



1700 G Street NW, Washington, DC 20552

CFPB Bulletin 2013-XX (Fair Lending)

Date: XXXXX XX, 2013

Subject: Proxy Methodology Employed in CFPB's Fair Lending Supervisory Activities

Several institutions and financial services industry groups have inquired about how the Consumer Financial Protection Bureau ("CFPB" or "Bureau") conducts fair lending analysis when critical demographic information—such as the race, ethnicity and sex of credit applicants or borrowers—is unavailable. This bulletin discusses techniques for proxying for such information, including the method the Bureau uses in its supervisory work, to assist lenders in monitoring for compliance with the Equal Credit Opportunity Act ("ECOA").¹

This bulletin is specific to the Bureau's fair lending supervisory activities. In an investigation or enforcement action, the Bureau may use similar or different proxy methods, depending on the specific facts and circumstances of the case.

Background on Proxies

The ECOA, as implemented by Regulation B, forbids creditors from inquiring about an applicant's demographic information, with very limited exceptions.² One important exception is the Home Mortgage Disclosure Act, which requires the collection and reporting of data on sex, race, and ethnicity by most mortgage originators. But outside of mortgage lending, fair lending statistical analyses must rely on proxies to assign race, ethnicity, or sex.

A proxy is a statistical method that uses available information to estimate unavailable information. For instance, a proxy based on first name might be used to estimate the probability that an applicant is female when the applicant's sex is not recorded. Proxies are commonly used by health scientists and economists to estimate differences across demographic groups.³ The federal banking regulators have long made clear that proxy methods may be used in fair lending examinations to estimate protected characteristics where

¹ See also CFPB Bulletin 2013-02 (Mar. 21, 2013), http://files.consumerfinance.gov/f/201303_cfpb_march_Auto-Finance-Bulletin.pdf, which discusses more generally indirect auto lenders' compliance with ECOA and suggests steps that they could take to manage fair lending risk including developing a robust fair lending compliance program and eliminating dealer pricing discretion. While the proxy issue discussed in this bulletin is more applicable to the development of a compliance program, it should not be interpreted as an endorsement of a particular method for managing fair lending risk.

² 12 C.F.R. § 1002.5(a), (b).

³ See, e.g., Marc N. Elliott et al., *A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity*, Health Services Research 43:5, Part 1 (Oct. 2008).

Comment [BES1]: If a "proxy" methodology (however defined) is debunked, does the use of "must" put us in a difficult position?

Comment [BES2]: I generally like this description. A thought: would this include/exclude exact matches to another database? Would this include the use of data from DMVs?

Comment [SW3]: I think it would be more helpful if we cite (econ and/or health science) articles here that use race/sex/ethnicity proxies to study a particular issue (the proxy need not be BISG though obviously that'd be better) rather than articles that focus on proxy methodologies themselves.

Bryce – are you aware of any such articles? could someone in OR do an article search?

Comment [BES4]: We have been conducting the research over the past year, including literature reviews. Academic economists are generally not publishing in this area with the exception of the use of multiple imputation methods for missing data.



1700 G Street NW, Washington, DC 20552

direct evidence of those characteristics is unavailable.⁴ Furthermore, for decades courts have accepted the use of reliable proxy methods in a variety of discrimination lawsuits, including ECOA actions.⁵

Proxy Methods

In general, how proxying is done depends on the characteristic being proxied. For example, to proxy for sex, the Bureau relies on a first-name database from the Social Security Administration that reports counts of individuals by sex and birth year for first names occurring at least five times for a particular sex in a birth year.⁶ The proxy method assigns a probability that a particular applicant is female based on the distribution of the population across sex categories (male or female) for the applicant's first name.

There are a greater variety of methods to proxy for race and ethnicity. The most common method for proxying the probability that an applicant is Hispanic or Asian is to use the surname database published by the Census Bureau.⁷ Another method to proxy for race and ethnicity—typically referred to as “geocoding”—uses the demographics of the census geography (e.g., census tract or block group) in which an individual's residence is located, and assigns probabilities about the individual's race or ethnicity based on the demographics of that area. This method is frequently used to proxy the probability that an applicant is African American, and it can be used to proxy for other racial and ethnic groups as well.

