

When Mortgage Data Can't Prove Discriminatory Lending

By **Abe Chernin, Shane Oka and Kevin Oswald** (November 19, 2025)

The Home Mortgage Disclosure Act and subsequent amendments expanding the universe of application data collected have been important steps forward in fair lending transparency.[1] But does HMDA data tell a complete story of a lender's credit decision?

In this article on detecting discriminatory mortgage lending, we discuss the importance of controlling for applicant characteristics, the limitations of HMDA data's ability to do so, and the need for manual loan file review to confirm indications of potential discriminatory lending.

There has been significant academic and regulatory discussion regarding the role statistical analysis should play in detecting discriminatory mortgage lending patterns and, more particularly, the use of public HMDA data in such analysis. While federal regulators have dialed back their fair lending enforcement under the Trump administration, private plaintiffs continue to use public HMDA data to initiate litigation.

For example, a 2022 Bloomberg article used public HMDA data to allege that Wells Fargo approved a lower percentage of refinancing applications for Black applicants compared to white applicants.[2]

Subsequently, In re: Wells Fargo Mortgage Discrimination Litigation, a class action filed in the U.S. District Court for the Northern District of California, alleged that Wells Fargo illegally discriminated against minority mortgage applicants.

On Aug. 5, 2025, the court denied class certification, finding that the plaintiff focused on the statistical disparity in application denial rates, but failed to present classwide evidence that Wells Fargo caused the disparity.[3]

In March 2023, the Consumer Financial Protection Bureau released a report discussing the effectiveness of the HMDA "in identifying possible discriminatory lending patterns." [4]

The report found the collection of additional loan-level risk characteristics beginning in 2018 "improve[d] the quality of statistical analyses conducted" [5] and that the "new HMDA data are often used to build evidence for legal cases and estimate appropriate remuneration amounts for harmed consumers." [6]

But a key question persists — can statistical analysis of HMDA data alone identify discriminatory lending to a sufficient degree of certainty?

When attempting to identify discriminatory lending patterns, careful empirical analysis is required to avoid erroneous conclusions. Importantly, a statistical analysis failing to account for relevant underwriting variables risks finding lending discrimination even when there is not any.



Abe Chernin



Shane Oka



Kevin Oswald

Statistical estimates of mortgage discrimination can be biased when an analysis does not control for underwriting factors that affect the lending decision and correlate with race.

For example, if it were a lender's policy to deny applicants who have had a recent bankruptcy, and minority applicants were more likely to have had a recent bankruptcy, then minorities may have a higher denial rate than other applicants, even if the lender does not explicitly consider race.

A statistical analysis that does not control for recent bankruptcies (i.e., compare denial rates of minority applicants against white applicants with similar bankruptcy histories) may therefore find discrimination even if there is none due to the omission of this underwriting factor. Such a phenomenon is known as omitted variable bias.

To illustrate the impact of omitted variable bias, Exhibit 1 below plots the excess Black denial rate, i.e., the difference between the denial rate of Black applicants and white applicants, on the vertical axis against the R-squared on the horizontal axis for 10 large mortgage lenders in 2024.[7]

R-squared is a standard statistical measure and in this context shows how well loan decisions are explained by the statistical model used, where an R-squared of 1 perfectly explains loan decisions. In Exhibit 1, three different models of increasing complexity are used for each lender.

