

WORKING PAPER · NO. 2024-18

Modernizing Fair Lending

Spencer Caro, Talia Gillis, and Scott Nelson

AUGUST 2024

MODERNIZING 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.

We challenge the conventional focus on equal outcomes, demonstrating how it has failed to address some of the most pernicious harms of traditional credit allocation and has stifled necessary machine-learning and alternative data innovations. We argue that disparities in validity of creditworthiness predictions—the accuracy with which a model identifies creditworthy applicants—severely impact equal access to credit and fail to equally extend credit to the creditworthy. Despite the mounting empirical evidence of the harm of validity disparities, traditional fair lending enforcement inadequately recognizes this disparity dimension, a gap that will become increasingly harmful as lending decisions rely on advanced statistical methods.

Our updated, holistic account of disparities engages with how the competing fairness notions of equal outcomes and equal validity may be fundamentally in tension. Reducing differential validity by improving data quality and models for protected groups could increase outcome inequality. On the other hand, decreasing outcome inequality could exacerbate differential validity when noisy and inaccurate lending decisions are made for the most vulnerable borrowers.

Using a lender simulation exercise, we demonstrate that these tensions can be addressed by balancing equal outcomes and validity, providing regulators the flexibility to weigh each dimension appropriately. We discuss how this approach can be practically implemented by lenders and regulators; we chart a path forward for considering other tensions within fair lending, such as balancing business benefit with discriminatory impact; and we address how fair-lending policies impact different protected groups. Our framework has direct implications for other domains—such as housing and employment discrimination—where competing disparity notions should be considered under discrimination doctrines.

* Spencer Caro, Independent.

Talia Gillis, Associate Professor, Columbia Law School.

Scott Nelson, Assistant Professor, Chicago Booth.

Many thanks to Dorothy Lund, Jared Mayer, and Bryce Stephens for helpful comments and discussions. Rotem Litinski and Sophia Stener provided excellent research assistance.

TABLE OF CONTENTS

INTRODUCTION	3
I. TRADITIONAL FAIR LENDING’S FOCUS ON EQUAL OUTCOMES.....	8
<i>A. Centrality of Disparities Metric to Disparate Impact Analysis.....</i>	<i>9</i>
1. Burden Shifting Framework	9
2. Disparity Metrics for Disparate Impact Analysis	12
<i>B. The Singular Focus on Equal Outcomes</i>	<i>13</i>
1. ECOA and FHA Regulations and Guidance Documents	14
2. Enforcement Actions	15
3. Private Litigation.....	17
II. EXPANDING DISPARITIES METRIC TO INCLUDE EQUAL VALIDITY	18
<i>A. Importance of Validity Disparities.....</i>	<i>18</i>
1. What is Differential Validity?.....	18
2. Significance of Differential Validity	21
<i>B. Recognizing Differential Validity within Fair Lending.....</i>	<i>22</i>
1. Consistency with Goals of Fair Lending	22
2. Prior Recognition of Additional Dimensions of Disparities.....	23
<i>C. Resolving Tensions between Equal Outcomes and Validity.....</i>	<i>24</i>
1. The Tradeoff Between Equal Outcomes and Equal Validity	25
2. Balancing Equal Outcomes and Validity	28
3. Importance of the Equal Outcomes-Validity Tradeoff with Big Data	29
III. IMPLEMENTING THE BALANCE BETWEEN EQUAL OUTCOMES AND VALIDITY.....	30
<i>A. Legal and Institutional Implementation.....</i>	<i>30</i>
1. Fair Lending’s Reliance on Public Enforcement	30
2. Implementing Disparity Measuring in LDA analysis.....	32
3. Implementation Disparity Analysis Through Other Regulatory Tools.....	33
<i>B. Challenges in Expanding Disparities Metric</i>	<i>37</i>
1. Missing Outcome Data Loan Outcomes	37
2. Missing Data on Demographics	39
3. The Use of Protected Characteristics.....	41
<i>C. Implications for Multiple Protected Groups.....</i>	<i>42</i>
<i>D. Implications for Disparity-Performance Tradeoff.....</i>	<i>43</i>
CONCLUSION	45

INTRODUCTION

There is emerging consensus that antidiscrimination law is in urgent need of reform.¹ Whether due to the increasing use of alternative data² or machine-learning tools,³ fair lending's disparate impact framework is ill-equipped to achieve fair outcomes in modern underwriting.⁴ Historically, disparate impact relied on scrutinizing the small set of inputs used by relatively simple screening technologies with the goal of detecting inputs that are prohibited proxies for protected characteristics and insufficiently causal of later outcomes accepted as business necessities, such as loan repayment.⁵

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., Robert Bartlett, Adair Morse, Nancy Wallace & Richard Stanton, *Algorithmic Discrimination and Input Accountability under the Civil Rights Acts*, 36 BERKELEY TECH. L.J. 675 (2020) (proposing an "Input Accountability Test" to ensure that algorithmic decision-making processes do not systematically discriminate against protected groups, advocating for input-based regulatory scrutiny over purely output-focused policies to better align with longstanding anti-discrimination law); Yang and Dobbie 2020, *supra* note 1.

However, modern underwriting has become too complex for such causal scrutiny to be feasible.⁶ Growing awareness of multifarious economic inequality has exposed the shortcomings of traditional disparate impact’s simplistic and indeterminate notions of discrimination.⁷ Disparate impact is not just becoming unworkable; its ambiguities prevent changes in business practices that are needed to address centuries-old sources of inequality.⁸

Inequality in credit markets stems not just from present day discriminatory behavior by individuals or single firms.⁹ The vestiges of historical discrimination are still with us today: in one prominent example, the descendants of victims of federally sanctioned redlining have lower wealth than their peers.¹⁰ Social practices and expectations also engender structural discrimination larger than individual people’s or firms’ actions.¹¹ These persistent and structural inequalities are evident in the

⁶ See generally Gillis, *supra* note 1; Laura Blattner, Scott Nelson, & Jann Spiess, *Unpacking the Black Box: Regulating Algorithmic Decisions*, ARXIV:2110.03443 (2021) (proposing a regulatory framework that balances the need for complex predictive models with the necessity of regulatory oversight, advocating for the use of targeted explainers that focus on aligning the preferences of regulators and agents).

⁷ Abbye Atkinson, *Rethinking Credit as Social Provision*, 71 STAN. L. REV. 1093, 1144 (2019) (arguing that credit has misguidedly become the principal mechanism of social mobility); see generally, KEEANGA-YAMAHTTA TAYLOR, *RACE FOR PROFIT* (2019) (documenting the exploitative lending practices that followed the ban on housing discrimination).

⁸ See e.g., STEPHEN L. ROSS & JOHN YINGER, *THE COLOR OF CREDIT: MORTGAGE DISCRIMINATION, RESEARCH METHODOLOGY, AND FAIR-LENDING ENFORCEMENT*, 311 (2002) (“[V]ariation in loan performance scoring systems across lenders could hide disparate-impact discrimination.”); see also Dan Black et al., *Discrimination in Mortgage Lending*, 68 AM. ECON. REV. 186, 189 (1978) (“[R]ace is an important determinant in the loan decision”); Helen Ladd, *Evidence on Discrimination in Mortgaged Lending*, 12 J. ECON. PERSPS. 41 (1998) (discussing statistical methods to detect discrimination in mortgage markets). See also Peter Swire, *The Persistent Problem of Lending Discrimination: A Law and Economics Analysis*, 73 TEX. L. REV. 787 (1995) (discussing the causes of persistent discrimination in credit markets).

⁹ In the context of employment discrimination, see 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).

¹⁰ 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” Maps*, 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. ELICKSON, *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.”).

¹¹ 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).

data used in underwriting to determine which individuals receive credit and on what terms. Credit reports, for example, contain less data for historically disadvantaged groups, such as racial and ethnic minorities, if for no other reason than the *inherited* lack of access to credit.¹² These disparities hinder access to credit markets—especially for individuals who need loans the most but who struggle to make their loan application attractive to lenders.

Our Article makes three contributions to the literature. Our first contribution is to highlight shortcomings in current fair lending enforcement, which focuses solely on outcome inequality while neglecting disparities in model accuracy—validity—for protected groups. This oversight has allowed some of the most pernicious lending practices to persist while stifling necessary machine-learning and alternative data innovations. Our second contribution is to demonstrate, using a novel simulation exercise, that there are tensions between improving model validity for protected groups and reducing decision disparities. Our third contribution is to propose an alternative framework for disparate impact that recognizes the dual importance of equal outcomes and equal validity, offering a balanced approach to achieving these goals.

