Enforcement work should focus on the source of the disparity (e.g., fees, points, etc.) rather than relying on APR anyway. Focusing on APR for supervisory work just makes conducting a large number of examinations possible. I think this point raises this concern: build better models that target the source of disparities.

**Comment [BES27]:** A court might raise the same concern with respect to other methodological choices.

**Comment [BES28]:** It may suggest something about the precision of the modeling.

<sup>11</sup> This risk is not merely hypothetical. We currently have an enforcement action based on APR disparities of less than 10 basis points. However, the average harm in that matter exceeds \$600 per loan.

<sup>12</sup> For an average auto loan of \$26,000 loan over 60 months, 5 basis points of disparity creates approximately \$0.60 of consumer harm each month.

**Comment [BES29]:** This seems reasonable.

DRAFT – SENSITIVE – PRE-DECISIONAL – ATTORNEY WORK PRODUCT

Bureau's enforcement action. Publishing an enforcement-based tolerance would be unwise not only because the facts and circumstances will affect our decision-making in each case, but also because our enforcement actions will likely err on the side of caution, such that any enforcement-based tolerance would be overly permissive of disparities. Instead, our communications would frame the announcement as another iteration of our guidance regarding robust fair lending compliance management systems to address fair lending risk in dealer markup policies. Framing the announcement in this way may reduce some of the risks mentioned above. It recognizes that each regulatory agency can have a different examination focus. Although there will still be some risk of precluding enforcement or supervisory actions below the published tolerance—especially given that higher tolerances can undermine lower ones—this risk can be mitigated by setting the tolerance at a sufficiently low level.

Comment [BES30]: Agreed.

We also would recommend that our tolerance be set at a level that balances at least two considerations: (1) that it be high enough to provide a meaningful safe harbor, and (2) that it be low enough to minimize the risk that we might be tempted to bring supervisory actions in response to disparities falling within the tolerance. If we choose to announce a supervisory tolerance in indirect auto markup, we would expect lenders to engage in active dealer monitoring and corrective action (e.g., dealer watch lists, lower caps on markup, etc.) when they identify disparities in dealer markup of 5 basis points or more, and commence stronger corrective action procedures (e.g., eliminating dealer discretion, remediating consumers, etc.) when they identify disparities in dealer markup of 10 basis points or more. Because the subject at issue here is dealer markup, which is a change in the APR of the loan that continues throughout the life of the loan, our tolerance would be expressed in basis points of disparity, not a dollar amount.

Comment [BES31]: In the context of mark-up in auto and compliance, a threshold is of little value in the absence of guidance on model specification (i.e., which controls to include).

Comment [BJK32]: Doesn't this arguably run contrary to our arguments about consumer harm?