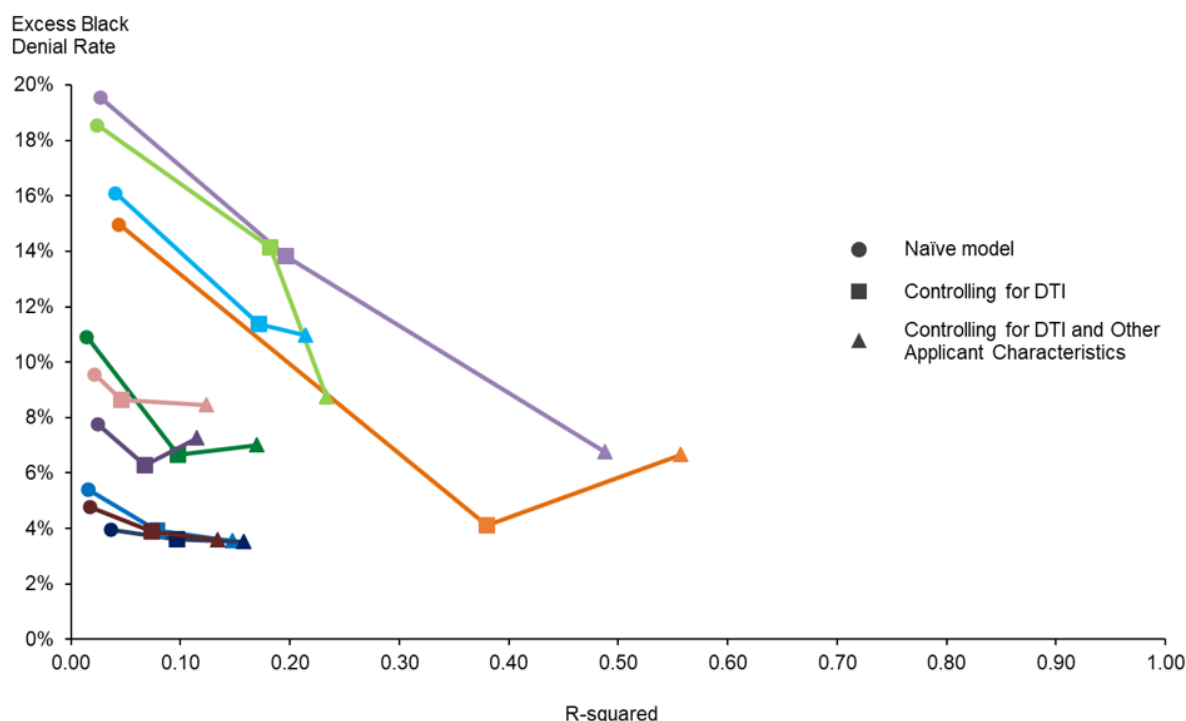
First, we plot the results of a naïve model that does not control for any applicant characteristics, and therefore, is simply plotting the raw difference in denial rate by race.

This results in high excess Black denial rates and low R-squareds, suggesting that race alone does not do a good job of explaining loan decisions.

Second, we plot the results of a model that controls for one important application characteristic, debt-to-income, or DTI. This results in a higher R-squared and smaller excess Black denial rates, meaning that this model is able to explain a portion of the disparity in denial rates as being based on DTI, which the naïve model inappropriately attributed to race.

Third, we plot the results of a model that controls for DTI and additional important applicant characteristics reported in public HMDA data, such as loan-to-value and income.[8] As with the second model, this results in a higher R-squared and on average results in a lower excess Black denial rate.

Exhibit 1: Illustration of Omitted Variable Bias Change in Excess Black Denial Rate by Lender



There are two important takeaways from Exhibit 1. First, the R-squared does not exceed 56% for any of these lenders. This means that for these lenders, public HMDA data explains just over half of the variation in lending decisions at most, and often much less. Second, adding applicant characteristics to the naïve model affects lenders to varying degrees.

As shown below, this heterogeneity is due in part to each lender's unique applicant population.

Thus, controlling for applicant characteristics can result in very different impacts across lenders, meaning that any comparisons across lenders must also account for differences in those lenders' applicant pools.