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. However, regulatory guidance, enforcement action, and private litigation have predominately focused on outcome equality in lending decisions, such as disparities in loan approval rates and terms. This singular focus on outcome equality has overlooked other disparity dimensions related to differential validity,¹³ highlighted in the algorithmic fairness literature, despite the growing body of evidence that longstanding structural inequalities across race, ethnicity, gender, and other protected classes make traditional business practices differentially valid across groups.

Loan underwriting often requires 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. In short,

¹² 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 Immigrants* (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 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.”).

¹³ These notions include calibration, “balance for the positive class,” “balance for the negative class,” etc. See Kleinberg et al., *Inherent Trade-Offs in the Fair Determination of Risk Scores*, 8TH INNOVATIONS IN THEORETICAL COMPUTER SCIENCE CONFERENCE (ITCS 2017), for a discussion of how these three are incompatible with each other and, typically, with equal outcomes as well.

the arity of disparity is more numerous than disparate impact analyses traditionally acknowledge, making disparate impact complex and fraught.

Prior work has demonstrated the prevalence of disparate validity in credit screening tools, for reasons including discrimination, data inequality, and statistical models that perform poorly for historically disadvantaged groups.¹⁴ Differential validity occurs when the correlation between screening criteria, such as credit score, and later outcomes, such as loan default, differs between groups.¹⁵ This is exactly what credit scores exhibit for minority and non-minority loan applicants.¹⁶ 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 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.¹⁷

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. Unlike outcome equality, which only compares observed lender decisions, measuring validity is more challenging because it requires estimating counterfactual loan outcomes for borrowers who were denied loans. To address this challenge, we propose several methods to estimate

¹⁴ 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”).

¹⁵ 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 measures 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 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.”).

¹⁶ 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).

¹⁷ 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).

differential validity by inferring these unobservable outcomes from complementary data, low-cost randomized trials, or changes in the regulatory system.

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.¹⁸

For our second contribution, we show that achieving equal validity in fair lending contexts is far more complex, however, than just enumerating it as a goal. Equal outcomes and equal validity can be – and often are – in tension with each other.¹⁹ Our analysis shows that innovation that addresses longstanding inequality in the validity of lending decisions may exacerbate inequality in outcomes. For instance, as new data or other innovations in the loan underwriting process provide more accurate correlations between approval decisions and individuals’ ability to repay a loan, thus improving the validity of underwriting decisions, some individuals’ loan applications might be more likely to be rejected.²⁰ On the other hand, as lenders forgo useful innovation in order to hew as closely as possible to fair lending’s singular goal of outcome equality, disparities in validity equality may increase. We illustrate these tensions between equal outcomes and equal validity through a simulation exercise, demonstrating that many models with lower outcome equality exhibit higher validity equality.

Our third contribution is to introduce an alternative framework for fair lending. Because equal outcomes and equal validity are both important but can be in tension with each other, it is imperative for the disparate impact framework to explicitly balance both goals. We argue that this balance must be quantifiable: enjoining a credit system to pursue a vague or undefined goal, such as often required under current regulations,²¹ is insufficient. The very problem that impedes advances in lending processes is the indeterminacy of what, exactly, would make a new business practice

¹⁸ 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]”).

¹⁹ Stefan Buijsman, *Navigating Fairness Measures and Trade-Offs*, 1 AI ETHICS (2023) (“[I]n virtually all cases, it is mathematically impossible to optimize for ... a wide range of statistical (and other, such as causal) measures . . . [I]nstead we will have to choose which fairness measure to prioritize in each case.”); Michelle Seng Ah Lee, Luciano Floridi & Jatinder Singh, *Formalising Trade-Offs Beyond Algorithmic Fairness: Lessons from Ethical Philosophy and Welfare Economics*, 1 AI ETHICS 529-544 (2021) (relating contextual considerations from ethical philosophy and welfare economics to the balancing of competing mathematical measures of fairness).

²⁰ See *infra* section II.C (discussing the tensions between equal outcomes and equal validity); see also Andreas Fuster et al., *Predictably Unequal? The Effects of Machine Learning on Credit Markets*, J. FINANCE (forthcoming).

²¹ See, e.g., the way the HUD’s Discriminatory Effect Rule (2023) describes the business justification is that the practice is necessary to achieve “one or more substantial, legitimate, non-discriminatory interests” without defining what would qualify as “substantial.” See 24 C.F.R. § 100.500 (2013).

permissible when it leads to different consumer outcomes than baseline lending practices. Our proposed framework offers flexibility for policymakers to determine the precise weights assigned to each dimension and to define the type of disparity measures to be used.

A weighting approach to competing fair lending goals sheds light on several additional shortcomings of the current regulatory framework. We explore the implications of our proposed approach for settings where multiple protected groups are affected by a lending policy and lenders may need to explicitly engage in a tradeoff of disparities for different groups. Additionally, we consider how an explicit balancing approach is able to address the tradeoffs between enhanced model performance and group disparities.²²

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 idea of validity disparities, discusses the tension between equal outcomes and validity measures, and presents a simulation exercise to demonstrate these tradeoffs, arguing for a balanced consideration of both disparity metrics. Part III explores the challenges and opportunities in implementing a weighing approach to disparities, addressing regulatory mechanisms, key challenges, and the broader implications for modernizing fair lending practices. We then conclude.

Our analysis extends beyond fair lending to other areas of discrimination law, including employment and housing, where disparate impact is required to consider multiple, often conflicting goals. The Supreme Court’s recent decision in *Students for Fair Admissions, Inc.*,²³ which struck down race conscious university admissions, suggests a shift in the permissible methods to achieve fairness, likely affecting other discrimination domains. Our Article contributes to this reevaluation by arguing for a clear definition of equally compelling, yet competing, goals in fair lending, and by proposing a framework to balance them effectively.

I. TRADITIONAL FAIR LENDING’S FOCUS ON EQUAL OUTCOMES

This section provides an overview to disparate impact under ECOA and FHA. 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.

²² See generally Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel & Aziz Huq, *Algorithmic Decision Making and the Cost of Fairness*, in 1 PROC. 23RD ACM SIGKDD INT’L CONF. ON KNOWLEDGE DISCOVERY & DATA MINING 797 (2017) (examining the trade-offs between algorithmic fairness and accuracy, highlighting how efforts to improve fairness in decision-making algorithms can sometimes result in increased costs or reduced predictive performance.); Annie Liang, Jay Lu & Xiaosheng Mu, *Algorithm Design: A Fairness-Accuracy Frontier*, ARXIV:2112.09975 (2023) (exploring the trade-offs between fairness and accuracy in algorithmic decision-making, proposing a framework to identify optimal solutions along the fairness-accuracy frontier).

²³ *Students for Fair Admissions, Inc. v. President & Fellows of Harvard Coll.*, 600 U.S. 181 (2023).

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 has primarily focused on the *equal-outcome* dimension of disparities—meaning differences in approval rates and loan terms—rather than *differential validity*, which is reflected in regulatory guidance and fair lending enforcement actions.

1. Burden Shifting Framework

The two federal statutes that form the core prohibition on discrimination in credit are the Fair Housing Act (FHA)²⁴ and the Equal Credit Opportunity Act (ECOA).²⁵ FHA, also known as Title VIII of the Civil Rights Act of 1968, protects renters and buyers from discrimination by sellers or landlords, 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.²⁶ ECOA extends this prohibition to all credit transactions,²⁷ not just those in the context of housing.²⁸

Both ECOA and FHA incorporate the doctrines of “disparate treatment” and “disparate impact.”²⁹ Disparate treatment addresses direct conditioning of a credit decision on a protected

²⁴ Fair Housing Act, Pub. L. No. 90-284, 82 Stat. 73 (1968) (codified as amended at 42 U.S.C. §§ 3601–3619 (2018)).

²⁵ 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)).

²⁶ 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).

²⁷ *See* 15 U.S.C. § 1691 (a)(1)–(2).

²⁸ 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. The empirical discussion of this Article relies on HMDA data. *See supra* section I.C.

²⁹ 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

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.

A disparate impact case typically follows the three-part burden-shifting framework developed in the Title VII employment discrimination context:³⁰ (1) the plaintiff must demonstrate that a certain factor disproportionately negatively impacts a protected group;³¹ (2) if successful, the creditor must show that the criterion makes the credit evaluation system more predictive or is justified by a legitimate business need;³² 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.³³

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,³⁴ preventing plaintiffs from furnishing relevant statistics to show prima facie

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 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).

³⁰ 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.”).

³¹ *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].”).

³² *Cherry*, 490 F. Supp. at 1031 (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.”).

³³ 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.”)

