WORKING PAPER · NO. 2025-112

Differential Validity in Fair Lending

Spencer Caro, Talia Gillis & Scott Nelson

AUGUST 2025

THE UNIVERSITY OF CHICAGO
BECKER FRIEDMAN INSTITUTE FOR ECONOMICS

5757 S. University Ave.
Chicago, IL 60637
Main: 773.702.5599
bfi.uchicago.edu

DIFFERENTIAL VALIDITY IN FAIR LENDING

Spencer Caro, Talia Gillis & Scott Nelson*

Fair lending's disparate impact doctrine aims to address lending disparities. But which disparities? Traditional fair lending has narrowly focused on equal outcomes—examining differences in loan approval rates or interest rates. However, this singular focus overlooks other critical dimensions of disparities that are essential for fair credit access.

This Article challenges the conventional emphasis on equal outcomes, demonstrating how it has failed to address deep-rooted inequalities in traditional credit allocation while also stifling innovation in machine-learning and alternative data. We argue that disparities in the validity of creditworthiness predictions —the accuracy with which a model identifies creditworthy applicants—severely impact equal access to credit and, in particular, the extension of credit to the creditworthy. Despite mounting empirical evidence of the harm of validity disparities, traditional fair lending enforcement inadequately recognizes this disparity dimension, a gap that may become increasingly harmful as lending decisions rely on advanced statistical methods. Future regulatory guidance, enforcement and supervision should explicitly recognize validity inequalities across protected groups while addressing the accompanying challenges of this more comprehensive perspective on disparities, which is essential for equitable credit allocation.

* Spencer Caro, Independent.
Talia Gillis, Associate Professor, Columbia Law School.
Scott Nelson, Assistant Professor, Chicago Booth.
Many thanks to Susan Block-Lieb, Dorothy Lund, Jared Mayer, Jacob Noti-Victor, Holli Sargeant, Bryce Stephens, Daniel Schwarcz, Rory Van Loo and participants in the Consumer Law Scholars Conference and Cardozo Metropolitan Junior Scholars' Workshop for helpful comments and discussions. Rotem Litinski, Sophia Stener and Ellery Weiner provided excellent research assistance.

Table of Contents

INTRODUCTION

There is emerging consensus that antidiscrimination law is in urgent need of reform.[1] Whether due to the increasing use of alternative data[2] or machine-learning tools,[3] fair lending's disparate impact framework is ill-equipped to achieve fair outcomes in modern underwriting.[4] At the same time, the "algorithmic fairness" literature, originating in computer science and statistics, has highlighted the many ways statistical predictions can result in and perpetuate disparities across groups,[5] as well as the

[1] *See, e.g.*, Talia B. Gillis, *The Input Fallacy*, 106 MINN. L. REV. 1175-1263 (2022) (arguing that focusing solely on the inputs to algorithms, rather than their outputs and the processes they implement, misses critical aspects of discrimination and fairness in automated decision-making); Pauline T. Kim, *Data-Driven Discrimination at Work*, 58 WM. & MARY L. REV. 857-935 (2017) (exploring how data-driven algorithms in employment can perpetuate and even exacerbate workplace discrimination, calling for stronger regulatory oversight and transparency); Pauline T. Kim, *Race-Aware Algorithms: Fairness, Nondiscrimination and Affirmative Action*, 110 CAL. L. REV. 1539-96 (2022) (discussing the potential for race-aware algorithms to enhance fairness and nondiscrimination in decision-making processes, while also engaging with the legal and ethical challenges they present); Daniel E. Ho & Alice Xiang, *Affirmative Algorithms: The Legal Grounds for Fairness as Awareness*, U. CHI. L. REV. ONLINE (2020), https://lawreview.uchicago.edu/online-archive/affirmative-algorithms-legal-grounds-fairness-awareness (analyzing race-aware algorithms through the lens of affirmative action and equal protection jurisprudence); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CAL. L. REV. 671-732 (2016) (arguing that big data analytics can produce disparate impacts that disadvantage protected groups, necessitating new legal frameworks to address these harms effectively); Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, 119 MICH. L. REV. 291-396 (2020) (proposing a novel statistical and legal framework to ensure that algorithmic decision-making processes comply with equal protection principles and reduce discriminatory outcomes); Aziz Z. Huq, *Racial Equity in Algorithmic Criminal Justice*, 68 DUKE L. J. 1043-1134 (2019) (examining the challenges and potential solutions for achieving racial equity in the use of algorithms within the criminal justice system, advocating for systemic reforms and cautious implementation of these technologies); Sonja B. Starr, *Statistical Discrimination*, 58 HARV. C.R.-C.L. L. REV 580 (2023) (arguing that the use of explicitly racialized algorithms in legal and medical contexts is pervasive and illegal).

[2] *See generally* Karan Kaul, *Adopting Alternative Data in Credit Scoring Would Allow Millions of Consumers to Access Credit*, URBAN INSTITUTE (Mar. 15, 2021), https://www.urban.org/urban-wire/adopting-alternative-data-credit-scoring-would-allow-millions-consumers-access-credit (discussing the potential benefits of incorporating alternative data into credit scoring systems to enhance access to credit).

[3] *See generally* FINREGLAB, THE USE OF MACHINE LEARNING FOR CREDIT UNDERWRITING 8-28 (2021) (discussing the challenges and opportunities of moving to machine-learning underwriting models).

[4] Talia B. Gillis & Jann L. Spiess, *Big Data and Discrimination*, 86 U. CHI. L. REV. 459-88 (2019) (arguing that big data technologies, while offering potential benefits, pose significant risks of discrimination that require careful legal and regulatory intervention to ensure fairness and equity in their application); Holli Sargeant, *Algorithmic Decision-Making in Financial Services: Economic and Normative Outcomes in Consumer Credit*, 3 AI & ETHICS 1295-1311 (2023) (examining the economic and normative implications of algorithmic decision-making in consumer credit markets, and arguing for the careful design and regulation of these systems to balance efficiency and fairness); Nikita Aggarwal, *The Norms of Algorithmic Credit Scoring*, 80 CAMBRIDGE L.J. 42-73 (2021) (examining the regulatory challenges of algorithmic credit scoring in the UK, arguing that current frameworks inadequately balance the core norms of allocative efficiency, distributional fairness, and consumer privacy); Christophe Hurlin, Christophe Pérignon, & Sébastien Saurin, *The Fairness of Credit Scoring Models*, ARXIV:2205.10200 (2021) (proposing a post-processing method to neutralize specific variables to enhance fairness without sacrificing predictive performance).

[5] *See, e.g.*, Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold & Richard Zemel, *Fairness Through Awareness*, PROCEEDINGS OF THE 3RD INNOVATIONS IN THEORETICAL COMPUTER SCIENCE CONFERENCE 214-226 (2012) (introducing a formal definition of fairness based on treating similar individuals similarly); Jiahao Chen, Nathan Kallus, Xiaojie Mao, Geoffry Svacha & Madeleine Udell, *Fairness Under Unawareness: Assessing Disparity When Protected Class Is Unobserved*, FAT* '19: PROCEEDINGS OF THE CONFERENCE ON FAIRNESS, ACCOUNTABILITY, AND TRANSPARENCY 339-348 (2019) (proposing a method for assessing disparities when protected group membership is unobserved); Matt J. Kusner, Joshua R. Loftus, Chris Russell, Ricardo Silva, *Counterfactual Fairness*, NIPS'17: PROCEEDINGS OF THE 31ST INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS 4069-4079 (2017) (defining fairness using counterfactuals that compare individuals across demographic groups); Moritz Hardt, Eric Price & Nati Srebro,

inherent tensions between competing fairness definitions.[6] However, much of this literature remains disconnected from the practical realities of specific decision-making contexts, particularly fair lending, and rarely addresses how law and regulation conceptualize, measure, and respond to prediction disparities.[7]

Persistent inequalities in credit markets—often rooted in historical patterns like redlining—compound the challenges of addressing disparities in lending. These inequalities take many forms: racial and ethnic minorities often have, on average, lower wages and wealth, reduced access to intergenerational financial support, and thinner or more limited credit histories. For example, the descendants of victims of federally sanctioned redlining have lower wealth than their peers,[8] while credit reports contain less data for historically disadvantaged groups frequently because of the *inherited* lack of access to credit.[9] As a result, racial and ethnic minorities are often less able to present the kinds

---

Equality of Opportunity in Supervised Learning, NIPS'16: PROCEEDINGS OF THE 30TH INTERNATIONAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS 3315 (2016) (defining fairness in terms of equal true positive rates across groups to address disparities in predictive performance).

[6] *See, e.g.*, Jon Kleinberg, Sendhil Mullainathan & Manish Raghavan, *Inherent Trade-Offs in the Fair Determination of Risk Scores*, PROC. OF THE 8TH INNOVATIONS IN THEORETICAL COMPUTER SCIENCE CONFERENCE 43:1-23 (2017) (demonstrating the impossibility of satisfying three fairness measures simultaneously outside of severely constrained edge cases); Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5(2) BIG DATA 153-163 (2017) (demonstrating that several fairness criteria cannot be satisfied simultaneously in the context of recidivism prediction); Sorelle A. Friedler, Carlos Scheidegger & Suresh Venkatasubramanian, *The (Im)possibility of Fairness: Different Value Systems Require Different Mechanisms for Fair Decision Making*, 64(4) COMMUNICATIONS OF THE ACM 136 (2021) ("Different value judgments can require satisfying contradicting fairness properties each leading to different societal outcomes.").

[7] *See* I. Elizabeth Kumar, Keegan E. Hines & John P. Dickerson, *Equalizing Credit Opportunity in Algorithms: Aligning Algorithmic Fairness Research with U.S. Fair Lending Regulation*, AIES '22: PROCEEDINGS OF THE 2022 AAAI/ACM CONFERENCE ON AI, ETHICS, AND SOCIETY 357-368 (2022) (arguing that the algorithmic fairness literature and fair lending regulatory practice should be better aligned); RELMAN COLFAX PLLC, FAIR LENDING MONITORSHIP OF UPSTART NETWORK'S LENDING MODEL, FIRST REPORT OF THE INDEPENDENT MONITOR at 7 (2021), https://www.relmanlaw.com/media/cases/1088_Upstart%20Initial%20Report%20-%20Final.pdf ("…[M]any 'fairness' proposals do not engage or align with the established three-step disparate impact analysis reflected in case law and regulatory materials.").

[8] *See* Disa Hynsjo & Luca Perdoni, *Mapping Out Institutional Discrimination: The Economic Effects of Federal "Redlining"* (CESifo Working Paper No. 11098, 2024), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4820845; Daniel Aaronson, Daniel Hartley, and Bhashkar Mazumder, *The Effects of the 1930s HOLC "Redlining" Map*s, 13 AM. ECON. J. ECON. POLICY 355-392 (2021); Robert C. Ellickson*, Stale Real Estate Covenants*, 63 WM. & MARY L. REV. 2033 (2022) (critiquing the persistence of outdated real estate covenants, examining their interaction with zoning ordinances and their impact on urban land use, while highlighting the potential for these covenants to perpetuate segregation and restrict land use flexibility). *See also* ROBERT C. ELLICKSON, AMERICA'S FROZEN NEIGHBORHOODS: THE ABUSE OF ZONING, 13 (2020) ("[L]ocal politics tends . . . to freeze the zoning of an established neighborhood of single-family houses. These neighborhoods blanket not only much of suburbia but also many central cities. . . . The freezing of zoning prevents a market economy from responding to changes in the forces of supply and demand. Frozen zoning prevents homebuilders from offering households new residential options.").

