fʌɪrplay

<u>Via electronic submission</u>

August 30, 2022

Federal Housing Finance Agency
Office of Financial Technology
400 7th Street SW, 5th Floor
Washington, DC 20219

RE: Fintech in Housing Finance: Request for Information, July 2022

Fairplay writes in response to the Office of Financial Technology's request for information (RFI) on the role of fintech in housing finance.

Fairplay is the world's first fairness-as-a-service company. In financial services, we use machine learning algorithmic techniques to evaluate the fairness of all aspects of the credit lifecycle, from marketing to credit underwriting to model governance. FairPlay supports and appreciates the RFI's recognition that, "*new techniques can be applied to fair lending testing to identify less discriminatory alteratives and ensure that proxies for protected class status are not included in a model*."

Fairplay has focused the comments in this letter to the public data made available by the Federal Housing Finance Agency (FHFA) and other financial regulators to monitor lending fairness. Our response directly addresses:

- **RFI Question B4**: "What are the existing data challenges that most prevent data-driven decision making in the mortgage lifecycle?"
- **RFI Question C1:** "What new fintech tools and techniques are emerging that could further equitable access to mortgage credit and sustainable home ownership?"
- **RFI Question C2**: "What emerging techniques are available to facilitate or evaluate fintech compliance with fair lending laws?"

Our central recommendation is that FHFA could enhance lending fairness by increasing the disclosure of records from its National Survey of Mortgage Originations (NSMO) public use file from the current 53,000 records to at least 100,000 records. The NSMO public use file is unique in that it is the only government-supplied mortgage data source that provides outcome data linked to protected status characteristics such as race, color, religion, national origin, sex, marital status and age. The Home Mortgage Disclosure Act (HMDA) database, for example, has millions of records, but, "*contains no information on loan performance, little information on borrower creditworthiness, and ha[s] up to a 21-month delay in release." (*source*:* National Mortgage Database (NMDB) Technical Report 1.1 (December 22, 2016)).

The NSMO public use file contains 52,792 sample mortgages originated from 2013 through 2021 based on 5,000 to 6,000 usable surveys per year. These surveys are linked to a supplemental file from Experian that allows for analyses of loan performance in connection with the race and other protected status characteristics of the applicant. The limited size of this data set, combined with the fact that it is a survey with self-selected respondents, constrains its utility.

Why do we need more public data on the performance of loans connected to applicants with protected status? The Equal Credit Opportunity Act (ECOA) limits lenders, subject to certain activities like self-testing, from collecting protected status information from credit applicants. Thus, government data sets are a valuable source of this information.

Unfortunately, the available evidence shows that current lending policies and regulations are not moving the needle on lending fairness, and as such we must find alternatives. Tracking a broader swath of performance data on loans issued to protected status applicants could help remove some of the fog in this area.

Recently, FairPlay did an analysis of 30 years of mortgage filings, relying principally on the HMDA database. In 2021, there were 23 million loan applications which resulted in 15 million mortgage loan originations, approximately a 65% approval rate. Our study found that lending fairness in 2021 — as measured by Adverse Impact Ratio (AIR) — remains stuck at 1990 numbers.[1] AIR for Native Americans was down from 94% in 1990 to 82% in 2021. AIR for Black applicants on mortgages registered at 80% in 1990, and was stuck in the high 70s through 2019, until a recent bump to 84% in 2021. AIR for HIspanic applicants was just 85% in 1990, and was still below 90% in 2021.

Increasing NSMO public use records to over 100,000 borrowers and expanding the number of data fields would be one way to meaningfully enhance the data resources available to tackle lending fairness.

How could increasing the size of the NSMO public use dataset help? There are multiple algorithmic methods which can be applied to reduce disparity in underwriting or pricing of mortgage loans. The most promising of these methods are modifications of classic machine learning systems (e.g. FairLearn, Adversarial Debiasing, and Distribution Matching).

Analytics companies currently can use NSMO train models to yield fairer lending outcomes. The predictive power of these models is limited by two factors: the restricted number of fields from NSMO which are actually usable during training, and the limited number of records in NSMO. In fact, there are fewer than 20 fields in the NSMO records

---

[1] AIR reflects the percentage of one group's loan approval rates relative to a control group (typically White applicants). So, for example, if blacks had a 60% credit approval rate, and whites had an 80% approval rate the AIR would be 60/80, or 75%. In general, AIR of less than 80% is an indicator of disparate impact.

which are available for predicting loan outcome at the time when loans are being approved or denied and when they are being priced.

       Bottom-line: NSMO is a great resource. By expanding the size of the data set that is released publicly, federal regulators could provide even more value to underserved populations and disadvantaged consumers.

       Thank you for your consideration.

Sincerely,

Thomas Oscherwitz

Thomas Oscherwitz
Vice President of Policy and Regulatory Affairs
FairPlay