³⁴ 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).

discrimination.³⁵ 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.³⁶ 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 1.³⁷

For Step 2, once a prima facie case of disparities is established, the defendant must prove that the practice is consistent with business necessity.³⁸ Regulatory guidance on ECOA states that creditors can defend a policy that produces disparity by showing a demonstrable relationship between the challenged policy and creditworthiness.³⁹ Some have argued that the variable must be related to creditworthiness and not merely included for the purpose of maximizing profit.⁴⁰ However, the justification could include some appeals to profitability if they are “manifest” and not “hypothetical” or “speculative.”⁴¹

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.’”⁴² The requirement to find an alternative practice was referred to as “other tests or selection devices, without a similarly undesirable racial effect”⁴³ in the *Albemarle* case.⁴⁴

³⁵ 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).

³⁶ 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.”).

³⁷ 64 Fed. Reg. 44582, 44586 (Aug. 16, 1999).

³⁸ 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 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.”).

³⁹ OFFICIAL STAFF COMMENTARY TO REGULATION B, 12 C.F.R. § 1002.6(A)-2.

⁴⁰ 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).

⁴¹ POLICY STATEMENT ON DISCRIMINATION IN LENDING, 59 FR 18266-01, 18269-70.

⁴² *Albemarle*, 422 U.S. at 425 (1975) (quoting *McDonnell Douglas Corp. v. Green*, 411 U.S. 792, 801 (1973)).

⁴³ *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).

⁴⁴ 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).

In general, there is very little guidance on the LDA stage of disparate impact.⁴⁵ 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 impact.”⁴⁶ In the context of creditworthiness models, an LDA must perform at or close to the level of the challenged practice.⁴⁷

Although traditionally considered the plaintiff’s burden, regulators have recently suggested that lenders must proactively search for LDAs. CFPB’s Associate Director of Fair Lending and Equal Opportunity, Patrice Ficklin,⁴⁸ and recent CFPB supervisory updates⁴⁹ 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. Equality 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

⁴⁵ 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.”).

⁴⁶ 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.”).

⁴⁷ See Mahoney, *supra* note 45, 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).

⁴⁸ 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>.

⁴⁹ 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.

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 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.⁵⁰ On the other hand, controlling for applicant characteristics may mask the true disparate impact of policies such that a raw, unconditional measure is preferable.⁵¹

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 default rates on loans are consistent with the predicted probabilities of default for different demographic groups, revealing disparities in how accurately the lending criteria predict loan performance across these groups. 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 outcomes 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

⁵⁰ 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>.

⁵¹ RELMAN COLFAX PLLC, FAIR LENDING MONITORSHIP OF UPSTART NETWORK'S LENDING MODEL, FIRST REPORT OF THE INDEPENDENT MONITOR at 10 (2021), https://www.reلمانlaw.com/media/cases/1088_Upstart%20Initial%20Report%20-%20Final.pdf.

groups.⁵² As discussed above, both metrics are equal-outcome measures are often estimated without accounting for differences in the baseline applicant pools.⁵³ We will examine legal and 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 any details of the analysis can be 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.⁵⁴ 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.”⁵⁵ One of the Rule’s illustrations of discrimination in a real-estate transaction refers to “denial rate” differences as a form of discrimination.⁵⁶ 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.⁵⁷

Regulation B, the implementing regulation of ECOA, adopts the “effects test” from employment law.⁵⁸ 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”, illustrated by an example of an income requirement for an overdraft line that leads to women and

⁵² 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, EXPLAINABILITY & FAIRNESS IN MACHINE LEARNING FOR CREDIT UNDERWRITING 59 (2023), https://finreglab.org/wp-content/uploads/2023/12/FinRegLab_2023-12-07_Research-Report_Explainability-and-Fairness-in-Machine-Learning-for-Credit-Underwriting_Policy-Analysis.pdf; 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).

⁵³ 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 . . .”).

⁵⁴ 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).

⁵⁵ 24 C.F.R. § 100.500(a).

⁵⁶ *Id.* § 100.120(b)(2).

⁵⁷ See 29 C.F.R. § 1607.3 (1978) (“Discrimination defined: Relationship between use of selection procedures and discrimination.”).

⁵⁸ 12 C.F.R. § 1002.6(a).

minority applicants being “rejected at a higher rate than men and nonminority applicants,” thereby using an equal outcome disparity metric.⁵⁹ 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 their measurement of disparities. These reports reflect the CFPB’s focus on denial rates disparities. For example, in its Fall 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.⁶⁰

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.⁶¹ In considering the disparity metric, these cases focus exclusively on what we term equality of outcomes, by considering disparities in loan approvals, interest rates and fees. Below we provide several examples of consent orders and settlements and the way they describe the disparities created by the lender to demonstrate the focus on equal-outcome metrics.

In a group mortgage lending cases in the wake of the 2007–2008 financial crisis, lenders were challenged for lending disparities that were deemed discriminatory.⁶² A typical example is the consent order from 2014 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

⁵⁹ *Id.*

⁶⁰ 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.

⁶¹ 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).

⁶² 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.

then giving loan officers and brokers discretion to deviate from those rates. According to the consent order, “National City Banks 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.” 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 loans than similarly situated White applicants.”⁶³ In other cases dealing with pre-financial crisis lending, complaints similarly focus on disparities in approval rates and loan costs, an equal-outcome metric.⁶⁴

A second group of cases involved the auto lending industry, where discretionary markup policies were not prohibited after the financial crisis.⁶⁵ 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,”⁶⁶ thereby focusing on unequal outcomes in the markup component of pricing. Other cases against Honda in

⁶³ See Consent Order at 2, *United States v. BancorpSouth Bank*, No. 16-118 (N.D. Miss. July 25, 2016).

⁶⁴ *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).

⁶⁵ These were a series of cases following the now-invalidated 2013 CFPB Bulletin, the CFPB challenged auto financiers’ 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); see also Mark Cohen, *Imperfect Competition in Auto Lending: Subjective Markups, Racial Disparity, and Class Action Litigation*, (Vanderbilt L. and Econ. Rsch. Paper No. 07-01, 2006).

⁶⁶ *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 contract, 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.

2015,⁶⁷ Fifth Third Bank in 2015,⁶⁸ and Toyota in 2016⁶⁹ contain similar language. A number of more recent cases brought by the Federal Trade Commission against auto dealers⁷⁰ use similar disparity metrics by claiming that racial minorities were charged different costs, fees and interest rates.⁷¹

3. Private Litigation

Courts have similarly focused on equal-outcome metrics for estimating disparities. In *Miller v. Countrywide Bank, N.A.*,⁷² for example, the District of Massachusetts court found that “African-American borrowers are charged higher fees and rates than similarly situated white borrowers,”⁷³ focusing on the outcome equality of fees and interest rates as the disparity metric. Disparate impact private litigation, both for auto lending⁷⁴ and mortgage lending,⁷⁵ has similarly focused on pricing disparities.

⁶⁷ *Honda*, 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.”

⁶⁸ *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.

⁶⁹ *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 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.

⁷⁰ 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.

⁷¹ 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).

⁷² 571 F. Supp. 2d 251, 253 (D. Mass. 2008).

⁷³ *Id.* at 259.

⁷⁴ 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 65, 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.ogosenet.com/litigation/case-index-closed-cases.html>.

⁷⁵ 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.

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.

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. 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. Importance of Validity Disparities

1. 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 group⁷⁷ – 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.

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).

⁷⁶ These notions include calibration, “balance for the positive class,” “balance for the negative class,” etc. See Kleinberg et al., *supra* note 13, 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.”).

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 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.

To keep our discussion simple and concrete, we focus hereafter on a notion of validity that quantifies the relationship between loan approval and loan repayment: specifically, a lender’s approval rate *among* the applicants who would repay a loan.⁷⁹ 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.”⁸⁰

In the example of the lender that underwrites one group using a highly accurate risk score 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.⁸¹

Figure 1 provides a graphical depiction of differential validity.⁸² 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

⁷⁹ 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>.

⁸⁰ 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 15, and Verma & Rubin, *supra* note 15. 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).

⁸¹ 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).

⁸² See also MICHAEL KEARNS & AARON ROTH, THE ETHICAL ALGORITHM: THE SCIENCE OF SOCIALLY AWARE ALGORITHM Design 77 (2020) (providing a similar diagrammatic representation).

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.