[9] *See* CFPB, *Who are the Credit Invisibles?* (Dec. 2016), https://files.consumerfinance.gov/f/documents/201612_cfpb_credit_invisible_policy_report.pdf (identifying disproportionate "credit invisibility" among consumers who are young, Black or Hispanic, or in low-income neighborhoods); *see also* UnidosUS, *Unscoreable: How The Credit Reporting Agencies Exclude Latinos, Younger Consumers, Low-Income Consumers, and Immigran*ts (Feb. 26, 2019), https://www.congress.gov/116/meeting/house/108945/witnesses/HHRG-116-BA00-Wstate-BrownJ-20190226.pdf ("Among the factors that contribute to the prevalence of credit invisibility for these populations is that mainstream credit scoring models rely on formulas and algorithms that fail to consider cultural norms, such as a reluctance to accumulate

of financial profiles that current underwriting systems reward.[10] These structural differences pose a core challenge for fair lending: how should credit allocation be judged fair when applicant groups are situated so differently from the start?

Fair lending has largely addressed these structural disparities through the doctrine of disparate impact, with a strong emphasis on outcome disparities—such as the differences in approval rates or interest rates across groups. In both litigation and enforcement, this focus is often operationalized through metrics such as the adverse impact ratio (AIR) and standardized mean differences (SMD).

This paper, drawing on research in algorithmic fairness and economic studies of lending disparities, argues that this singular focus on outcome disparities misses a crucial dimension: the accuracy of creditworthiness predictions across groups, which we term *differential validity*. This oversight has allowed some of the most harmful lending practices to persist while stifling necessary innovations in machine learning and alternative data.

Our Article makes three contributions to address these challenges. First, we analyze how current fair lending enforcement has come to rely heavily on outcome metrics, such as loan approval rates and pricing differentials, to identify disparate impact. Second, we show that this approach overlooks a separate form of disparity: when predictive models perform worse for some groups than others, which we call differential validity. Third, we argue that addressing validity disparities aligns with the goals of fair lending, offering a more legally and operationally robust framework for equitable outcomes in lending.

Regarding our first contribution, we argue that traditional fair lending's shortcomings stem from its narrow focus on a single dimension of lending disparities. Measuring and evaluating these disparities is crucial for fair lending compliance. Initiating a disparate impact claim requires demonstrating that a seemingly neutral lending policy results in disparities for a protected group, and prevailing in such a claim requires assessing whether there is a feasible, less discriminatory alternative. In practice, regulatory guidance, enforcement action, and private litigation have predominately focused on quantitative disparities in loan approval rates and terms to establish claims and defenses.

debt, reliance on cash, credit history from other countries, or failure to account for other methods of making on-time payments—which keep these individuals out of the financial mainstream.").

[10] In the context of employment discrimination, s*ee* Patrick Kline, Evan K. Rose & Christopher R. Walters, *Systemic Discrimination Among Large U.S. Employers*, 137 Q.J. ECON. 1963-2036 (2022) (investigating systemic discrimination by large U.S. employers through a large-scale correspondence experiment); Patrick Kline et al., *A Discrimination Report Card* (NBER Working Paper No. 32313, Apr. 2024), https://www.nber.org/papers/w32313 (analysis of firm-specific discrimination metrics suggests that rigorous, systematic auditing can identify and mitigate discriminatory practices effectively). *See* Luke Herrine, *Credit Reporting's Vicious Cycles*, 40 N.Y.U. REV. L. & SOC. CHANGE 305-35 (2016) (analyzing the self-reinforcing nature of credit reporting systems, arguing that they perpetuate economic inequalities and hinder financial mobility for marginalized communities). *See also* Kerwin K. Charles & Erik Hurst, *The Transition to Home Ownership and the Black-White Wealth Gap*, 17 REV. ECON. & STAT. 281-307 (2003) (exploring how differences in home ownership rates contribute to the persistent wealth gap between Black and White households in the United States); Moritz Kuhn, Moritz Schularick, & Ulrike I. Steins, *Income and Wealth Inequality in America*, *1949-2016*, 75 Q.J. ECON. 1419-1450 (2020) (providing a comprehensive analysis of the evolution of income and wealth inequality in the U.S., highlighting the growing disparities and their implications for economic stability); Kerwin K. Charles & Erik Hurst, *The Correlation of Wealth across Generations*, 32 J. POL. ECON. 115-137 (2002) (investigating the intergenerational transmission of wealth, showing that parental wealth significantly influences the economic outcomes of their children).

Our second contribution is to highlight a different kind of disparity—differences in predictive accuracy across groups.[11] Loan underwriting often involves predicting future loan repayment—estimating whether a loan will be repaid or if a borrower will default. Although creditworthiness predictions necessarily come with some uncertainty—some approved borrowers might have defaulted and some rejected applicants might have repaid—when prediction inaccuracies disproportionately occur for protected groups, they lead to differential validity, a distinct type of disparity from outcome disparities. Differential validity arises when a model's predictive performance varies across groups, leading to systematic disadvantages for certain populations.

In recent years, the algorithmic fairness literature has emphasized the ways in which statistical predictions can create disparities and how these disparities may arise under different definitions of error rates and accuracy. However, fair lending in practice has yet to engage meaningfully with the implications of these findings. Our purpose is to show how validity disparities in lending create unique harms, particularly in the context of longstanding structural inequalities tied to race, ethnicity, gender, and other protected classes. Despite this growing body of evidence on the importance of the multi-faceted nature of disparities, fair lending regulation remains narrowly focused on outcome inequality.

Prior work has demonstrated the prevalence of differential validity in credit screening tools, for reasons including discrimination, data inequality, and statistical models that perform poorly for historically disadvantaged groups.[12] Differential validity occurs when the correlation between screening criteria, such as credit score, and later outcomes, such as loan default, differs between groups.[13] This is exactly what credit scores exhibit for minority and non-minority loan applicants.[14] While concerns about validity have gained some recognition in the context of employment discrimination—the touchstone of fair lending disparate impact—they have been largely neglected in

[11] These notions include calibration, "balance for the positive class," "balance for the negative class," etc. *See* Kleinberg et al., *supra* note 6, for a discussion of how these three are incompatible with each other and, typically, with equal outcomes as well.

[12] *See, e.g.*, Muhammad Bilal Zafar et al., *Fairness Beyond Disparate Treatment & Disparate Impact*, PROC. OF THE 26TH INT'L CONFERENCE ON WORLD WIDE WEB (2017) (proposing a notion of "disparate mistreatment" defined in terms of the rates of misclassification across social groups); Robert B. Avery et al., *Credit Risk, Credit Scoring, and the Performance of Home Mortgages*, FED. RES. BULL. 621, 630 (July 1996) ("If the baseline population used to generate the scoring index is not sufficiently diverse, then scores may lack predictive power for the underrepresented segments of the overall population. For example, rent, utility, and other nonstandard payment histories, which are often considered important for low-income populations, are frequently left out of scoring models. Thus, scores for these populations may not reliably assess individual risk."); Marco Di Maggio et al., *Invisible Primes: Fintech Lending with Alternative Data* (NBER Working Paper No. 29840, 2022), https://www.nber.org/system/files/working_papers/w29840/w29840.pdf (finding that new underwriting data can reveal that some nominally "subprime" consumers are actually "prime" and are in this sense "invisible primes").

[13] As we explain in more detail in Section II.A., *infra*, we use a modular definition for validity but focus for concreteness on a particular notion of differential validity: the true positive rate gap, meaning the gap between groups the approval rate of true loan repayers. For a discussion of other possible validity measured discussed extensively in the algorithmic fairness literature, *see* Dana Pessach & Erez Shmueli, *Algorithmic Fairness*, ARXIV PREPRINT AT ARXIV:2001.09784 (2020); Sahil Verma & Julia Rubin, *Fairness Definitions Explained*, FairWare'18: PROC. OF THE IEEE/ACM INT'L WORKSHOP ON SOFTWARE FAIRNESS at 1-7 (2018). *But see* Sam Corbett-Davies et al., *The Measure and Mismeasure of Fairness*, 24 J. MACHINE LEARNING FAIRNESS 1-117, 5 (2023) for extended critical discussion (arguing that many mathematical definitions of fairness can yield strongly Pareto-dominated decision policies and consequently "can, perversely, harm the very groups they were designed to protect.").

[14] *See* Laura Blattner & Scott Nelson, *How Costly is Noise? Data and Disparities in Consumer Credit*, arXiv:2105.07554 (2021) (examining how data noise in consumer credit reporting can exacerbate disparities in credit access and outcomes, particularly affecting disadvantaged groups).

lending discrimination regulation. In practice, fair lending prioritizes outcome equality over validity, overlooking the question of whether loans and loan terms allocated to the consumers who are the best match for them.[15]

Addressing differential validity is critical for fair lending, as it reflects a lender's failure to equally extend credit to the creditworthy. Whether because credit data is noisier for racial minorities and other protected groups or because models are unable to account for the complex ways in which features predict creditworthiness, differential validity leads to worse lending decisions for protected groups. The risks of ignoring differential validity are evident when lenders satisfy equal outcome measures by randomly approving loans for a protected group, resulting in a misallocation of funds and harms to defaulting borrowers.

Our third contribution is to show that despite fair lending law historically having focused on equality in outcomes, there are statutory and regulatory bases for also prioritizing validity in fair lending. The Equal Credit Opportunity Act (ECOA) aims to ensure that credit is "equally available to all *credit-worthy* customers" (emphasis added). The ECOA provides that credit must not just be allocated equally across groups but also must be correlated with creditworthiness. Lending policies, in other words, must involve equally valid tests for all groups in addition to ensuring equality in outcomes. The Consumer Financial Protection Bureau (CFPB) has also taken stances that emphasize equal validity as an important fair lending outcome: the no-action letter (NAL) provided to Upstart, a FinTech lender discussed later in this Article, required Upstart to assess their model's predictive accuracy, demonstrating the validity of their lending process by customer group.[16]

Yet despite this foundation, lenders may be hesitant to adopt more predictive models, particularly those using machine learning or alternative data, because of concerns that they could increase measured outcome disparities and invite regulatory scrutiny.[17] This reluctance can discourage innovation that might improve assessments of creditworthiness for disadvantaged groups. Incorporating differential validity into compliance frameworks could provide a clearer pathway for the possibility of demonstrating that new models may promote both fairness and accuracy.

Our analysis extends beyond fair lending to other areas of discrimination law, including employment and housing, where disparate impact must increasingly contend with multiple, often conflicting goals. The Supreme Court's recent decision in *Students for Fair Admissions, Inc.*,[18] which

---

[15] *See, e.g.*, a recent Request for Information ("RFI") from the Consumer Financial Protection Bureau, Docket No. CFPB-2020-0026, demonstrating the Bureau's approach to disparate impact analysis under the Equal Credit Opportunity Act emphasized the objective of utilizing innovation to increase access to credit for all consumers (the agency's "approach to disparate impact analysis under ECOA," for example, emphasized a goal of "ensur[ing]… innovation can increase access to credit for *all* consumers.") (emphasis added).