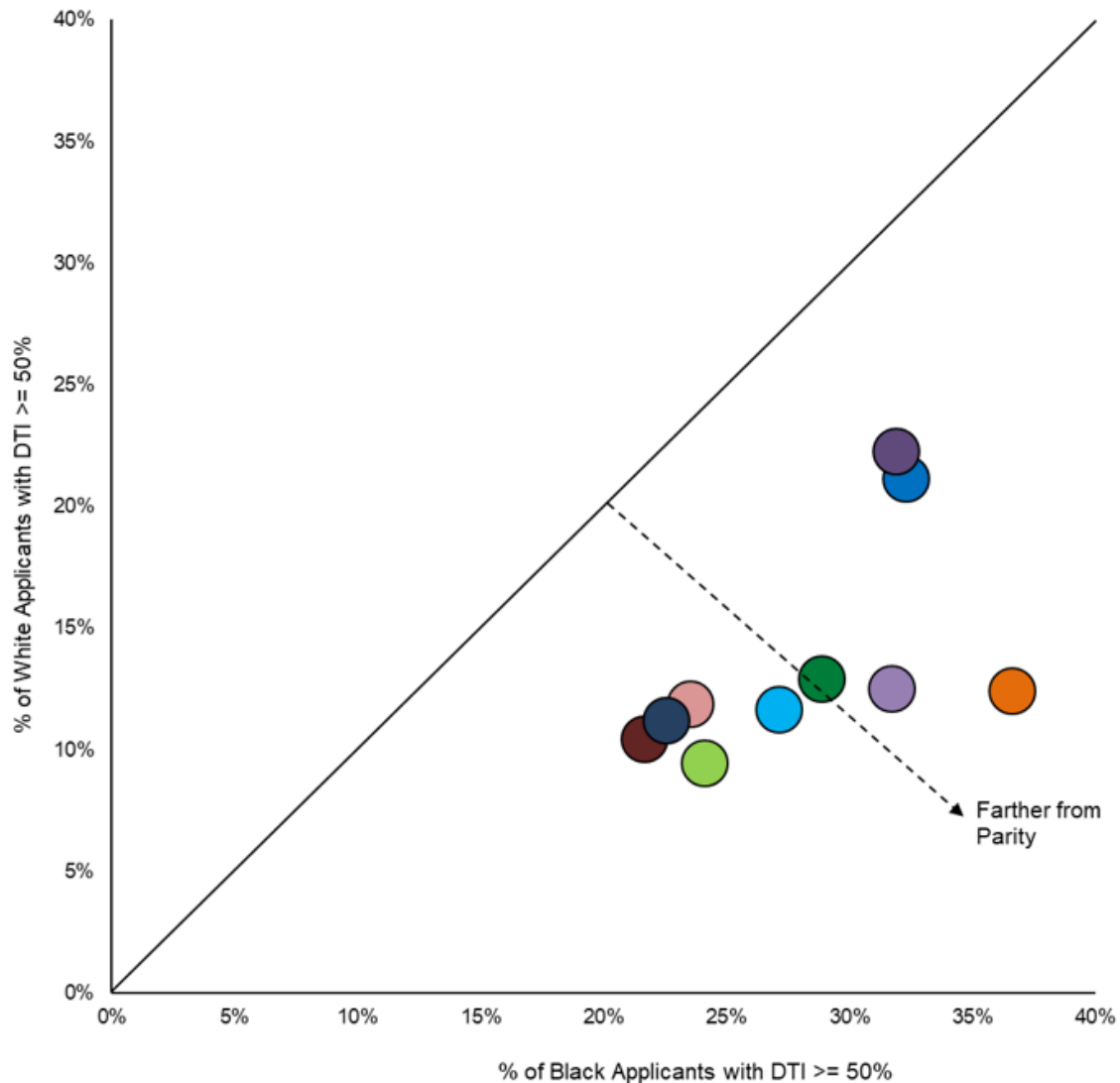
We highlight differences in applicant profiles across lenders in Exhibit 2 below. One important lending criteria for many lenders is whether applicants have a DTI ratio of 50% or higher, with a higher DTI ratio generally signifying higher credit risk.

Exhibit 2 shows that while the proportion of Black applicants having a DTI of 50% or higher is larger than the same proportion for white applicants for each of the 10 lenders, there is variation in this proportion across lenders.

For example, the lender represented by an orange circle has the largest disparity in DTI between its Black applicants and white applicants.