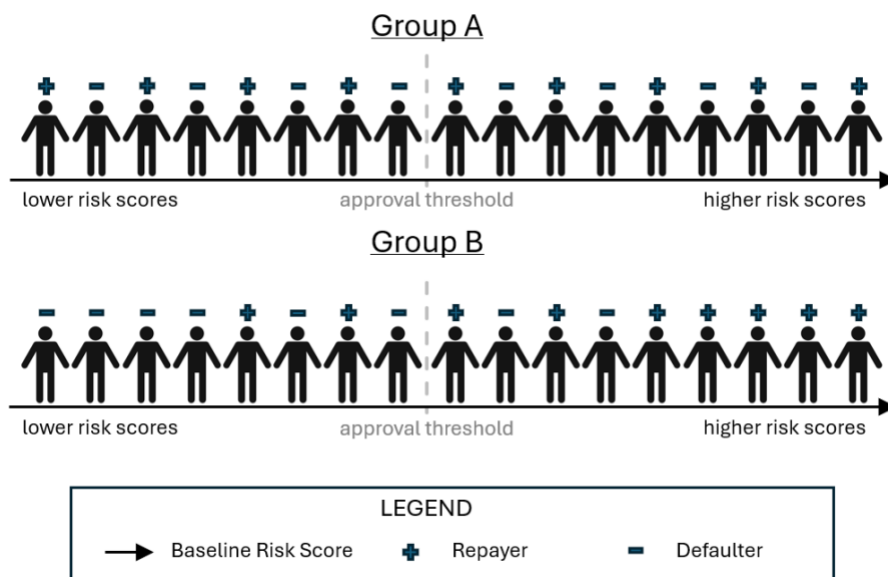


Figure 1: Differential Validity

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.⁸³ 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.⁸⁴ These concerns echo those raised in other contexts where the *misallocation* of opportunities like loans, education, or jobs is seen as holding first-order importance.⁸⁵

⁸³ See Blattner & Nelson, *supra* note 16.

⁸⁴ See Blattner & Nelson, *supra* note 16.

⁸⁵ 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 16; Disa M. Hynsjö & Luca Perdoni, *The Effects of Federal “Redlining” Maps: a Novel Estimation*

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.⁸⁶

2. 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. 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 randomly selecting borrowers 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 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 or being assigned an interest rate that may induce over-borrowing relative to what the consumer’s inherent creditworthiness can support. Because of the myriad adverse consequences of loan default for consumers—which can extend even to incarceration in jurisdictions with particularly aggressive debt collection practices⁸⁷—being granted a loan that one is not able to repay

Strategy (2022), <http://congress-files.s3.amazonaws.com/2022-07/The%2520Effects%2520of%2520Federal%2520Redlining%2520Maps.pdf>.

⁸⁶ 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 50.

⁸⁷ *See generally* ACLU, A POUND OF FLESH: THE CRIMINALIZATION OF PRIVATE DEBT (2018), <https://www.aclu.org/report/pound-flesh-criminalization-private-debt>.

can commonly lead to harm, regardless of whether the inaccurately high credit score led to apparently favorable loan terms.

B. 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 language and goals of fair lending. In particular, the goal of fair lending in providing credit to applicants who are creditworthy can only be accomplished through validity in lending policies, as a low-validity policy would be closer to allocating loans without regard to creditworthiness—even if at equal approval rates across groups. Correspondingly, any failure to have equally valid lending policies across demographic groups is inconsistent with the goals of fair lending.

1. 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:⁸⁸

“[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. Economic stabilization would be enhanced and competition among the various financial institutions and other firms engaged in the extension of credit would be strengthened by an absence of discrimination on the basis of sex or marital status, as well as by the informed use of credit which Congress has heretofore sought to promote. 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 enhance economic stabilization, promote the informed use of credit, and 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. In sum, the goal of ECOA is to ensure equal opportunity to credit among the creditworthy, and we provide a framework that implements this goal amid the unique challenges of modern-day underwriting.

⁸⁸ 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 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.”

2. Prior Recognition of Additional Dimensions of Disparities

Fair lending regulation has also been sensitive in the past to whether information is empirically valid for predicting creditworthiness. This shows fair lending’s attention to validity in assessing creditworthiness, albeit only implicitly and only in certain circumstances.

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. In 1976, two years after ECOA’s passage, Congress amended it to shield lenders from discrimination liability for including age as a variable in “any empirically derived credit system that is “demonstrably and statistically sound.”⁸⁹ 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.”⁹⁰ 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,⁹¹ and that age can be only used if an elderly applicant is not assigned a negative factor,⁹² it reflects fair lending’s awareness of what is required from a valid model that is not merely “judgmental.”⁹³

At times, Regulation B interprets ECOA in a way that actively proscribes a credit policy likely to reduce differential validity. One pertinent example is in Section 1002.6(b)(5), which instructs that creditors consider “alimony, child support, or separate maintenance payments” as income when applying for credit.⁹⁴ Reliable income sources such as alimony or child support genuinely inform the recipient’s ability to repay, but they might be overlooked by a lender who fails to account for these data’s relevance for certain subgroups, such as female applicants.⁹⁵ Even if the lender otherwise ensured equality in approval rates for male and female applicants, additionally considering these other income sources can ensure that a lending policy allocates loans to those who are best positioned to repay. Similarly, Section 1002.6(b)(6), mandates that creditors distinguish “accounts that the applicant and the applicant’s spouse are permitted to use or for which both are contractually liable”⁹⁶ and consider information that shows that information of credit history does not reflect the applicant’s creditworthiness.⁹⁷ Removing data that “[did] not accurately reflect” female applicant’s creditworthiness, advanced the validity of lending policies for women.”⁹⁸

⁸⁹ Pub. L. No. 93-495, § 503, 88 Stat. 1500, 1522 (1974).

⁹⁰ 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”).

⁹¹ The inclusion of other protected characteristics is always prohibited. *See* the Official Interpretation for 12 C.F.R. § 1002.2(p) (“Besides age, no other prohibited basis may be used as a variable.”).

⁹² *Id.* (“...provided that age of an elderly applicant is not assigned a negative factor or value”).

⁹³ Hsia, *supra* note 90, at 406. *See also* Victoria Angelova et al., *Algorithmic Recommendations and Human Discretion*, (NBER Working Paper No. 31747, 2023), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4589709.

⁹⁴ 12 C.F.R. § 1002.6(b)(5)

⁹⁵ *See generally* CAROLINE CRIADO PÉREZ, *INVISIBLE WOMEN, DATA BIAS IN A WORLD DESIGNED FOR MEN* (2019).

⁹⁶ 12 C.F.R. § 1002.6 (b)(6)(i).

⁹⁷ 12 C.F.R. § 1002.6 (b)(6)(ii).

⁹⁸ Taylor, *supra* note 29, at 614.

Sensitivity to the validity of a test, rather than just selection rates, can be also found in employment discrimination, which is the touchstone of fair lending disparate impact. Under the federal government’s Uniform Guidelines on Employee Selection Procedures (“UGESP”), the concept of validity refers to a satisfactory correlation between the predictor and outcome, such as job performance⁹⁹ (e.g., completing a bachelor’s degree).¹⁰⁰ Differential validity, then, is an unsatisfactory difference in this correlation coefficient between groups (e.g., white and nonwhite). Employment associations and practitioners have also laid down principles for the validation and use of selection methods,¹⁰¹ rejecting equal outcomes as the sole basis for selection method fairness.¹⁰²

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.¹⁰³ 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.”¹⁰⁴

C. Resolving Tensions between Equal Outcomes and Validity

In the previous Section, 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. In this Section we confront an uncomfortable truth—there are 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.¹⁰⁵

⁹⁹ Uniform Guidelines on Employee Selection Procedures, 29 C.F.R. § 1607 (1978). In the context of college admissions, the College Board has discussed the requirement that there be a correlation between the predictor (e.g., SAT score) and outcome (e.g., completing a bachelor’s degree). See THE COLLEGE BOARD, DIFFERENTIAL VALIDITY, DIFFERENTIAL PREDICTION, AND COLLEGE ADMISSION TESTING: A COMPREHENSIVE REVIEW AND ANALYSIS, Research Report No. 2001-6 at 4 (2001), <https://files.eric.ed.gov/fulltext/ED562661.pdf>.

¹⁰⁰ 29 C.F.R. § 1607 (1978).

¹⁰¹ The Society for Industrial and Organizational Psychology (SIOP), a professional association for practitioners and researchers in the field of industrial-organizational psychology, takes validity to be “a unitary concept with different sources of evidence contributing to an understanding of the inferences that can be drawn from a selection procedure.” See SOCIETY FOR INDUS. ORGAN. PSYCH., PRINCIPLES FOR THE VALIDATION AND USE OF PERSONNEL SELECTION PROCEDURES 38 (5th ed. 2018), <https://www.apa.org/ed/accreditation/about/policies/personnel-selection-procedures.pdf>.

¹⁰² SIOP flat-out “rejects the concept of parity of outcomes as the main definition of test fairness.” Kelly Trindel et al., *Fairness in Algorithmic Employment Selection: How to Comply with Title VII*, 35 ABA J. LABOR & EMP. L. 241, 268 (2021).

¹⁰³ See, e.g., Verma & Rubin, *supra* note 15, at 4 (“The main idea behind this definition is that applicants with a good actual credit score and applicants with a bad actual credit score should have a similar classification, regardless of their gender.”).