[16] *See* CFPB, No-Action Letter to Upstart (Nov. 30, 2020), https://files.consumerfinance.gov/f/documents/cfpb_upstart-network-inc_no-action-letter_2020-11.pdf (conditioning Upstart's NAL on Upstart's implementation of the Model Risk Assessment Plan (MRAP), requiring Upstart to, *inter alia*, "[t]est Upstart's model and/or variables or groups of variables on a periodic basis for adverse impact and predictive accuracy by group, with results provided to the [CFPB]").

[17] *See* FINREGLAB, EXPLAINABILITY & FAIRNESS IN MACHINE LEARNING FOR CREDIT UNDERWRITING 46 (2023), https://finreglab.org/wp-content/uploads/2023/12/FinRegLab_2023-12-07_Research-Report_Explainability-and-Fairness-in-Machine-Learning-for-Credit-Undewriting_Policy-Analysis.pdf (describing concerns that more predictive models could raise measured disparities and reputational risk, even if they improve predictions for disadvantaged groups).

[18] *Students for Fair Admissions, Inc. v. President & Fellows of Harvard Coll.*, 600 U.S. 181 (2023).

struck down race-conscious university admissions, suggests a shift in the permissible methods to achieve fairness, likely affecting other discrimination domains. More recently, an April 2025 Executive Order directed federal agencies, including the Department of Housing and Urban Development (HUD) and the Consumer Financial Protection Bureau (CFPB), to deprioritize enforcement actions based on disparate impact liability under statutes such as the Fair Housing Act and the Equal Credit Opportunity Act.[19] While we do not engage directly with the implications of the Supreme Court decision or administrative directive for disparate impact doctrine, our argument that fair lending should incorporate differential validity may resonate with a legal landscape that is increasingly skeptical of outcome balancing. Attention to predictive validity can be understood as a call to improve the accuracy of decision-making, which may appeal to critics of outcome-based adjustments (regardless of whether such criticism has merit or not).

We proceed as follows. Part I provides an overview of fair lending disparate impact, demonstrating the regulatory focus on equal outcome notions of disparities. Part II introduces the concept of validity disparities and explains their significance for fair lending. Part III discusses how equal validity can be recognized under the existing legal framework and identifies crucial questions for future work to fully incorporate validity considerations into fair lending enforcement. We then conclude.

I. TRADITIONAL FAIR LENDING'S FOCUS ON EQUAL OUTCOMES

This section provides an overview of disparate impact under the Fair Housing Act (FHA)[20] and ECOA.[21] It begins by discussing disparate impact's burden shifting framework, highlighting how a crucial aspect of both challenging and defending a lending practice is the definition and measurement of lending disparities. We then argue that regulations, regulatory guidance documents, and enforcement actions have nearly exclusively focused on *equal outcome* notions of disparities.

A. Centrality of Disparities Metric to Disparate Impact Analysis

How to measure and what metric is used for estimating disparities is central to fair lending's disparate impact doctrine. This Part argues that traditionally fair lending has adopted a particular measure of disparities—equal outcomes—and overlooked other dimensions such as the validity of a particular lending policy. We start by providing an overview of disparate impact and the burden-shifting framework, highlighting how measuring disparities is central in both establishing a *prima facie* case and analyzing whether there are less discriminatory alternatives. We then argue that fair lending, as reflected in regulatory guidance and fair lending enforcement actions, has primarily focused on the

---

[19] Exec. Order No. 14,127, Restoring Equality of Opportunity and Meritocracy (Apr. 23, 2025), https://www.whitehouse.gov/presidential-actions/2025/04/restoring-equality-of-opportunity-and-meritocracy/. In addition, on May 12, 2025, the CFPB withdrew a 2012 bulletin that recognized disparate impact as evidence of lending discrimination for purposes of ECOA and Reg B, further signaling an administrative shift away from the traditional disparate impact doctrine in the context of fair lending. CFPB, INTERPRETIVE RULES, POLICY STATEMENTS, AND ADVISORY OPINIONS; WITHDRAWAL, 90 FED. REG. 20087 (May 12, 2025), https://www.govinfo.gov/content/pkg/FR-2025-05-12/pdf/2025-08286.pdf.

[20] Fair Housing Act, Pub. L. No. 90-284, 82 Stat. 73 (1968) (codified as amended at 42 U.S.C. §§ 3601–3619 (2018)).

[21] Equal Credit Opportunity Act, Pub. L. No. 93-495, 88 Stat. 1500 (1974) (codified as amended at 15 U.S.C. § 1691 *et seq.* (2018)).

differential-outcomes dimension of disparities—meaning differences in approval rates and loan terms—and not on *differential validity*.

1. Burden Shifting Framework

The two federal statutes that form the core prohibition on discrimination in credit are the FHA and ECOA. FHA, also known as Title VIII of the Civil Rights Act of 1968, protects renters and homebuyers from discrimination by landlords or sellers, covering a range of housing-related conduct and prohibiting discrimination in setting housing-related credit terms on the basis of race, color, religion, sex, disability, familial status, and national origin.[22] ECOA extends this prohibition to all credit transactions,[23] not just those in the context of housing.[24]

Both ECOA and FHA incorporate the doctrines of "disparate treatment" and "disparate impact."[25] Disparate treatment addresses direct conditioning of a credit decision on a protected characteristic, often involving intent to discriminate. Disparate impact, the focus of this Article, covers cases where a facially neutral rule has an impermissible disparate effect on protected groups.

---

[22] In 1988, the Fair Housing Amendments Act was passed, adding sex, disability, and family status as protected groups. *See* Fair Housing Amendments Act of 1988, Pub. L. No. 100-430, 102 Stat. 1619 (codified as amended at 42 U.S.C. § 3601).

[23] *See* 15 U.S.C. § 1691 (a)(1)–(2).

[24] Initially, ECOA only covered discrimination on the basis of sex and marital status. It was amended in 1976 to prohibit discrimination because of race, color, religion, and other grounds. *See* Pub. L. No. 94-239, 90 Stat. 251 (1977) (codified as amended at 15 U.S.C. § 1691). There are other laws that have additional provisions relating to credit pricing discrimination that are not the subject of this Article. The Community Reinvestment Act (CRA), 12 U.S.C. § 2901 (2018), encourages banks and other lenders to address the needs of low-income households within the areas they operate. The CRA does not create a private right of action, instead instructing the relevant supervisory agency on how to oversee that institutions are serving the lending needs of their communities. HMDA, 12 U.S.C. § 2801, requires that certain financial institutions make regular disclosures to the public on mortgage applications and lending. Although HMDA does not contain any explicit discrimination provisions, one of its purposes is to allow the public and regulators to consider whether lenders are treating borrowers differently.

[25] While the texts of ECOA and FHA do not explicitly recognize the two discrimination doctrines, the disparate impact doctrine has long been recognized in credit pricing cases by courts and agencies alike. The Supreme Court recently affirmed that disparate impact claims could be made under FHA in *Texas Dep't of Hous. & Cmty. Affs. v. Inclusive Communities Project, Inc.*, 576 U.S. 519 (2015), confirming the position of eleven appellate courts and various federal agencies including the Department of Housing and Urban Development (HUD), the agency primarily responsible for enforcing FHA. *See* Robert G. Schwemm, *Fair Housing Litigation after Inclusive Communities: What's New and What's Not*, COLUM. L. REV. SIDEBAR 106, 106 (2015) ("The Court's 5-4 decision in the *ICP* case endorsed forty years of practice under the FHA, during which the impact theory of liability had been adopted by all eleven federal appellate courts to consider the matter."). There is not an equivalent Supreme Court case with respect to ECOA, but the Consumer Financial Protection Bureau (CFPB), the agency primarily responsible for enforcing the ECOA, and lower courts have found that the statute allows for a claim of disparate impact. *See, e.g.*, *Ramirez v. GreenPoint Mortgage Funding, Inc.*, 633 F. Supp. 2d 922, 926–27 (N.D. Cal. 2008); *Equal Credit Opportunity Act (ECOA)*, CFPB CONSUMER LAWS AND REGULATIONS (2013), https://files.consumerfinance.gov/f/201306_cfpb_laws-and-regulations_ecoa-combined-june-2013.pdf ("The ECOA has two principal theories of liability: disparate treatment and disparate impact."). *See also* Winnie F. Taylor, *The ECOA and Disparate Impact Theory: A Historical Perspective*, 26 J.L. & POL'Y 575, 600 (2018).

During the first Trump Administration, the CFPB proposed abandoning disparate impact liability under the ECOA. *See* Press Release, Mick Mulvaney, Statement of the Bureau of Consumer Financial Protection on Enactment of S.J. Res. 57 (May 21, 2018) (stating that the CFPB will reexamine its guidance on disparate impact liability under the ECOA). For a skeptical view of whether the statutory language of ECOA supports disparate impact, *see generally* Peter N. Cubita & Michelle Hartmann, *The ECOA Discrimination Proscription and Disparate Impact—Interpreting the Meaning of the Words That Actually Are There*, BUS. LAW. 829 (2005).

A disparate impact case typically follows the three-part burden-shifting framework developed in the Title VII employment discrimination context:[26] (1) the plaintiff must demonstrate that a certain factor disproportionately negatively impacts a protected group;[27] (2) if successful, the creditor must show that the criterion makes the credit evaluation system more predictive or is otherwise justified by a legitimate business need;[28] and (3) the plaintiff then has the opportunity to show that the creditor's legitimate business needs could be met by a less discriminatory alternative.[29]

Historically, due to lack of data, Step 1 was less plaintiff-friendly than today. Regulation B, ECOA's implementing regulation, once prohibited inquiries about an applicant's protected characteristics,[30] preventing plaintiffs from furnishing relevant statistics to show prima facie discrimination.[31] However, the Home Mortgage Disclosure Act (HMDA) was amended in 1989 to mandate that lenders collect and disclose data on race, sex, and income to monitor mortgage approval rates.[32] Ten years later, Regulation B was amended to allow voluntary collection of data on race, color, religion, or national origin for non-mortgage credit, helping plaintiffs meet Step 1.[33]

For Step 2, once a prima facie case of disparities is established, the defendant must prove that the practice is consistent with business necessity.[34] Regulatory guidance on ECOA states that creditors

---

[26] Regulation B [hereinafter Reg B], the regulation implementing ECOA, instructs that employment law jurisprudence forms the basis of disparate impact analysis under ECOA. § 1002.6(a) n.2 ("The legislative history of the act indicates that the Congress intended an 'effects test' concept, as outlined in the employment field by the Supreme Court in the cases of *Griggs v. Duke Power Co.* and *Albemarle Paper Co. v. Moody*, to be applicable to a creditor's determination of creditworthiness.") (citations omitted). Disparate impact first entered American law in the 1971 landmark case *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971) (concerning a legal challenge to hiring requirements of a high school diploma and aptitude test). A formal burden-shifting framework was articulated in the subsequent employment decision *Albemarle Paper Co. v. Moody*, 422 U.S. 405 (1975), which became the three-step burden-shifting approach that is applied today. The burden-shifting framework was codified in Title VII in Section 703(k), added by the Civil Rights Act of 1991. Similar language exists in the 2013 HUD Disparate Impact Rule. *See* 12 C.F.R. § 1002.6 n.2 (2003) (discussing the relevance of Title VII for interpreting fair lending disparate impact); *see also* Equal Credit Opportunity, 41 Fed. Reg. 29,870, 29,874 (proposed July 20, 1976) (codified at 12 C.F.R. § 202) ("Congress intended certain judicial decisions enunciating this 'effects test' from the employment area to be applied in the credit area.").