Over the last decade, another method to proxy for race and ethnicity has been developed that integrates the surname and geographical approaches described above. This method was developed by health research economists,⁸ and it combines the respective probabilities generated by the surname and geographical proxies. Published research has found that the integrated approach produces proxies that correlate highly with self-reported race and

Comment [BESS]: An observation: the only information that we have on residency is based on reported applicant address. I am just reacting to claims that are circulating that we are building a master linked database containing information about individuals. Should think about any sensitivities.

⁴ See *Interagency Fair Lending Examination Procedures*, at 12-13, available at <http://www.ffiec.gov/PDF/fairlend.pdf> (explaining that “[a] surrogate for a prohibited basis group may be used” in a comparative file review 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 (temporarily adopting the FFIEC Interagency Fair Lending Examination Procedures).

⁵ See, e.g., *United States v. Union Auto Sales, Inc.*, 490 F. App'x 847, 849 (9th Cir. 2012) (“classification of ‘Asians’ and ‘non-Asians’ did not render the ECOA claim any less plausible” because “[t]he link between names and racial categorization for the purposes of both antidiscrimination law and discriminatory conduct is well-established”). Examples of other types of discrimination suits in which proxies have been used and accepted include: *Benavidez v. City of Irving*, 638 F. Supp. 2d 709, 717 (N.D. Tex. 2009) (Voting Rights Act), *I.M.A.G.E. v. Bailar*, 518 F. Supp. 800, 807 (N.D. Cal. 1981) (employment discrimination) and *United States v. Reyes*, 934 F. Supp. 553, 560-62 (S.D.N.Y. 1996) (constitutional challenges to jury pool selections).

⁶ <http://www.ssa.gov/oact/babynames/limits.html>.

⁷ <http://www.census.gov/genealogy/www/data/2000surnames/index.html>.

⁸ Marc N. Elliott et al., *A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity*, HEALTH SERVICES RESEARCH 43:5, Part 1 (Oct. 2008).



1700 G Street NW, Washington, DC 20552

ethnicity data and is more accurate than using surname or geography alone.⁹ The Bureau is using the integrated proxy as the primary method in our non-mortgage analyses, though we typically test the robustness of our findings using alternative methods, such as surname or geography individually.¹⁰

There are proxy methods for race and ethnicity that use nonpublic information, such as data or photographs from Department of Motor Vehicle databases or proprietary databases matching first or middle names to certain racial or ethnic groups. For the purpose of conducting our supervisory work, we have chosen not to use proxy methods that rely on nonpublic data so that lenders can replicate our methods without purchasing proprietary databases.

Final Remarks

Depending on the size and complexity of the financial institution, a well-developed fair lending compliance program should include regular statistical analysis of loan data for potential disparities on a prohibited basis.¹¹ Outside of the mortgage context, such analysis will typically require proxies. The Bureau recognizes that there are multiple reasonable ways to proxy for protected characteristics, and lenders are not required to adopt any particular method, including the integrated proxy method described above, in their fair lending compliance programs. However, where lenders' business practices—such as compensating auto dealers via discretionary interest rate markup—present significant fair lending risk, the Bureau will consider the existence and nature of any statistical analysis in determining whether lenders have adequately managed and monitored that fair lending risk.¹²

Comment [SW6]: The word "efficient" was used here. I don't think the technical definition of efficiency is well known so it is probably better just to say "accurate."

Bryce – your thoughts?

Comment [BES7]: I am ok with accuracy. Could also skirt the issue and say something like "performs better?"

Comment [SW8]: Should we be clear that we use census tract-level information, not census block?

Comment [SW9]: Should this be "mortgage origination?" Do servicers typically have GMI when they did not originate the loans themselves?

⁹ Marc N. Elliott et al., *Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities*, HEALTH SERVICES & OUTCOMES RESEARCH METHODOLOGY (2009) 9:69-83.

¹⁰ In addition to trying different proxy methods, we also check the robustness of our findings by using the proxied probabilities in different ways when estimating disparities. We estimate models that use the probabilities directly; we also consider models that apply threshold-based classification rules that assign an application to a single racial or ethnic group by determining whether the probability associated with each classification meets or exceeds a particular threshold.

¹¹ See Supervisory Highlights: Fall 2012 at 6-7, http://files.consumerfinance.gov/f/201210_cfpb_supervisory-highlights-fall-2012.pdf.

¹² For more information on managing fair lending risk, see CFPB Bulletin 2013-02, *supra* at n. 1.