¹⁰⁴ 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>.

¹⁰⁵ 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.

Below we explore the tensions among different disparity notions. Modern screening technologies present an unprecedented number of sensible business choices. Should disparate impact regulation always prioritize equal outcomes over validity? Intuitively, no. Given the growing body of research discussed above on the costs of differential validity and the clear instances of unfairness discussed above that arise from differential validity, fair-lending regulation needs to thoughtfully balance violations of equal outcomes and validity rather than focus on unalloyed parity in outcomes alone. We formalize this intuition and provide a workable protocol for resolving competing disparities that balances different notions.

1. The Tradeoff Between Equal Outcomes and Equal Validity

In many real-world settings, lenders can only reduce differential validity by also affecting outcome equality.¹⁰⁶ Moreover, when considering the many models that are otherwise similar, they may have different levels of unequal outcomes and differential validity, raising questions about which combinations of the two should be seen as permissible.

The tension between equal outcomes and equal validity in fair lending arises from several interconnected sources. When a model is optimized for lending decisions, the primary goal is often to minimize prediction errors in the population of applicants overall, thereby maximizing overall performance. This optimization process might overlook patterns in the data that are particularly relevant for some subgroups that are a relatively small share of the overall population, precisely because the model is optimized for the population as a whole and not for small subgroups. This source of differential validity, which elsewhere has been termed “aggregation bias,” is known to be important in settings as diverse as the design of urban transit networks, pharmaceutical research, medical advice, and consumer lending.¹⁰⁷

Another source of tension stems, ironically, from lender efforts to comply with a disparate impact framework that often is focused narrowly on equality in outcomes. Because demographic groups have differing risk profiles due to socio-economic factors and historical disadvantage, a lending policy may be designed to avoid using certain datapoints that would exacerbate differential outcomes across groups. However, this avoidance can exacerbate differential validity. Enforcing equal outcomes by treating groups with different risk profiles similarly can misrepresent the actual risk, thereby impacting validity. Additionally, the application of statistical constraints to ensure equal outcomes can alter the distribution of errors (such as false positives and false negatives) across demographic groups. For example, equalizing approval rates might increase the false positive rate for a higher-risk group while decreasing it for a lower-risk group, improving equal outcomes but compromising equality of outcomes.¹⁰⁸

¹⁰⁶ This impossibility result is complementary to but distinct from that in Kleinberg et al., *supra* note 13, 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).

¹⁰⁷ See generally Blattner & Nelson, *supra* note 16; Pérez, *supra* note 95; 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).

¹⁰⁸ 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

In some situations, validity and equal outcomes are in tension because fair lending may limit the use of empirically relevant differences for normative reasons. For example, in predicting future income for the purposes of determining creditworthiness, a statistically valid model may use gender in predicting income. Applicants who otherwise look similar may have different income trajectories because of their gender, whether resulting from labor market discrimination or other reasons. Therefore, fair lending’s limitations on the use of gender in lending decision creates a chasm between empirically and normatively relevant information, partially forgoing validity for the sake of equal outcomes.

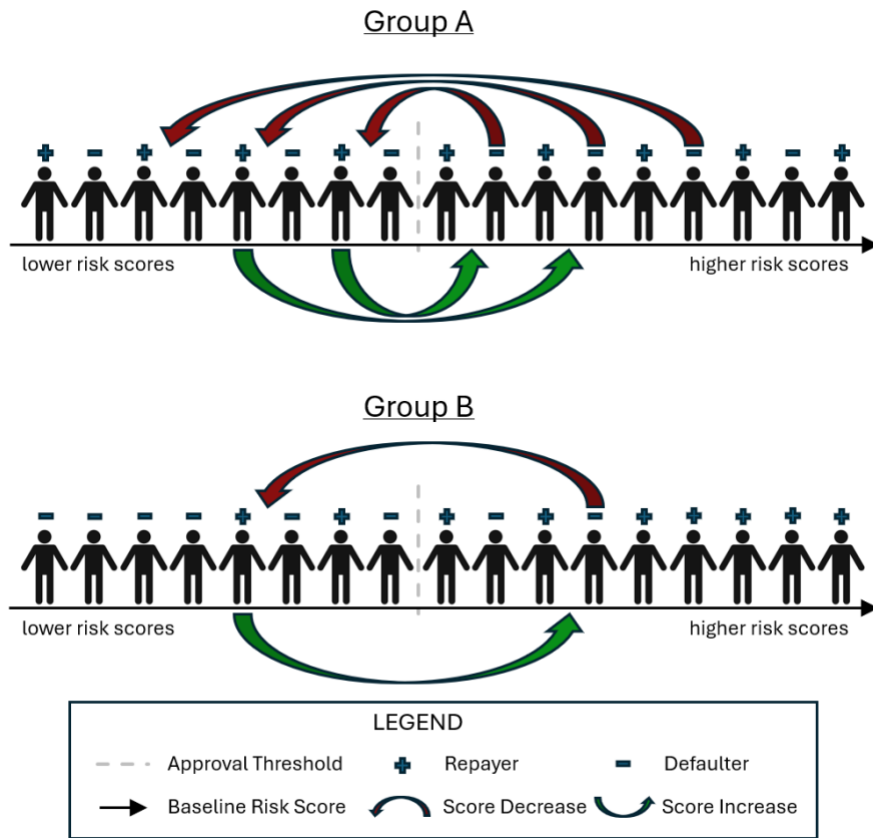


Figure 2: The Tradeoff Between Equal Outcomes and Equal Validity

To illustrate this tension graphically, we return to the example from Figure 1 and consider, in Figure 2, the introduction of a new lending policy, for example a new credit risk tool or new data to be used in underwriting. As before, the effects of the existing lending policy are shown by ordering loan applicants in each of two groups based on their perceived creditworthiness (e.g., as measured by a risk score). Applicants to the right of the cutoff denoted by the dashed line in the figure are approved for a loan. The effects of the new lending policy are shown by the red and green arrows: under this policy, some applicants are perceived to have higher creditworthiness than before (green arrows) and some are perceived to have lower creditworthiness than before (red arrows).

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.)

The figure illustrates how equal outcomes and equal validity can be in tension. Under the new lending policy, Group A’s approval rate declines: three previously approved applicants are newly rejected, while only two previously rejected applicants are newly approved. However, validity for Group A is enhanced: more of the applicants who would later repay a loan (denoted “+”, as before) are approved. The new lending policy thus addresses the disparity in validity that was evident in Figure 1, while introducing inequality in approval rates.

To more concretely demonstrate the possible tradeoffs between increasing outcome equality and decreasing differential validity, we develop an empirical example on simulated data. The example illustrates quantitatively and graphically how certain models or underwriting criteria introduce some quantum of additional outcome inequality at the potential benefit of reducing disparities in validity.

This empirical example focuses on a lender’s loan approval decisions. The lender faces a new data source to potentially use to make these decisions and needs to evaluate whether using the data in a credit risk model is permissible under fair lending regulation. The new data source can variously be understood as any of the variables not traditionally used in US underwriting that lenders increasingly have a new ability to access, for example cash-flow data from consumer bank accounts, a history of utility bill payments, or a history of payday loan use.¹⁰⁹ These variables in the simulation exhibit different levels of correlation with protected class, and they also differ in the extent to which they are informative about default for different groups; we simulate 81 such potential variables. To make loan approval decisions using a new variable, the lender develops (or equivalently, “trains” or “estimates”) a predictive model using data on past loan applicants. Once developed, the model can use similar input data from future loan applicants to predict their default probability, and the lender approves loans for all applicants with predicted default probabilities below some threshold. Further details about the simulation can be found in Appendix A.

Figure 3 shows the change in differential validity and the change in outcome inequality for each of these 81 potential variables, relative to the simulated lender’s existing credit risk model. For example, the first of these 81 potential variables might increase the gap in approval rates between two groups by 0.5 percentage points and might decrease the gap in validity between the two groups by 1.2 percentage points; the next variable might induce a different change in differential validity and in outcome inequality; and the third variable might induce still different changes, and so on. Which of these new variables should be permissible for the lender to use? The red (dotted) line in Figure 3 and the green (dashed) line in the figure show two potential answers.

Under fair lending regulation that focuses exclusively on outcome equality—denoted by the red (dotted) line—any new lending policy that exacerbates outcome inequalities such as unequal approval rates is suspect, modulo a business-necessity defense.¹¹⁰ In other words, any new variable located below the red (dotted) line is one a lender may hesitate to incorporate in a new scoring model

¹⁰⁹ For example, given the correlation of payday loan use with racial and ethnic group membership, a simulated data feature that corresponds to past payday loan use and performance might be positively correlated with minority status and might have higher variance for members of a minority group. *See generally* Paige Marta Skiba & Jeremy Tobacman, *Do Payday Loans Cause Bankruptcy?*, 62 J. L. ECON. 485-519 (2019); Mario L. Small et al., *Banks, Alternative Institutions and the Spatial-Temporal Ecology of Racial Inequality in U.S. Cities*, 5 NATURE HUM. BEHAV. 1622-8 (2021).