[27] *A.B. & S. Auto Service, Inc. v. South Shore Bank of Chicago*, 962 F.Supp. 1056, 1060 (N.D.Ill.1997) (citation omitted) (First, a plaintiff needs to demonstrate that the challenged practice "has a significantly greater discriminatory impact on [a protected group].").

[28] *Cherry v. Amoco Oil Co.*, 490 F. Supp. 1026, 1031 (N.D. Ga, Jun. 11, 1980) (Second, if the plaintiff successfully convinces the court that a given policy creates disparities, then "the creditor must show that the criterion makes the credit evaluation system "more predictive than it would be otherwise.").

[29] OFFICIAL STAFF INTERPRETATION, § 1002.6(A)(2), 50 FED. REG. 48,050 (1985) (Third, the plaintiff receives "the opportunity to show that the creditor's legitimate business needs could be met by a less discriminatory alternative.")

[30] 12 C.F.R. § 1002.5. (Reg B categorically proscribed "inquir[ing] about the race, color, religion, national origin, or sex of an applicant or any other person in connection with a credit transaction."); *see also* Matheson, *The Equal Credit Opportunity Act: A Functional Failure*, 21 HARV. J. LEGIS. 371, 382-91 (1984).

[31] *See, e.g.*, *Cherry*, 490 F. Supp. at 1030 (plaintiff did not have adequate data to show that zip code criterion had a disparate impact on African Americans who applied for credit but were rejected).

[32] *See* Home Mortgage Disclosure, 54. Fed. Reg. 51,356, 51,359-60 (1989) (codified at Equal Credit Opportunity Act (Reg B), 12 C.F.R. § 203.4(a)(10) (2011)) (noting the HMDA's amended mandate of collecting and disclosing "data on the race, sex, and income of applicants and borrowers, in addition to the geographic itemization of loans that is currently required.").

[33] 64 Fed. Reg. 44582, 44586 (Aug. 16, 1999).

[34] Susan S. Grover, *The Business Necessity Defense in Disparate Impact Discrimination Cases*, 30 GA. L. REV. 387, 403 n.53 (1996) ("[When] placing a burden of production on the defendant, the defendant is compelled to introduce evidence sufficient

can defend a policy that produces a disparity by showing a demonstrable relationship between the challenged policy and creditworthiness.[35] Some have argued that this relationship between the policy and creditworthiness must be based in the borrower's likelihood of repaying the loan, rather than the profitability of the loan more broadly.[36] However, the justification could include some appeals to profitability if they are "manifest" and not "hypothetical" or "speculative."[37]

Assuming such necessity is established, next is Step 3: the challenged policy is still prohibited if there is a "less discriminatory alternative" (LDA) that also accomplishes the business necessity. Shortly after *Griggs*, the Supreme Court held in *Albemarle Paper Co. v. Moody*, that "[i]f an employer does then meet the burden of proving that its tests are 'job related,' it remains open to the complaining party to show that other tests or selection devices, without a similarly undesirable racial effect, would also serve the employer's legitimate interest in 'efficient and trustworthy workmanship.'"[38] The requirement to find an alternative practice was referred to as "other tests or selection devices, without a similarly undesirable racial effect"[39] in the *Albemarle* case.[40]

In general, there is very little guidance on the LDA stage of disparate impact.[41] The Commentary to Regulation B provides that a practice that meets a legitimate business need is permissible if it "cannot be reasonably be achieved as well by means that are less disparate in their

---

to permit an inference of the fact it is attempting to prove. This is a lesser burden than the burden of persuasion, which means that if the defendant fails to prove the existence of the fact at issue, the plaintiff immediately prevails.").

[35] OFFICIAL STAFF COMMENTARY TO REGULATION B, 12 C.F.R. § 1002.6(A)-2.

[36] *See* Talia Gillis, *"Price Discrimination" Discrimination* (2024). *See, e.g.*, *Walker v. Bank of Am. Corp.*, No. 8:18-CV-02466-PWG, 2019 WL 3766824 at *3 (D. Md. Aug. 8, 2019).

[37] POLICY STATEMENT ON DISCRIMINATION IN LENDING, 59 FR 18266-01, 18269-70.

[38] *Albemarle*, 422 U.S. at 425 (1975) (quoting *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 801 (1973)).

[39] *Albemarle*, 422 U.S. at 425; *accord, e.g.*, *Jones*, 752 F.3d at 53 (citing *Albemarle* to interpret the 1991 Act's text); *see also Allen v. City of Chicago*, 351 F.3d 306, 312 (7th Cir. 2003) (same using a "see also" signal).

[40] The 1991 Act codified this burden-shifting regime as the "alternative employment practice" requirement. 42 U.S.C. § 2000e-2(k)(1)(A) (2012). Congress did not define the phrase, and its substantive meaning remains uncertain. The LDA test has not always been treated as a separate step. *See, e.g.*, *Wards Cove Packing Co. v. Atonio*, 490 U.S. 642, 659 (1989) (treating the alternative employment practice test as part of the "business justification" phase); *see also Dothard v. Rawlinson*, 433 U.S. 321, 332 (1977) (treating the alternative employment practice test as a narrow tailoring requirement for the business necessity defense).

[41] ECOA cases "rarely have discussed, much less reached, the 'third prong' of less discriminatory alternatives analysis." Peter E. Mahoney, *The End(s) of Disparate Impact*, 47 EMORY L.J. 409, 490 (1998). This dearth of case law is especially unfortunate given recent scholarship arguing that, for nearly any underwriting model, an LDA likely exists. *See* Emily Black et al., *Less Discriminatory Algorithms* (2023) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4590481. In the context of fair lending, *see* Implementation of the Fair Housing Act's Discriminatory Effects Standard, 78 Fed. Reg. 11,460, 11,470 (Feb. 15, 2013) (codified at 24 C.F.R. § 100) ("If a practice has a discriminatory effect, it may still be lawful if supported by a legally sufficient justification, which includes proving that no less discriminatory alternative exists to achieve the same business objective.").

impact."[42] In the context of creditworthiness models, an LDA must perform at or close to the level of the challenged practice.[43]

Although traditionally considered the plaintiff's burden,[44] regulators have recently suggested that lenders must proactively search for LDAs. CFPB's Associate Director of Fair Lending and Equal Opportunity, Patrice Ficklin,[45] and recent CFPB supervisory updates[46] have emphasized lenders' responsibility to demonstrate their search for alternative models.

In conclusion, measuring disparities is central to a disparate impact claim and defense. While showing disparities is essential for challenging lending policies initially, consideration of disparities is also required for defending the policy, especially under recent interpretations of the LDA requirement. The next section examines the equal-outcomes metrics regulators and courts have used to test for lending disparities.

2. Disparity Metrics for Disparate Impact Analysis

Given the centrality of disparity analysis in disparate impact claims, both at the initial and final stages of the burden shifting framework, the metric used to estimate disparities is crucial. As we will argue below, the current focus on *equal outcomes* is misguided, as it overlooks other significant disparities, specifically *validity* disparities. In this section, we provide an overview of outcome equality measures and how they differ from validity measures.

We use the term "equal outcomes" to refer to a situation in which lending decisions are similar on average across different groups and the term "outcome inequality" to refer to a situation where lending decisions are not similar across different groups. Equality of outcomes can either deal with the extensive margin of lending—meaning whether a loan application is approved or denied—or with the intensive margin of lending—meaning the terms of a loan contract, conditional on receiving a loan. These loan terms are often the interest rate of the loan, fees such as loan closing fees, or the amount of credit extended. Unequal outcomes in this setting mean that either approval rates or loan terms are different for a protected group and a reference group.

In the context of fair lending, outcome equality metrics are often referred to as the adverse-impact-ratio (AIR), in the case of loan approval rates, and standardized-mean-differences (SMD), for interest rates, fees, or credit limits. The AIR compares the rate of approval for different demographic

---

[42] *See* 12 C.F.R. § 1002.6(a), Supplement I to Part 202, Comment 6(a)-2 (1994) (A disparate output yields liability "unless the creditor practice meets a legitimate business need that cannot reasonably be achieved as well by means that are less disparate in their impact.").

[43] *See* Mahoney, *supra* note 41, at 490. Reg B requires credit scoring systems to be "[d]eveloped for the purpose of evaluating the creditworthiness of applicants with respect to the legitimate business interests of the creditor utilizing the system (including, but not limited to, minimizing bad debt losses and operating expenses in accordance with the creditor's business judgment)." *See also* 12 C.F.R. § 1002.2(p)(1)(ii) (1997).

[44] *See* Michael Selmi, *Algorithms, Discrimination and the Law*, 82 OHIO STATE L. J. 406, 644 (2020) (discussing whether the burden exceeds intention of the law and imposes undue burden on the plaintiff.)

[45] Practical Law Finance, *CFPB Clarifies Duty to Perform Fairness Testing on Lending Models*, WESTLAW (Apr. 19, 2023), https://content.next.westlaw.com/practical-law/document/I7770d7c1da5611ed8921fbef1a541940/CFPB-Clarifies-Duty-to-Perform-Fairness-Testing-on-Lending-Models?transitionType=Default&contextData=%28sc.Default%29.

[46] *See* CFPB, FAIR LENDING REPORT OF THE CONSUMER FINANCIAL PROTECTION BUREAU (June 2024), https://files.consumerfinance.gov/f/documents/cfpb_fair-lending-report_fy-2023.pdf.

groups, ensuring that one group is not disproportionately disadvantaged. For instance, if the approval rate for minority applicants is significantly lower than that for non-minority applicants, this would indicate a potential disparity under disparate impact using the equal outcomes metric. On the other hand, SMD measures the differences in the mean values of continuous variables, such as interest rates, across different groups. This helps in identifying whether certain groups are being systematically charged higher fees or interest rates.

Both types of outcome equality—AIR and SMD—can be measured unconditionally or conditionally. An unconditional measurement of AIR and SMD looks at the raw differences of approval rates or interest rates without controlling for any characteristics of the borrowers. A conditional measurement accounts for factors that may be relevant to differences in loan eligibility and creditworthiness, such as income, loan amount, or credit history. Accounting for these differences may be important when there are underlying differences in the applicant pool among demographic groups.[47] On the other hand, controlling for applicant characteristics may mask the true disparate impact of policies such that a raw, unconditional measure is preferrable.[48]

Equal outcome measures of disparity can be contrasted with validity measures of lending decisions. Instead of asking whether lending approvals are equal across groups, validity measures evaluate if the extent to which loan outcomes align with the predicted probabilities of default are different among demographic groups. For example, a validity measure might examine whether the ultimate default rates on loans are consistent with the predicted probabilities of default used in lending decisions; *differential* validity arises when validity is greater for one demographic group than another. Below we argue that traditional disparate impact frameworks focus on an equal-outcome notion of fairness, overlooking differential validity, which we discuss in more detail below in Section II.A.1.