Unsurprisingly, moving from a naïve model that controls only for race to a model that also controls for DTI also has the biggest impact on this lender, as can be seen from the large increase in R-squared for the orange line in Exhibit 1.

Exhibit 2: Illustration of Differences in Applicant Pool by Lender
Percent of Applicants with DTI Ratio Greater Than 50% by Race and Lender



While there is a confidential, private version of HMDA data that contains credit score and many other relevant variables, in our experience, neither version of HMDA data can account for all important underwriting variables.

Critically, HMDA data lacks important quantitative data (e.g., prior bankruptcies) and qualitative data (e.g., responsiveness of the applicant to requests for more information).

As further illustration of the limitations of a statistics-only approach, Neil Bhutta, Aurel Hizmo and Daniel Ringo conducted an analysis, published in 2022, of 9 million mortgage applications across lenders that use the confidential, private version of HMDA data. They calculated an R-squared of 39.8%, which means that the majority of the variation in loan outcomes is still unexplained.

Drawing conclusions about discrimination based solely on statistical analysis that fails to

explain significant variation in loan outcomes ignores the potential for omitted variable bias.

To address the inherent risk of omitted variable bias, individual loan file review is often required because it can incorporate both quantitative and qualitative information relevant to the loan decision. The Office of the Comptroller of the Currency describes statistical analysis, such as regressions, as the initial "scoping" step in a fair lending examination to identify potential discrimination.[9]

After scoping, the OCC guidelines state that examiners should conduct a detailed review of a sample of loan applications, including information on the applicant's qualifications, the level of assistance received during the application process, the reasons for denial, the loan terms and other information.[10] Only after moving beyond a purely statistics-only approach should the examiner potentially conclude that discrimination has occurred.

In summary, when dealing with allegations of mortgage lending discrimination, it is important for litigators to consider the limitations of a statistics-only approach.

A careful statistical regression based on HMDA data and internal lender data can be useful in finding potential indications of discrimination, and with advances in AI technology, a statistics-only approach is increasingly able to incorporate some qualitative information.

However, manual loan file review is usually still needed to confirm such indications of discrimination.

Abe Chernin is a senior vice president, and Shane Oka and Kevin Oswald are principals, at Cornerstone Research.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of their employer, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] HMDA requires many financial institutions to maintain, report, and publicly disclose loan-level information about mortgages. See <https://www.consumerfinance.gov/data-research/hmda/>.

[2] <https://www.bloomberg.com/graphics/2022-wells-fargo-black-home-loan-refinancing/>.

[3] In re Wells Fargo Mortgage Discrimination Litigation, Order Re Class Certification, filed August 5, 2025.

[4] https://files.consumerfinance.gov/f/documents/cfpb_hmda-voluntary-review_2023-03.pdf, pp. 26–35, 97.

[5] https://files.consumerfinance.gov/f/documents/cfpb_hmda-voluntary-review_2023-03.pdf, p. 106.

[6] https://files.consumerfinance.gov/f/documents/cfpb_hmda-voluntary-review_2023-03.pdf, p. 107.

[7] There were 168 lenders in 2024 that had at least 10,000 observations. 10 lenders were

randomly chosen from among this set and have been anonymized.

[8] The variables controlled for are race, lender fixed effects, DTI, LTV, total number of units, loan term, whether a government AUS was run on the application, whether the application has a co-applicant, loan type, conforming loan limit, whether the purpose of the property is business or commercial, whether the property is a primary residence, log loan amount, log income, whether the property is a manufactured home, whether it is a fixed rate mortgage, lien status, and whether there is a prepayment penalty.

[9] Federal Deposit Insurance Corporation, "Interagency Fair Lending Examination Procedures," available at <https://www.fdic.gov/regulations/examinations/fairlend.pdf>, p. 1.

[10] Federal Deposit Insurance Corporation, "Interagency Fair Lending Examination Procedures," available at <https://www.fdic.gov/regulations/examinations/fairlend.pdf>, p. 17.