¹¹⁰ *See* Section III.A.3, *infra* (discussing the notion of performance trade-offs in the context of Upstart’s underwriting); Section III.D, *infra* (discussing how accounting for performance trade-offs broadens the notion of disparate impact beyond its historically binary treatment).

under fair lending’s current focus on outcome equality. Below we consider what models might be compliant if a regulator considered differential validity in addition to outcome equality.

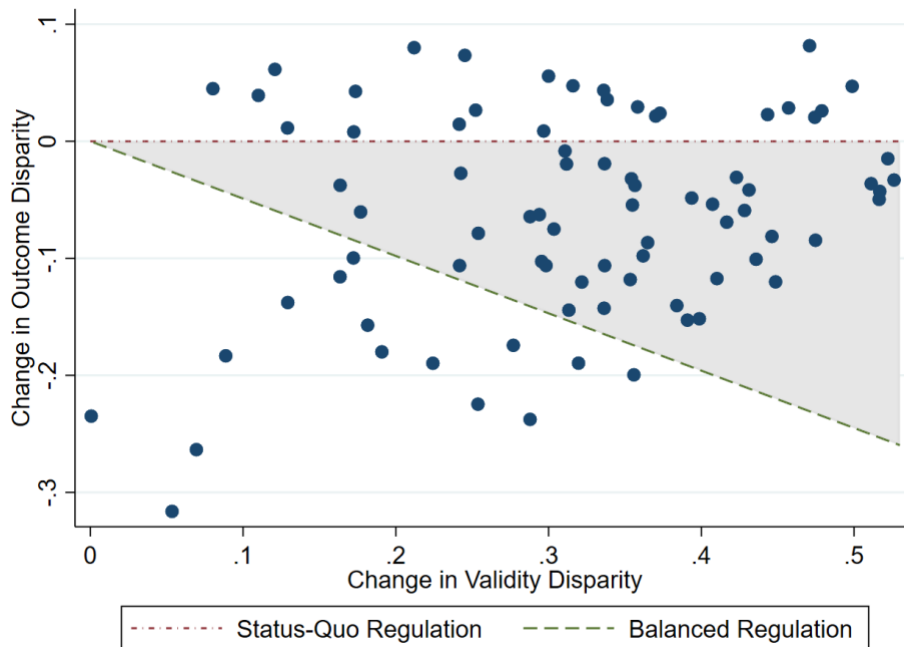


Figure 3: Balancing Differential Validity and Outcome Inequality

2. Balancing Equal Outcomes and Validity

Figure 3 shows that many lending policies that increase outcome inequality nonetheless present considerable potential gains in terms of reducing disparities in model validity. If there are tradeoffs between the goal of decreasing disparities in outcomes, on the one hand, and decreasing differential validity, on the other hand, what should be the obligations of a lender in selecting a lending model?

We argue that given the importance of both disparity measures, fair lending regulation should be reconceptualized to explicitly recognize the tradeoff between multiple disparity metrics. The sloped, green line in Figure 3 provides one graphical representation of a new regulatory policy that does so, by explicitly and quantitatively trading off between the two types of disparity. The slope of the line can be an explicit lever of fair-lending policy: a regulator can decide how to balance these two competing goals by calibrating the slope of the line. A line with zero slope—which places no value on reducing differential validity—illustrates how a status-quo regime focused exclusively on outcome equality is at an extreme of potential fair lending approaches within this policy space. The shaded area of the figure illustrates the space of models that are newly compliant as the regulator adjusts this slope. The key takeaway from the sloped line is that a regulator can facilitate the use of alternative models that alleviate other sources of unfairness in lending besides unequal outcomes.

A formalization of the balancing of unequal outcomes and differential validity, on which we provide more detail in Appendix B, allows the lender to know from the outset whether its underwriting is fair-lending compliant when selecting among underwriting models, rather than determining compliance in an ex-post regulatory compliance review.

3. Importance of the Equal Outcomes-Validity Tradeoff with Big Data

The need to consider differential validity in evaluating discriminatory lending practices has become increasingly important as lenders incorporate alternative data that have not traditionally been used in credit decisions. The inclusion of additional data, such as educational backgrounds and cash-flow information, in creditworthiness assessments should not only be evaluated based on approval rates but also on the extent to which these additional data sources reduce differential validity for protected group applicants.

Several claims have been made regarding the ability of alternative data to enhance creditworthiness assessments by lenders in recent years. One prominent example is Upstart, an alternative lender that uses educational data, such as SAT scores, to screen potential borrowers.¹¹¹ Although Upstart faced criticism for potentially exacerbating unequal outcomes through the use of SAT scores,¹¹² the company argued that additional educational data allowed for the more accurate assessments of credit risk, especially for borrowers lacking a traditional credit history.¹¹³

Similar arguments have been made with respect to the use of utility bill payment history in consumer credit.¹¹⁴ The addition of utility bill payment history is promising for reducing differential validity and improving the signal-to-noise ratio in credit files, but there are well-documented racial disparities in utility bill delinquency and default, arguably due to exogenous forces such as energy unreliability and shoddy infrastructure.¹¹⁵ So, even though utility bill payment history can be highly informative in the aggregate, thus drastically improving differential validity, the data may introduce nontrivial inequality in approval rates or other outcomes that tracks protected characteristics, like race.

Another example is the inclusion of payday loan repayment data in credit files. Traditionally, payday loan repayment data has been excluded from mainstream credit assessments. Recent debates suggest that excluding payday loan data may disadvantage vulnerable consumers. Although payday loan usage can predict future loan defaults, it is also strongly correlated with membership in protected

¹¹¹ The CFPB issued a NAL under ECOA, to Upstart in 2017. CFPB, *CFPB Announces First No-Action Letter: Upstart Network* (Sept. 14, 2017), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-announces-first-no-action-letter-upstart-network/>. Upstart's NAL terminated, by its own request, in 2022. CFPB, *CFPB Issues Order to Terminate Upstart No Action Letter* (June 8, 2022), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-order-to-terminate-upstart-no-action-letter/>.

¹¹² Katherine Welbeck & Ben Kaufman, *Fintech Lenders' Responses to Senate Probe Heighten Fears of Educational Redlining*, STUDENT BORROWER PROTECTION CENTER, (July 31, 2022), <https://protectborrowers.org/fintech-lenders-response-to-senate-probe-heightens-fears-of-educational-redlining>. In particular, only 4% of HBCUs had an average SAT score in the top 50% of Upstart's SAT groupings.

¹¹³ Di Maggio et al., *supra* note 14.

¹¹⁴ Brianna McGurran, *How Utility Bills Can Boost Your Credit Score*, EXPERIAN (Aug. 16, 2023), <https://www.experian.com/blogs/ask-experian/does-paying-utility-bills-help-your-credit-score/>.

¹¹⁵ See generally *Taylor v. City of Detroit*, No. 2:20-cv-11860-SDD-APP, 2020 WL 3883610 (E.D. Mich. July 9, 2020).

classes, particularly racial or ethnic minorities.¹¹⁶ Incorporating payday loan repayment data into mainstream credit scoring—a process already underway with at least two of the three major U.S. credit bureaus—could improve the predictive accuracy of credit scores while simultaneously exacerbating disparities between white and non-white loan applicants.

These examples—educational data, utility payment history, and payday loan repayment data—illustrate that expanding the data used to assess creditworthiness can create a trade-off between equality of outcomes and model validity. Therefore, evaluating the desirability of using alternative data necessitates quantifying this trade-off to realize the potential of improvements in validity, while also appropriately balancing the goals of equal validity and equal outcomes across groups. Understanding and addressing this trade-off is crucial for developing fair and equitable lending practices in the era of big data.

III. IMPLEMENTING THE BALANCE BETWEEN EQUAL OUTCOMES AND VALIDITY

This section discusses the challenges and opportunities in implementing a balancing approach to disparities, as well as the broader implications for other aspects of fair lending that require modernization. Given that most fair lending enforcement occurs through regulatory oversight rather than private litigation, we start by exploring how a balanced approach can be incorporated through LDA requirements or other regulatory mechanisms, such as no-action letters and regulatory sandboxes.

Next, we examine key challenges for our approach, including the permissibility and availability of protected characteristics and strategies to address the issue of missing counterfactual loan outcomes.