B. The Singular Focus on Equal Outcomes

In this section, we discuss that while there is a lack of explicit regulatory guidance and clear case law on the disparity metrics used in disparate impact analysis, lender practice and regulatory enforcement have focused on equal-outcome disparity metrics. Traditionally, lenders have assessed disparate impact using AIR, calculated by dividing the loan approval rate for the protected group by that of the reference group and SMD, comparing and standardizing the interest rates between groups.[49] As discussed above, both metrics are equal-outcome measures and are often estimated without accounting for differences in the baseline applicant pools.[50] We will examine legal and

---

[47] Richard Pace, *Fool's Gold?, Assessing the Case For Algorithmic Debiasing*, PACE ANALYTICS CONSULTING LLC (Jan. 27, 2023), https://www.paceanalyticsllc.com/post/fools-gold-algorithmic-debiasing.

[48] RELMAN COLFAX PLLC, FAIR LENDING MONITORSHIP OF UPSTART NETWORK'S LENDING MODEL, FIRST REPORT OF THE INDEPENDENT MONITOR at 10 (2021), https://www.relmanlaw.com/media/cases/1088_Upstart%20Initial%20Report%20-%20Final.pdf.

[49] *See id.* (discussing the use of AIR and SMD for the first stage of disparate impact and arguing that the measures should not control for other factors); FINREGLAB (2023), *supra* note 18; *see also* Nicholas Schmidt & Bryce Stephens, *An Introduction to Artificial Intelligence and Solutions to the Problems of Algorithmic Discrimination*, arXiv:1911.05755 143 (2019), https://arxiv.org/abs/1911.05755; Patrick Hall et al., *A United States Fair Lending Perspective on Machine Learning*, 4 FRONTIERS IN ARTIFICIAL INTELLIGENCE 3 (2021).

[50] Both the measure of AIR and SMD could be calculated in a way that accounts for applicant pool differences. For example, the AIR could be calculated separately by credit score bins. *See* practical guidance provided to lenders in *Charles River Associates*, at 3 ("The AIR can be calculated for any value of the score and then plotted against a potential score threshold. This can be useful for understanding the impact of the score throughout its distribution . . . .").

regulatory sources related to disparity measures and demonstrate their alignment with the current emphasis on equal outcomes.

1. ECOA and FHA Regulations and Guidance Documents

Despite the general language used in regulatory and guidance documents, the focus in practice remains on equal-outcome measures of disparities, such as loan denial rates, when assessing disparate impact in fair lending practices.

Neither ECOA nor FHA make explicit reference to disparate impact and so the details of the requisite analysis are found in the implementing regulations and other regulatory documents. HUD implemented FHA through its 2013 Disparate Impact Rule, which was partially motivated by existing variation in the application of the disparate impact doctrine in case law.[51] The Rule defines disparities generally by stating that a "practice has a discriminatory effect where it actually or predictably results in a disparate impact on a group of persons."[52] One of the Rule's illustrations of discrimination in a real-estate transaction refers to "denial rate" differences as a form of discrimination.[53] The conclusion is that, while the language in the Rule is general and may potentially include other disparity metrics, its explicit reference is to outcome equality only—denial rates. This view of disparities is consistent with the primary disparity metric used under Title VII, employment discrimination, which often discusses employment selection rates.[54]

Regulation B, the implementing regulation of ECOA, adopts the "effects test" from employment law.[55] The Official Interpretation only makes a general reference to a practice that is "discriminatory in effect because it has a disproportionately negative impact on a prohibited basis."[56] This proposition is illustrated by an example of an income requirement for an overdraft line that leads to women and minority applicants being "rejected at a higher rate than men and nonminority applicants," thereby using an equal outcome disparity metric.[57] Similarly, the 1994 Joint Statement on Fair Lending provides an example of disparate impact in which a minimum loan size cutoff excludes minority applicants, using an equal-outcome measure of disparities.

More recent regulatory documents also focus on equal-outcome measures of disparities. The CFPB periodically publishes summaries of its regulatory work that can shed light on its measurement of disparities. These reports reflect the CFPB's focus on denial rate disparities. For example, in its Fall

---

[51] *See* 24 C.F.R. § 100 (2013) ("Through four decades of case-by-case application of the Fair Housing Act's discriminatory effects standard by HUD and the courts, a small degree of variation has developed in the methodology of proving a claim of discriminatory effects liability."). The 2013 Disparate Impact Rule was reinstated in 2023 following an attempt in 2020 to alter the rule. *See* Reinstatement of HUD's Discriminatory Effects Standard, 88 Fed. Reg. 19,450, 19,489 (Mar. 31, 2023) (to be codified at 24 C.F.R. § 100).

[52] 24 C.F.R. § 100.500(a).

[53] *Id.* § 100.120(b)(2).

[54] *See* 29 C.F.R. § 1607.3 (1978) ("Discrimination defined: Relationship between use of selection procedures and discrimination.").

[55] 12 C.F.R. § 1002.6(a).

[56] *Id.*

[57] *Id.*

2015 report, the Bureau explained that underwriting reviews evaluate "potential disparities in denial rates," which often looks at the ratio of loan application denials probabilities between groups.[58]

These examples demonstrate that while the official language used in regulation and guidance documents does not make any reference to a particular disparity measure and uses general language of "discriminatory effect," the examples and illustrations focus on disparities in loan denial rates—an equal-outcome metric.

2. Enforcement Actions

There are a number of fair lending enforcement actions that have challenged lending practices arguing that they caused impermissible disparate impact. [59] In considering the disparity metric, these cases focus exclusively on equality of outcomes, by considering disparities in loan approvals, interest rates, and fees. Below we provide several examples of consent orders and settlements where the description of relevant disparities makes clear this focus on equal-outcome metrics.

In a group of mortgage lending cases in the wake of the 2007–2008 financial crisis, lenders were challenged for lending disparities that were deemed discriminatory.[60] A typical example is the 2014 consent order resulting from a challenge of National City Bank's practice of setting lending "base-rates"—the interest rate determined by borrower creditworthiness, also known as "par rates"—and then giving loan officers and brokers discretion to deviate from those rates. According to the consent order, the CFPB found that "[National City Bank's] policies and practices result[ed] in African-American and Hispanic Borrowers paying higher interest rates, fees, and other costs than similarly-situated non-Hispanic White borrowers."[61] A similar complaint against Bancorp South found that the bank had "discriminated against African-American applicants by denying their mortgage loan applications more often than similarly situated non-Hispanic White ("White") applicants; and discriminated against African-American applicants by charging them higher prices on their mortgage

[58] CFPB, SUPERVISORY HIGHLIGHTS (FALL 2015), https://files.consumerfinance.gov/f/201510_cfpb_supervisory-highlights.pdf. The Report also discusses the use of marginal effects analysis, which is a different way to consider differences in denial rates.

[59] Enforcement of ECOA and FHA—and fair lending law more generally—is spread across several agencies. With the creation of the CFPB in 2011, the Bureau assumed enforcement responsibility over ECOA with respect to entities within its jurisdiction. This loosely covers institutions like banks and lending companies. For a full discussion of the various CFPB authorities and institutions that they cover, *see* Adam J. Levitin, *The Consumer Financial Protection Bureau: An Introduction*, 32 REV. BANKING & FIN. L. 321, 343 (2013). Other federal agencies share enforcement authority with respect to institutions over which they have supervisory authority. With respect to FHA, HUD shares enforcement authority with the DOJ. *See* Jonathan Zasloff, *The Secret History of the Fair Housing Act*, 53 HARV. J. ON LEGIS. 247, 250 (2016). Although both HUD and the DOJ can bring enforcement action, they are subject to different statutes of limitations. On the enforcement of FHA, *see generally* ADAM LEVITIN, CONSUMER FINANCE: MARKETS AND REGULATION 455 (2d ed. 2023). My focus here is primarily on disparate impact claims, although there have been enforcement action cases that allege discriminatory conduct more directly. *See, e.g.*, Complaint at 3 ¶ 21, *United States v. Long Beach Mortg. Co.*, No. 96-CV-6159 (C.D. Cal. Sept. 5, 1996).

[60] There were a number of enforcement actions that predated the financial crisis. *See, e.g.*, *United States v. Huntington Mortg. Co.*, Case No. 1:95-cv-02211 (N.D. Ohio Date, 1995); Complaint, *Long Beach Mortg. Co.*, No. 96-CV-6159 (C.D. Cal. 1996). The vast majority of cases that challenge discretionary markups, however, followed the financial crisis.

[61] *See* Consent Order at 3, *United States v. Nat'l City Bank*, No. 13-1817 (W.D. Pa. Jan. 9, 2014).

loans than similarly situated White applicants."[62] In other cases dealing with pre-financial crisis lending, complaints similarly focus on disparities in approval rates and loan costs, an equal-outcome metric.[63]

A second group of cases involved the auto lending industry, at a time after the financial crisis when discretionary markup policies were not yet prohibited.[64] These cases similarly focus on equal-outcome metrics for the purposes of establishing that the lending policies created disparities. The case against Ally Financial in 2013, for example, noted that the "CFPB's and the DOJ's markup analyses focused on the interest rate difference between each borrower's contract rate and Ally's buy rate,"[65] thereby focusing on unequal outcomes in the markup component of pricing. Other cases against Honda in 2015,[66] Fifth Third Bank in 2015,[67] and Toyota in 2016[68] contain similar language. Several

---

[62] *See* Consent Order at 2, *United States v. BancorpSouth Bank*, No. 16-118 (N.D. Miss. July 25, 2016).

[63] *United States v. Provident Funding Associates*, No. 15-2373 (N.D. Cal. May 28, 2015); *see also* Complaint at 8; *United States v. Countrywide Fin. Corp.*, No. 11-10540 (C.D. Cal. Dec. 21, 2011); Consent Order at 3, *United States v. JPMorgan Chase Bank, N.A.*, No. 1:17-CV-00347-AJN (S.D.N.Y. Jan. 20, 2017); Consent Order at 4; *United States v. Primelending*, No. 3:10-CV-2494-P (N.D. Tex. Jan. 11, 2011).

[64] These were a series of cases following the now-invalidated 2013 CFPB Bulletin, the CFPB challenged auto financers' markup and compensation policies. The auto lending industry had been under significant scrutiny in the years leading up to the 2013 CFPB Bulletin. Early studies by Professor Ian Ayres demonstrated how discretionary dealer markups disproportionately impacted racial minorities. *See* Ian Ayres, *Fair Driving: Gender and Race Discrimination in Retail Car Negotiations*, 104 Harv. L. Rev. 817-872 (1991); Ian Ayres, *Further Evidence of Discrimination in New Car Negotiations and Estimates of Its Cause*, 94 MICH. L. REV. 109 (1995); Ian Ayres & Peter Siegelman (1995), *Race and Gender Discrimination in Bargaining for a New* Car, 85 AM. ECON. REV. (1995). These audit studies were supported by a large observational study in 2003. *See* Mark A. Cohen, *Imperfect Competition in Auto Lending: Subjective Markup, Racial Disparity, and Class Action Litigation*, 8 REV. L & ECON. 21 (2012). This research resulted in a series of cases between 2003 and 2007 in which the National Consumer Law Center settled several class actions. *See* NAT'L CONSUMER L. CTR., AUTO ADD-ONS ADD UP: HOW DEALER DISCRETION DRIVES EXCESSIVE, ARBITRARY, AND DISCRIMINATORY PRICING (Oct. 2007), https://filearchive.nclc.org/car_sales/report-auto-add-on.pdf. Ian Ayres, Expert Report, Willis et al. v. American Honda Finance Corp. No. 3-02-0490 (M.D. Tenn. Jun. 30, 2004).