Finally, we discuss the implications of our proposed shift towards balancing competing fair lending goals for other areas where traditional fair lending doctrine needs updating. This includes the necessity of balancing performance with disparities and managing trade-offs among multiple protected groups.

A. Legal and Institutional Implementation

1. Fair Lending’s Reliance on Public Enforcement

Although both ECOA and FHA grant consumers a private right of action, in practice disparate impact cases are rarely litigated in courts. Courts generally do not consider a group-level disparity to constitute a “concrete injury” redressable in a private ECOA or FHA action.¹¹⁷ A group-level disparity in approvals and rejections is a mere “abstract” injury because it does not indicate that any specific

¹¹⁶ See generally Skiba & Tobacman, *supra* note 109; Small et al., *supra* note 109.

¹¹⁷ A creditor is broadly defined to be “a person who, in the ordinary course of business, regularly participates in a credit decision, including setting the terms of the credit.” 12 CFR § 1002.2(l). To sue a creditor for discrimination under ECOA, the consumer “must have (1) suffered an injury in fact, (2) that is fairly traceable to the challenged conduct of the defendant, and (3) that is likely to be redressed by a favorable judicial decision.” *Spokeo, Inc. v. Robins*, 136 S.Ct. 1540, 1547 (2016). The standing analysis often short-circuits at whether there exists a real problem in the first place, i.e., an “injury in fact” that must be “concrete and particularized” rather than “abstract.” *Id.* at 1548 (quoting *Lujan v. Defenders of Wildlife*, 504 U.S. 555, 560, 112 S.Ct. 2130, 119 L.Ed.2d 351 (1992)).

individual plaintiff “suffered an injury that was the result of . . . discrimination at the time that she entered the loan.”¹¹⁸ Accordingly, it is unlikely that a court would confer standing for a private ECOA action merely on the basis of group-level disparities in differential inaccuracy.

When considering the type of harm that arises from differential validity, the consequences of loan denial for a creditworthy borrower, or of loan nonpayment for subsequent defaulter can unfold over a lifetime. The process of collecting on an unpaid debt can take years or, in cases where the consumer may inadvertently reaffirm the debt during the debt collection process,¹¹⁹ can even take decades.¹²⁰ Moreover, because it may be unobservable what a consumer’s fully accurate score would have been in the absence of differential validity, harms from differential validity may inherently only be quantifiable at the group level rather than the individual level.¹²¹ Being disadvantaged by differential validity in credit scores may therefore lead to unobservable risk of later harm, creating barriers for private action.¹²²

Plaintiffs also face strong difficulties with attaining class certification, particularly with respect to establishing the “commonality” in harm and questions of law and fact.¹²³ Courts require plaintiffs to show evidence licensing an inference that members of the class suffered from a common policy of

¹¹⁸ *Gilmore v. Ally Fin. Inc.*, No. 15-CV-6240 (RER), 2017 WL 1476596, at *6 (E.D.N.Y. Apr. 24, 2017). *Cf. Masudi v. Ford Motor Credit Co.*, No. 07-CV-1082 (CBA)(LB), 2008 WL 2944643, at *4 (E.D.N.Y. July 31, 2008) (dismissing an ECOA complaint because it “d[id] not allege a single event, policy or action taken by either of the defendants regarding plaintiff’s car loan, nor d[id] plaintiffs allege how the finance charges imposed on their loan were discriminatory”) (emphasis in original).

¹¹⁹ *See, e.g., Berthoud Nat. Bank v. Dunn*, 762 P.2d 759 (Colo. Ct. App. 1988) (“[An] express promise to pay a debt acts to take the case out of the statute of limitations.”).

¹²⁰ The fact that the harm materializes after the borrower receives the loan and not while being an applicant does not seem to be a barrier to private action. With respect to who qualifies as an “applicant” under ECOA, the CFPB has repeatedly argued that this term “is not expressly limited to those currently in the process of seeking credit” but instead “include[s] both those who are currently seeking credit and those who sought and have now received credit.” *See, e.g.*, Brief for the CFPB as Amicus Curiae, p. 4, *Fralish v. Bank of Am., N.A.*, No. 3:20-CV-418 RLM-MGG, 2021 WL 4453735 (N.D. Ind. Sept. 29, 2021), *appeal dismissed*, No. 21-2846, 2022 WL 1089194 (7th Cir. Jan. 28, 2022). This distinction matters because discriminatory practices can happen even after the application is closed and loan money is disbursed, such as later revoking the credit on a prohibited basis. Recognizing the risk of future harm as an injury-in-fact is also contiguous with practice in ECOA’s touchstone of employment law—the Supreme Court has held 9-0 that “employees” protected against discrimination in Title VII includes those who are no longer employees when the discrimination occurs. *Robinson v. Shell Oil Co.*, 519 U.S. 337, 341 (1997). Justice Thomas, writing for the majority, noted that there is “no temporal qualifier in the statute such as would make plain that § 704(a) protects only persons still employed at the time of the retaliation.” *Id.*

¹²¹ In other words, economic analysis is able to quantify differential validity across groups by showing that a credit score has greater statistical noise for one group than for another (Blattner & Nelson, *supra* note 16), but it may be unobservable which particular individuals in the group faced a score that was “too high” versus “too low.”

¹²² And in 2021, the Supreme Court held that “in a suit for damages, the mere risk of future harm, standing alone, cannot qualify as a concrete harm.” *TransUnion LLC v. Ramirez*, 141 S.Ct. 2190, 2209-10 (2021). So, even though group-level inequities raise the risk of future harm, such as an individual being denied a loan and unable to access the concrete things the loan would have provided, the probabilistic inequities alone will not confer standing to a private consumer under ECOA. At the very least, getting a fuller picture of credit discrimination will require broadening the time scale to recognize the longer-term harms imposed by differential validity. In particular, the trouble with FPs is that they are not free—they are *inefficient approvals*. Because an FP is a defaulted loan, minimizing FPs means sparing consumers from the process of loan delinquency and eventual default, which imposes costs stemming from various sources such as late fees, predatory interest rates, and possible court expenses. *See* Blattner & Nelson, *supra* note 16, for a general approach to quantifying the costs of noise.

¹²³ *See Cason v. Nissan Motor Acceptance Corp.*, 212 F.R.D. 518, 521 (M.D. Tenn. 2002).

discrimination,¹²⁴ but it seems creditors commonly argue successfully that plaintiffs cannot satisfy the commonality prong. The failure generally turns on the individual nature of the issues, such as the discretion of individual employees that make decisions on credit applications.¹²⁵ Additionally, in opposing class certification in discrimination class actions, counsel for creditors commonly argue that individual issues concerning each class member’s experiences with the creditor predominate over common questions of law and fact, rendering class resolution inappropriate.¹²⁶ The capping of punitive damages can also create barriers for class actions.¹²⁷

The conclusion is that fair lending, and disparate impact in particular, is enforced primarily through public enforcement. Most complaints originate with the CFPB or other agencies, such as the Department of Justice, which bring lawsuits against institutions for discrimination. These complaints often result in consent orders. Supervisory activities also play a key role in overseeing lender compliance efforts. In the next section, we will discuss how a weighting approach can be implemented by lenders and overseen by regulators through the analysis and the search for less discriminatory alternatives (LDAs).

2. Implementing Disparity Measuring in LDA analysis

The balancing of disparity measures is likely to be most meaningful in the context of LDA search and analysis. As discussed in Section I.A.1, a lender policy justified by business necessity is still impermissible if a “less discriminatory alternative” meets the business need. Although this third and last stage of disparate impact analysis has existed for decades in fair lending and other domains, it has traditionally received less attention and little regulatory guidance.¹²⁸ Given that it was considered the plaintiff’s burden to demonstrate an LDA, it is unsurprising that third parties have been limited in their ability to identify alternatives.

Recent regulatory developments suggest that the search for alternatives will become the duty of lenders, with fair lending compliance and regulatory supervision including analysis for whether alternatives exist. The CFPB recently published its annual Fair Lending Report for 2023, discussing LDAs for the first time. According to the CFPB, compliance with fair lending law requires lenders “to develop a process for the consideration of a range of less discriminatory models.”¹²⁹ The CFPB

¹²⁴ See *Love v. Jobanns*, 439 F.3d 723, 728 (D.C. Cir. 2006).

¹²⁵ See, e.g., *In re Countrywide Fin. Corp. Mortg. Lending Practices Litig.*, 708 F.3d 704, 707 (6th Cir. 2013); *Rodriguez v. Nat’l City Bank*, 726 F.3d 372, 381-386 (3d Cir. 2013); *In re Wells Fargo Residential Mortg. Lending Discrimination Litig.*, No. 08-WL-1930 (MMC), 2011 WL 3903117, at *1-4 (N.D. Cal. Sept. 6, 2011).