[65] *Ally Financial Inc.*, CFPB No. 2013-0010 (Dec. 19, 2013). Consent Order at 6; *United States v. Pacifico Ford, Inc.*, No. 2:07-CV-3470(PBT) (E.D. Pa. Sept. 4, 2007). According to the consent order, Ally used its underwriting model to determine the base-rate—the minimum interest rate at which it would finance or purchase an installment contact, also known as the "buy rate." *Id.* Ally would then allow dealers discretion in setting the markup interest rate above the base-rate and would compensate dealers for the spread. *Id.* CFPB analysis of the markup rates revealed that Black borrowers were paying on average 29 basis points more, Hispanic borrowers were paying 20 basis points more, and Asian and Pacific Islanders were paying 22 basis points more than white borrowers with similar base-rates. *Id.* The consent order explicitly states that the markup pricing disparities were not based on risk-based considerations like "creditworthiness or other objective criteria related to borrower risk," concluding that "Ally's specific policy and practice are not justified by legitimate business need …." *Id.* at 8.

[66] *In the Matter of: American Honda Finance Corporation*, CFPB No. 2015-0014 (July 14, 2015). The CFPB found that Honda allowed dealers to increase a consumer's interest rate above Honda's own "established risk-based buy rate." This resulted in a dealer markup that was 36 basis points higher for Black borrowers than similarly situated non-Hispanic white borrowers—with the difference not "based on creditworthiness or other objective criteria related to borrower risk."

[67] *In the Matter of: Fifth Third Bank*, CFPB No. 2015-0024 (Sept. 28, 2015). Black borrowers were charged 35 basis points more, and Hispanic borrowers were charged 36 basis points more, in dealer markup than similarly situated non-Hispanic white borrowers for retail installment contracts.

[68] *In the Matter of: Toyota Motor Credit Corp.*, CFPB No. 2016-0002 (Feb. 2, 2016). There were a number of enforcement actions against auto lenders that predate the CFPB. *See, e.g.*, Partial Consent Decree, *United States v. Nara Bank*, No. 2:09-CV-7124 (RGK)(JC) (C.D. Cal. Nov. 18, 2009). Toyota was accused of allowing auto dealers to offer discretionary interest rates above its buy rate based on "individual borrowers' creditworthiness and other objective criteria related to borrower risk." This discretion resulted in higher rates charged to racial minorities. The CFPB examination revealed that, on average, Black borrowers were charged 27 basis points more, and AAPI borrowers were charged 18 basis points more, than

more recent cases brought by the Federal Trade Commission against auto dealers[69] use similar disparity metrices by claiming that racial minorities were charged different costs, fees and interest rates.[70]

3. Private Litigation

Courts have similarly focused on equal-outcome metrics for determining a finding of disparities. In *Miller v. Countrywide Bank, N.A.*,[71] for example, the District of Massachusetts court found that "African-American borrowers are charged higher fees and rates than similarly situated white borrowers,"[72] focusing on the outcome equality of fees and interest rates as the disparity metric. Disparate impact private litigation, both for auto lending[73] and mortgage lending,[74] has similarly focused on pricing disparities.

In short, while official regulation and guidance documents refer generally to discriminatory impact and hint to equal-outcomes in the examples they provide, enforcement actions and private litigation focus more explicitly and exclusively on equal-outcome disparity metrics.

---

similarly-situated non-Hispanic white borrowers. Again, the CFPB concluded that Toyota's "policy and practice" of compensating dealers from markup revenue was not justified by legitimate business purpose.

[69] Section 1029 of Dodd-Frank excluded auto dealers from the CFPB's direct oversight. Thus, the cases brought by the CFPB deal with indirect auto lending activity.

[70] *See* Complaint for Permanent Injunction and Other Equitable Relief, *Fed. Trade Comm'n v. Liberty Chevrolet, Inc.*, No. 20-CV-3945 (S.D.N.Y. May 21, 2020). *See* Complaint at 12, *Liberty Chevrolet, Inc.*, No. 1:20-CV-3945 (S.D.N.Y. May 21, 2020). The case ultimately settled the following year. *See* Stipulated Order for Permanent Injunction and Other Equitable Relief, *Liberty Chevrolet, Inc.*, No. 1:20-CV-3945 (S.D.N.Y. May 22, 2020), ECF Nos. 9, 10. *See* Complaint for Permanent Injunction, Monetary Relief, and Other Relief, *Fed. Trade Comm'n v. N. Am. Auto. Servs.*, No. 1:22-cv-01690 (N.D. Ill. Mar. 31, 2022). *See* Complaint for Permanent Injunction, Monetary Relief, and Other Relief, *Fed. Trade Comm'n v. Passport Auto. Grp.*, No. 8:22-cv-02670-GLS (D. Md. Oct. 18, 2022).

[71] 571 F. Supp. 2d 251, 253 (D. Mass. 2008).

[72] *Id.* at 259.

[73] *See, e.g.*, *Willis v. Am. Honda Fin. Corp.*, No. 3-02-0490 (M.D. Tenn. Nov. 8, 2004); *Claybrooks v. Primus Auto. Fin. Serv., Inc.*, 363 F. Supp. 2d 969 (M.D. Tenn. 2005). For further discussion of these cases, *see* Cohen, *supra* note 64, at 49. Between 2003 and 2007, the National Consumer Law Center (NCLC) settled several auto finance cases. A list of these cases and settlements can be found at NAT'L CONSUMER L. CTR, CASE INDEX – CLOSED CASES, https://nclc-old.ogosense.net/litigation/case-index-closed-cases.html.

[74] *See, e.g.*, *Miller v. Countrywide Bank, N.A.*, 571 F. Supp. 2d 251 (D. Mass. 2008); *Ramirez v. GreenPoint Mortg. Funding, Inc.*, 633 F. Supp. 2d 922 (N.D. Cal. 2008); *Ware v. Indymac Bank*, 534 F. Supp. 2d 835 (N.D. Ill. 2008); *Zamudio v. HSBC North America Holdings, Inc.*, No. 07-C-4315, 2008 WL 517138 (N.D. Ill. Feb. 20, 2008); *Martinez v. Freedom Mortg. Team, Inc.*, 527 F. Supp. 2d 827 (N.D. Ill. 2007); *Newman v. Apex Financial Group, Inc.*, No. 07 C 4475, 2008 WL 130924 (N.D. Ill. Jan. 11, 2008); *Jackson v. Novastar Mortg., Inc.*, 645 F. Supp. 2d 636 (W.D. Tenn. 2007); *Guerra v. GMAC, LLC.*, 2009 WL 449153 (E.D. Pa. Feb. 20, 2009) (denying a motion to dismiss); *Taylor v. Accredited Home Lenders, Inc.*, 580 F. Supp. 2d 1062 (S.D. Cal. 2008) (denying a motion to dismiss); *Garcia v. Countrywide Fin. Corp.*, No. 07-1161 (C.D. Cal. Jan. 15, 2008) (denying motion to dismiss, including claims of discretionary markups).

II. EXPANDING DISPARITIES METRIC TO INCLUDE EQUAL VALIDITY

This section introduces the idea of *validity* disparities and argues for why they are important and consistent with the goals of fair lending.[75] We then discuss how the disparity measures of equal outcomes and validity may be in tension with one another and that reducing disparities on one dimension may come at a cost to the other dimension, mirroring recent work in the algorithmic fairness literature which has demonstrated the impossibility of satisfying multiple fairness conditions in many circumstances.[76] The section then presents a simulation exercise in which these tradeoffs are demonstrated. We end by proposing that the two disparity metrics be balanced and jointly considered when determining which lending practices are permissible.

A. What is Differential Validity?

Consider a hypothetical lender that uses highly accurate risk scores to evaluate one group of loan applicants, while for another group of applicants the lender uses coin flips to decide who is approved and who is denied. Even if the two groups have equal approval rates—for example, if the lender approves half of the applicants in each group[77]—this underwriting process is evidently unfair to the second group, whose access to credit is decided by a mere game of chance. [78] *Validity* formalizes this notion of fairness.

Formally, validity can be defined as the extent to which the decisions in a lending policy correlate with the characteristics or outcomes that determine a successful loan, such as later repayment. This is an intentionally modular definition: the decisions in question can comprise loan approvals, interest rates, and other loan terms; loan "success" can incorporate other outcomes besides just loan

---

[75] Our focus is on disparate impact approaches that treat disparity itself—as measured by some metric—as sufficient to trigger scrutiny under the doctrine, rather than approaches that require a judgment about the source or cause of the disparity. For example, Prince and Schwarcz emphasize the importance of assessing whether a facially neutral variable is predictive because it serves as a proxy for a protected characteristic. *See* Anya E.R. Prince & Daniel Schwarcz, *Proxy Discrimination in the Age of Artificial Intelligence and Big Data*, 105 IOWA L. REV. 1257 (2020). For a skeptical perspective on statistical methods that aim to identify and neutralize the use of proxies for protected characteristics, *see* Talia Gillis, *Orthogonalizing Inputs*, ACM Symposium in Computer Science and Law (2024).

[76] These notions include calibration, "balance for the positive class," "balance for the negative class," etc. *See* Kleinberg et al., *supra* note 6, for a discussion of how these three are incompatible with each other and, typically, with outcome equality as well.

[77] Many algorithmic fairness approaches, like adversarial de-biasing have a similar result of equalizing approval rates, potentially at the cost of validity.

[78] Importantly, for this metaphor, we assume that the coinflips were collected correctly (e.g., a proverbial "heads" was not counted as a proverbial "tails") and solely focus on the issues with using coinflips in the first place to predict the unrelated property of creditworthiness. In stipulating that any underlying data be correctly collected, we also bracket the issue of whether credit scores can, by their nature, be "true" or "false" and in this sense more or less "accurate." This question is especially important when courts interpret the Fair Credit Reporting Act (FCRA), a federal statute mandating "reasonable steps" to ensure "maximum possible accuracy" in credit reporting.

Courts have held that FCRA does not regulate the accuracy of credit *scores*, specifically, because algorithmically generated credit scores are mere opinions about the probability of default and thus cannot be true or false in the first place. *See, e.g.*, *Jefferson Cnty. Sch. Dist. No. R–1 v. Moody's Investor's Servs., Inc.*, 175 F.3d 848, 854–56 (10th Cir. 1999) (affirming dismissal on grounds that credit ratings "did not contain a provably false factual connotation" and such evaluation "could well depend on [assessing] a myriad of factors"); *see also Plumbers' Union Local No. 12 Pension Fund*, 632 F.3d at 775–77 ("The ratings are opinions purportedly expressing the agencies' professional judgment about the value and prospects of the certificates.").

repayment; and the correlation at the heart of this definition can be quantified as needed in various applications, for example as a regression coefficient, an error rate, or a rank correlation. For example, as discussed at length in the algorithmic fairness literature, the validity of a prediction can be measured in terms of balance for the positive class—among those who actually experience loan "success," prediction scores are similarly distributed across groups—and balance for the negative class—among those who do not experience loan "success," prediction scores are similarly distributed across groups—among other metrics.[79]

To keep our discussion simple and concrete, we focus hereafter on a notion of validity that quantifies the relationship between loan approval decision and true loan repayment outcome: specifically, a lender's approval rate *among* the applicants who would repay a loan.[80] This type of equality—that is, equal approval rates in the set of applicants who would ultimately repay—has been termed in the algorithmic fairness literature "equality of opportunity."[81]

In the example of the lender that underwrites one group using a highly accurate risk score that precisely identifies borrowers who will repay a loan, and another group using coin flips, validity readily captures what is intuitively unfair. In the first group, the highly accurate risk score allows the lender to provide loans to precisely the applicants that will ultimately repay the loan. In the second group, only half of the borrowers who could ultimately repay the loan receive it. Therefore, the lender's policy is more valid for one group than for another.

Such cases of disparities in validity can also be termed *differential validity*. If creditworthiness is understood in terms of later loan repayment, then our particularized definition that focuses on repayment and approval rates maps closely to ECOA's mandate that credit be "equally available to all *credit-worthy* customers"; statutorily, ECOA calls for minimizing differential validity.[82]

---

79 For further discussion of these definitions, see papers mentioned in *supra* note 13.

80 Other particularizations of our definition of validity could similarly focus on the extent to which a lending policy avoids approving loans for borrowers who ultimately default. This notion of fairness also has clear intuitive appeal, especially in view of how mortgage foreclosures have historically exacerbated racial wealth disparities in the US. *See* Amir Kermani & Francis Wong, *Racial Disparities in Housing Returns* (NBER Working Paper No. 29306, 2021), https://www.nber.org/papers/w29306.

81 Moritz Hardt et al., *Equality of Opportunity in Supervised Learning*, ARXIV:1610.02413 (2016), https://arxiv.org/abs/1610.02413 (proposing a fairness criterion that requires equal true positive rates across demographic groups to ensure that predictions are not biased against any particular group). The term "equalized odds" has also been used to refer to equality of opportunity. While many determinants of loan repayment may be outside of an individual's control, conceptually it may be helpful to understand the "equality of opportunity" label as stemming from a premise that individual effort determines repayment. Under this premise, equality of opportunity comes from guaranteeing equal outcomes for individuals who end up able to repay, rather than for all individuals unconditionally. For a summary of the different fairness metrics discussed in the algorithmic fairness literature *see* FINREGLAB, THE USE OF MACHINE LEARNING FOR CREDIT UNDERWRITING 126-130 (2021) (providing an overview of 21 fairness metrics, of which "equal opportunity" is one). *See also* Pessach & Shmueli, *supra* note 13, and Verma & Rubin, *supra* note 13. In the legal context, *see* Deborah Hellman, *Measuring Algorithmic Fairness*, 106 VIRGINIA L. REV. 811 (2020) (arguing that a measure of the ratio between false positives and false negatives is a normatively meaningful measure that suggests unfairness).

82 We use the term "creditworthy" only in the technical sense of denoting future repayment, even though the term carries fraught moral connotations that we categorically disown. *See generally* JOSH LAUER, CREDITWORTHY: A HISTORY OF CONSUMER SURVEILLANCE AND FINANCIAL IDENTITY IN AMERICA (2017).

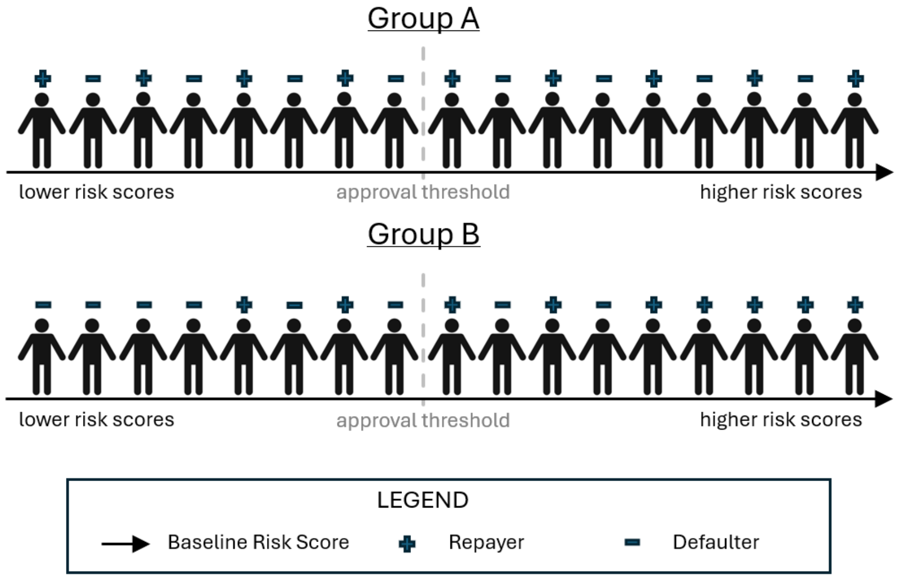

Figure 1: Differential Validity

Figure 1 provides a graphical depiction of differential validity.[83] Two groups of loan applicants, denoted A and B, are ranked by a lender's lending policy; higher ranked applicants are farther to the right, and those to the right of the vertical dashed line are approved for a loan. To demonstrate the concept of validity, we assume that we know the underlying truth about the applicant type—whether they are a loan defaulter or a repayer. The "+" icons denote which applicants would repay a loan if approved, and the "-" icons denote which applicants would default. In this graphical example, the lending policy has greater validity for Group B than for Group A: more applicants who would repay the loan are approved in Group B than in Group A, even as the overall approval rate and the number of applicants who would repay are the same in both groups. Indeed, because the lending policy for Group A appears to sort applicants in a way completely uncorrelated from their ultimate repayment potential, the lending policy in Group A is tantamount to lending by coin flip.

While lenders of course do not use coinflips in practice, the presence of statistical noise in credit scoring means that most modern underwriting relies on inputs that can be, in whole or in part, tantamount to coinflips. Recent research points to modern credit scores being more like a coinflip for members of racial and ethnic minorities than for majority group members, principally because of disparities in the length or richness of individuals' credit histories.[84] Reliance on underwriting inputs that are statistically noisier for some protected class subgroups has both immediate costs, in the sense of inducing disparities in how much credit flows to individuals that plausibly have the most productive use for a loan, and dynamic costs, in the sense of further contributing to noise in the data used to train the credit scoring models of the future.[85] These concerns echo those raised in other contexts where

---

[83] *See also* MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM Design 77 (2020) (providing a similar diagrammatic representation).

[84] *See* Blattner & Nelson, *supra* note 14.

[85] *Id.*

the *misallocation* of opportunities like loans, education, or jobs is seen as holding first-order importance.[86]

Depending on how our definition is particularized, measuring differential validity may be understood counterfactually—the applicants *would not have* defaulted if they had been granted the loan. We discuss such measurement questions in more detail in Section III.B.2.

The example above also demonstrates how overlooking differential validity, while focusing solely on increasing outcome equality, may expose lenders to other legal risks. Credit and model safety and soundness requirements focus on principled lending decisions that are violated when outcome equality is achieved randomly. Additionally, ability-to-repay requirements, which require lenders to assess borrower ability to pay before lending, are violated when coin-toss lending takes place.[87]

B. Significance of Differential Validity

Overlooking differential validity and focusing solely on equal-outcome measures of disparities can lead to lending decisions that are harmful for borrowers and lenders.[88] This is because focusing on equal outcomes, without considering differential validity, may lead to lending decisions that either achieve disparity through reducing the lending threshold for a protected group or by selecting borrowers in a relatively indiscriminate way from a protected group to achieve outcome equality.

Unequal outcomes and unequal validity cause distinct harms to the consumer. The harms from unequal outcomes are familiar: a disadvantaged group faces a greater likelihood of being denied a loan or of paying higher interest rates. The harms from differential validity are more subtle. When a screening technology such as a credit score exhibits differential validity across groups, the group with lower validity may both have more individuals for whom the credit score is "too high" relative to the individual's inherent creditworthiness and more individuals for whom the credit score is "too low." In other words, the score exhibits greater variance, in a statistical sense, or equivalently contains greater signal noise around what the fully accurate credit score for each consumer would be. This means some applicants whose score is too high will be approved for a loan when they would not otherwise be approved if the score were perfectly accurate, or will pay an interest rate lower than what they

---

[86] *See* Ayşegül Şahin et al., *Mismatch Unemployment*, 104 AM. ECON. REV. 3529-64 (2014); Chang-Tai Hsieh et al., *The Allocation of Talent and U.S. Economic Growth*, 87 ECONOMETRICA 1439-74 (2019); Erik Hurst et al., *The Distributional Impact of the Minimum Wage in the Long Run* (2022), https://www.nber.org/system/files/working_papers/w30294/w30294.pdf; Blattner & Nelson, *supra* note 14; Disa M. Hynsjö & Luca Perdoni, The *Effects of Federal "Redlining" Maps: a Novel Estimation Strategy* (2022), http://congress-files.s3.amazonaws.com/2022-07/The%2520Effects%2520of%2520Federal%2520Redlining%2520Maps.pdf.

[87] For example, Regulation Z, the implementing regulation of the Truth in Lending Act, requires that card issuers consider the consumer's ability to make the required periodic payments. *See* discussion in Pace, *supra* note 47.

[88] Outside of lending, there has been greater emphasis on of validity measures of disparities, particularly in the context of insurance. In insurance law, the concept of unfair discrimination is typically understood as deviation from actuarial fairness: differential treatment of individuals that is not justified by statistically sound risk-based distinctions. *See* Kenneth S. Abraham, DISTRIBUTING RISK: INSURANCE, LEGAL THEORY, AND PUBLIC POLICY 64 (1986). This differential validity is closest to "calibration" definitions of algorithmic fairness. However, this notion does not generally account for the reliability of predictions, such as whether the model performs more or less accurately for different groups. As a result, a model could be well-calibrated—matching average outcomes—but still exhibit unequal validity along the dimensions of equal false positive or false negative rates, with higher error variance or misclassification rates for certain groups. We focus on differences in false negative rates and false positive rates, although we recognize that calibration could be an additional dimension of disparities that could concern regulators.

otherwise would pay; other applicants whose score is too low will be rejected when they would not otherwise be rejected if the score were perfectly accurate, or will pay an interest rate higher than what they otherwise would pay.

Hence for a group disadvantaged by differential validity, consumers whose score is "too low" face harm similar to that of consumers disadvantaged by unequal outcomes: higher likelihood of rejection or of facing an interest rate higher than what a fully accurate credit score would imply. Moreover, this disparity is particularly concentrated among the borrowers most able to repay the loan. Consumers whose score is "too high" face the possibility of being granted a loan they ultimately are not able to repay. Because of the myriad adverse consequences of loan default for consumers—which can extend even to incarceration in jurisdictions with particularly aggressive debt collection practices[89]—being granted a loan that one is not able to repay can commonly lead to harm, regardless of whether the inaccurately high credit score led to apparently favorable loan terms.[90]

While it is often assumed that lenders have sufficient incentives to develop valid models—an assumption that may partly explain the traditional focus of disparate impact law on unequal outcomes—this view overlooks important limitations. Even if lenders are generally motivated to improve model accuracy, this does not mean that *differential* validity will decrease. In particular, minority groups may be less profitable or more costly to model accurately due to limited data, reducing the incentive to ensure predictive validity across groups. Moreover, although lenders internalize the cost of lending to borrowers who default, they are far less likely to internalize the social cost of denying loans to creditworthy applicants.

III. RECOGNIZING DIFFERENTIAL VALIDITY WITHIN FAIR LENDING

Although fair lending has traditionally focused on equality of outcomes, expanding the measure of disparities to also consider differential validity is consistent with the goals of fair lending. In particular, a central goal of fair lending is to provide credit to applicants who are creditworthy, and ensuring the validity of lending policies is crucial for advancing this goal. Conversely, any failure to have equally valid lending policies across demographic groups is inconsistent with the goals of fair lending.

A. Consistency with Goals of Fair Lending

Considering validity is crucial to the achieve the goals of fair lending. ECOA's articulated goal is to provide equal access to credit among all creditworthy individuals. Statutory clarification instructs as follows: [91]

---

[89] *See generally* ACLU, A POUND OF FLESH: THE CRIMINALIZATION OF PRIVATE DEBT (2018), https://www.aclu.org/report/pound-flesh-criminalization-private-debt.

[90] *See* Holli Sargeant, *Algorithmic Decision-Making in Financial Services: Economic and Normative Outcomes in Consumer Credit*, 3 AI & ETHICS 1295 (2023) (discussing how false positives in loan predictions can raise banks' costs due to increased credit risk and potential defaults, while consumers may face financial strain from loans they may not be able to repay).

[91] Pub. L. 93-495, title V, § 502, Oct. 28, 1974, 88 Stat. 1521. ECOA's legislative history demonstrates that it was enacted to combat discriminatory barriers in lending by providing everyone with an equal opportunity to credit according to their creditworthiness. Although Congress rejected proposals to protect against discrimination because of race, color, national origin, age, and religion, Congress opted to protect women from such discrimination and passed the Equal Credit Opportunity Act as part of the Depository Institutions Amendments Act of 1974. According to Congresswoman Sullivan, "the discrimination provision is not as strong as we would have gotten if this had been handled as a separate bill . . . [i]t's

> [T]here is a need to ensure that the various financial institutions and other firms engaged in the extensions of credit exercise their responsibility to make credit available with fairness, impartiality, and without discrimination on the basis of sex or marital status… It is the purpose of this Act … to require that financial institutions and other firms engaged in the extension of credit make that credit equally available to all credit-worthy customers without regard to sex or marital status.

ECOA's goals to make credit equally available to all credit-worthy customers require that a protected class of creditworthy customers not be disproportionately impacted by what amounts to randomness in the credit determinations. A renewed focus on equal validity in lending decisions is directly responsive to this statutory goal.

Fair lending regulation has also been sensitive in the past to whether information is empirically valid for predicting creditworthiness, albeit only implicitly and only in certain circumstances. One of the most salient example of fair lending's consideration of validity is in the context of when it is permissible to consider age in a lending model, shielding lenders from discrimination liability for including age as a variable in "any empirically derived credit system" that is "demonstrably and statistically sound."[92]

In the context of fair lending, there has also been increasing attention to disparity metrics that are closely related to validity, such as model performance.[93] For example, a recent report by Charles Rivers Associates, who often oversee fair lending compliance for clients, suggests that in addition to traditional measures of AIR and SMD, "one can look at the performance of the model among demographic groups."[94]

B. Key Questions for Future Work

Fully integrating validity disparities into fair lending enforcement requires addressing some core challenges in expanding the definition of disparities. Below we outline two of these key issues that should be addressed in future work.

---

a start, but it's not as far as it should have gone." Congress went further two years later and amended ECOA to extend protection against discrimination because of "race, color, age, religion, national origin, status as a public benefit recipient, or victim of creditor retaliation for suing under the Consumer Credit Protection Act."

[92] Equal Credit Opportunity Act Amendments of 1976, Pub. L. No. 94-239, § 3, 90 Stat. 251, 253 (1976) (codified at 15 U.S.C. § 1691(b)(3)) (permitting consideration of age in credit systems that are empirically derived and demonstrably and statistically sound). According to Regulation B, to be an empirically derived scoring system, it must be based on empirical comparisons and periodically revalidated to "maintain predictive ability." Although the requirement to show a valid empirical model is only for the purpose of including age in the model and not for including other protected characteristics, and that age can be only used if an elderly applicant is not assigned a negative factor, it reflects fair lending's awareness of what is required from a valid model that is not merely "judgmental." 12 C.F.R. § 1002.2(p). *See also* discussion in David C. Hsia, *Credit Scoring and the Equal Credit Opportunity Act*, 30 HASTINGS L.J. 371, 406 (1978) (noting that that these three standards "impose only minimal statistical standards for predictive accuracy" and "any system with a modicum of predictive power should easily pass government specifications").

[93] *See, e.g.*, Verma & Rubin, *supra* note 13, at 4 ("The main idea behind this definition is that applicants with a good actual credit scope and applicants with a bad actual credit score should have a similar classification, regardless of their gender.").

[94] CHARLES RIVER ASSOCIATES, *What is Disparate Impact Testing?*, CRA INSIGHTS: FINANCIAL ECONOMICS 1, 3 (Jan. 2023), https://media.crai.com/wp-content/uploads/2023/01/30101848/FE-Insights-What-is-Disparate-Impact-Testing.pdf.

1. Tensions across disparity definitions

In the previous sections, we argued that reducing differential validity of lending decisions is crucial for equal access to credit and that equal outcomes alone do not capture these dimensions. However, the uncomfortable truth is that there are often trade-offs among different notions of fairness. For example, violations of outcome equality can be inversely related to inequality across groups in the *validity* of a screening technology.[95] The tension between equal outcomes and equal validity in fair lending arises from several interconnected sources. Models optimized for overall population performance often overlook patterns specific to smaller subgroups, leading to differential validity, a phenomenon linked to "aggregation bias."[96] Furthermore, efforts to enforce equal outcomes, such as avoiding certain data points or applying statistical constraints, can unintentionally distort error distributions across groups, exacerbating validity disparities even as they improve outcome equality.[97] This observation highlights the complex trade-offs between achieving fairness in outcomes and ensuring predictive accuracy across demographic groups.

These tensions between different definitions of disparities have been recognized and explored within the algorithmic fairness literature.[98] Expanding fair lending's consideration to include differential validity requires acknowledging these conflicting objectives, as efforts to improve validity disparities may exacerbate outcome disparities. Future work should focus on developing methods to quantify these tensions and creating frameworks that balance the dual goals of addressing both validity and outcome disparities when they conflict.

2. Measuring loan outcomes

For fair lending to consider differential validity, a necessary step is to estimate counterfactual default outcomes among rejected loan applicants. Estimating validity requires computing the count of true positives, true negatives, false positives, and false negatives for each protected class group. While calculating true positives and false positives is straightforward, tantamount just to calculating default rates on originated loans, calculating the count of true negatives and false positives requires discerning

[95] An extreme example is using a weighted coinflip for one protected class group that achieves equality in loan approval rates with another protected class group, while the latter group is afforded a screening technology that more accurately—less randomly—allocates loans to borrowers who are a good match for a loan. But, affording less random underwriting decisions to the former group may then lead to violations of equal outcomes.

[96] *See generally* Blattner & Nelson, *supra* note 14; CAROLINE CRIADO PÉREZ, INVISIBLE WOMEN, DATA BIAS IN A WORLD DESIGNED FOR MEN (2019); David C. Chan, Matthew Gentzkow, & Chuan Yu, *Selection with Variation in Diagnostic Skill: Evidence from Radiologists*, 137 Q. J. ECON. 729-783 (2022); Robert Temple & Norman L. Stockbridge, *BiDil for Heart Failure in Black Patients: The US Food and Drug Administration Perspective*, 146 ANNALS OF INTERNAL MED. 57-62 (2007); Michael W. Sjoding, Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, & Thomas S. Valley, *Racial Bias in Pulse Oximetry Measurement*, 383 NEW ENG. J. OF MED. 2477-8 (2020).

[97] *See* Vitaly Meursault, Daniel Moulton, Larry Santucci & Nathan Schor, *One Threshold Doesn't Fit All: Tailoring Machine Learning Predictions of Consumer Default for Lower-Income Areas*, arXiv:2208.11116 (2022) (investigating the need for adapting machine learning models to better predict consumer default rates in lower-income areas, suggesting that using a single threshold across different socioeconomic groups can lead to suboptimal outcomes.)

[98] *See, e.g.*, Kleinberg et al., *supra* note 6, which focuses on calibration rather than equal outcomes ("parity"). The most closely related impossibility result is that from Alexandra Chouldechova, which implies the impossibility result stated here. *See* Alexandra Chouldechova, *Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments*, 5 BIG DATA 153-163 (2017).

how rejected loan applicants would have performed if granted a loan.[99] Future work should therefore consider approaches for assessing these counterfactual if-not-rejected loan outcomes that would allow for a comprehensive measure of differential validity.

CONCLUSION

At the core of this Article is the recognition that credit market disparities, which fair lending aims to address, are complex and multifaceted. These disparities stem from underlying differences among demographic groups, shaped by historical discrimination and other social forces. Fair lending does not operate in a vacuum; it intervenes in a context where applicant features and credit histories are deeply influenced by past inequalities. In this setting, discrimination doctrines like disparate impact seek to prevent further exacerbation of pre-existing disadvantages and increase fair access to credit, even when statistical differences among groups exist. This means that fair lending should care both about equalizing lending outcomes, and ensuring that loans are extended to those most likely to repay.

Historically, fair lending has failed to fully recognize the challenge of addressing disparities in an unequal world. We argue that fair lending should be modernized in critical ways. It should expand the dimensions of disparity that lenders measure and address. The traditional emphasis on outcome equality has overlooked the crucial dimension of equal validity, thereby undermining the equitable extension of credit and perpetuating the disadvantage faced by protected groups. We argue that incorporating differential validity into disparate impact analysis is possible within the existing legal framework, even though it has been largely neglected in enforcement. However, incorporating differential validity alongside differential outcomes requires addressing key challenges, including resolving conflicts between disparity goals and how to measure differential validity.

Our focus on the measurement of disparities—the lynchpin of disparate impact under fair lending—has direct implications for other areas of discrimination law, such as employment and housing. These domains also grapple with pre-existing differences among protected groups, leading to similarly complex and multifaceted tensions.

By explicitly recognizing the multiple goals of fair lending and addressing the accompanying questions in implementation, discrimination law can be better equipped to address contemporary challenges. While not the primary focus of this Article, we note that courts are increasingly scrutinizing outcome-based frameworks, as reflected in the Supreme Court's decision in *Students for Fair Admissions*. For those who view outcome metrics as problematic, a focus on differential validity—grounded in the accuracy of individual-level predictions—may offer a more compelling and legally durable path forward.

[99] This problem is often referred to as the selective labels problem. *See generally* Amanda Coston, Ashesh Rambachan & Alexandra Chouldechova, *Characterizing Fairness over the Set of Good Models Under Selective Labels*, PROC. OF THE INT'L CONF. ON MACHINE LEARNING 2144 (2021) (exploring how selective labels—situations where the outcomes are only observed for a subset of the population—affect the fairness of machine learning models and proposing a framework to evaluate fairness across different models that are consistent with the available data).

Modernizing fair lending in the ways we propose will lead credit to be extended more equitably and validly, ultimately fostering a more comprehensive form of equality than otherwise possible in the modern financial marketplace.