¹²⁶ See *Love*, 439 F.3d at 726.

¹²⁷ Although ECOA authorizes the recovery of class action damages for discrimination (15 U.S.C. § 1691e(b).), class action filings in general, and those for discrimination especially, seem uncommon because the maximum class recovery is low. ECOA caps the punitive damages that a plaintiff class may receive in an ECOA class action at \$500,000 or 1% of the net worth of the creditor, whichever is less (15 U.S.C. § 1691e(b)). Plaintiffs can also recover actual damages. But because plaintiffs must specifically prove their actual damages (See *Anderson*, 666 F.2d at 1277-78.), courts are unlikely to certify classes seeking recovery of actual damages due to the individualized nature of that inquiry. See, e.g., *Coleman v. Gen. Motors Acceptance Corp.*, 296 F.3d 443, 447 (6th Cir. 2002) (certifying only an injunctive class); *Cason*, 212 F.R.D. 518, 521 (M.D. Tenn. 2002) (same).

¹²⁸ See Mahoney 1998, *supra* note 45, at 490; see also Black et al. 2024, *supra* note 45.

¹²⁹ Black et al. 2024, *supra* note 45.

has essentially elevated the LDA search, which has been lacking regulatory guidance, to a key component of fair lending, determining that lenders themselves must conduct this search.

As proactively searching for LDAs becomes a regulatory requirement on lenders, the measurement of disparities becomes central to this search. We therefore argue that our balancing framework, recognizing the explicit trade-offs between one disparity (differential validity) and another (equal outcomes), should become part of the LDA search. Unlike traditional LDA analysis that is “one-way” in the sense that it minimizes outcome equality violations among all sensible business choices, our framework allows the consideration of policies that may not be strictly superior on the outcome equality dimension but may create meaningful validity gains for protected groups.

One way to formalize LDA analysis, which we expand upon in Appendix B, is as an explicit constraint on lenders’ choices over underwriting models and data inputs. This provides a framework for evaluating alternative models that is consistent with the multiple goals of fair lending. Unlike the traditional approach to LDA, where this type of analysis takes place *ex post*, our framework allows lenders to efficiently trade-off between disparities from the outset. By implementing an LDA from the beginning, lenders can provide evidence of compliance to regulators.

Our proposal for a regulatory compliance regime aims to avoid the harms of outcome inequality and differential validity at the outset, rather than navigate the thorny distinctions between “abstract” and “concrete” injuries. Under the disparate impact interpretation of ECOA’s antidiscrimination provision, firms must comply by ensuring both validity and outcome equality. However, it is not obvious how much is enough, and firms may hesitate to use an underwriting strategy that could achieve valuable improvements in equality of outcomes or validity due to concerns about getting the balance wrong. If regulators such as the CFPB use our LDA constraint to describe exactly how to make this trade-off, firms will not need to guess or risk avoidable inefficiencies.

3. Implementation Disparity Analysis Through Other Regulatory Tools

Other regulatory tools also lend themselves to the types of balancing analysis of disparity measures we have outlined above. In recent years, the CFPB has launched several initiatives to encourage marketplace innovation for the benefit of consumers. One such initiative is the CFPB’s “No-Action Letters” (NAL) program.¹³⁰ A NAL, created as part of the CFPB’s Project Catalyst, allows lenders to apply for a regulatory NAL under ECOA. With a NAL, the CFPB refrains from bringing an action under ECOA for a limited period of time, so that the lender can implement a pre-approved experimental program. In 2019, the CFPB updated its NAL policy as it found its prior policy “not an adequate response to the extent of innovation occurring in markets for consumer financial products and services.”¹³¹ The new policy provides a more streamlined review process for products and

¹³⁰ The authorities given to the CFPB under Title X of Dodd-Frank form the basis for the CFPB’s original NAL policy. 81 FR 8686 (Feb. 22, 2016). The SEC also has a mechanism for NAL, which may be considered as a more case-by-case adjudication on the part of the regulation. Although Dodd-Frank does not explicitly provide the authority for NAL, they are considered under the CFPB’s supervisory and enforcement discretion.

¹³¹ CFPB, POLICY ON NO-ACTION LETTERS, 84 FED. REG. 48229 (Sept. 13, 2019) [hereinafter New NAL Policy]. The Final Rule can be found at: https://files.consumerfinance.gov/f/documents/cfpb_final-policy-on-no-action-letters.pdf and <https://www.govinfo.gov/content/pkg/FR-2019-09-13/pdf/2019-19763.pdf>.

services.¹³² It also no longer requires applicants to show that they are likely to provide “substantial” benefit to consumers but rather a “potential” benefit.¹³³

Under current procedure, lenders can apply for a NAL at any time and can expect to receive a decision within 60 days.¹³⁴ The CFPB reserves sole discretion in making this decision.¹³⁵ Regulation provides some examples of factors in the CFPB’s NAL analysis, such as “the quality and persuasiveness of the NAL request, with particular emphasis on the proposed products’ or services’ potential consumer benefits and risks,” “how the applicant intends to mitigate any potential risks,” “the applicant’s explanation of why the NAL is needed and the underlying statutory and regulatory provisions” and “potential litigation risk.”¹³⁶

The criteria used by the CFPB in assessing the risks and benefits of a new product are currently opaque,¹³⁷ as is demonstrated by the only NAL issued by the CFPB under ECOA, given to Upstart in 2017¹³⁸ with a follow-up letter issued in 2020.¹³⁹ Although the letters require Upstart to engage in discrimination testing, the methodology for and results of the testing are not publicly available. Upstart has faced significant criticism for its potential to exacerbate outcome equality, most notably through its use of SAT scores to screen potential borrowers.¹⁴⁰ In particular, only 4% of HBCUs had an average SAT score in the top 50% of Upstart’s SAT groupings.¹⁴¹ Upstart’s NAL terminated, by its own request, in 2022,¹⁴² leaving the legality of using SAT data to remain uncertain. Beyond the criticism

¹³² For example, the new policy states the intention of the CFPB to grant or deny an application within 60 days and includes a policy for obtaining NAL modifications. *Id.*

¹³³ *See* New NAL Policy. Since its adoption of the new policy, the CFPB has provided NALs with respect to housing counseling agreements by Housing Counseling Agencies (HCA), which are regulated by the Department of Housing and Urban Development (HUD). The first NAL was issued in September 2019 and covered HUD supervised HCAs that enter agreements with consumers, as long as they comply with certain requirements such as providing consumers with a memorandum of understanding with the consumer reflecting the terms of the housing counselling. CFPB, No-Action Letter to U.S. Dep’t of Hous. & Urban Dev. (Sept. 10, 2019), https://files.consumerfinance.gov/f/documents/cfpb_HUD-no-action-letter.pdf. It provided a similar NAL to Bank of America in January 2020. *See* CFPB, No-Action Letter to Bank of America (Jan. 10, 2020), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-no-action-letter-to-facilitate-housing-counseling-services/>.

¹³⁴ New NAL Policy at 48229-01.

¹³⁵ *Id.*

¹³⁶ *Id.*

¹³⁷ For example, what makes for a request high in “quality”? What constitutes “persuasiveness” in assessing consumer benefits and risks? *See id.*

¹³⁸ CFPB, CFPB Announces First No-Action Letter: Upstart Network, (Sept. 14, 2017), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-announces-first-no-action-letter-upstart-network/>. For the general policy, *see* CFPB, POLICY ON NO-ACTION LETTERS, 81 FR 8686 (Feb. 22, 2016).

¹³⁹ 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.

¹⁴⁰ Katherine Welbeck & Ben Kaufman, *Fintech Lenders’ Responses to Senate Probe Heighten Fears of Educational Redlining*, STUDENT BORROWER PROTECTION CENTER, (July 31, 2022), <https://protectborrowers.org/fintech-lenders-response-to-senate-probe-heightens-fears-of-educational-redlining>.

¹⁴¹ *Id.*

¹⁴² CFPB, CFPB Issues Order to Terminate Upstart No Action Letter, (June 8, 2022), <https://www.consumerfinance.gov/about-us/newsroom/cfpb-issues-order-to-terminate-upstart-no-action-letter/>.

that Upstart’s methodology creates outcome inequality, there remains a question of whether Upstart’s use of educational data improves validity disparities. As Upstart is meant to cater to populations that do not have traditional credit files, it may have the potential to decrease validity disparities. In fact, the 2020 updated NAL requires Upstart to report to the CFPB on the model’s predictive accuracy by group.¹⁴³

The NAL program could be a more robust regulatory tool if it is scaled-up and explicit in the type of disparity testing that is required by adopting our balancing approach to disparities. To view NALs as opportunities to test specific, quantified trade-offs between outcome equality and validity would provide a clear set of criteria for both lenders and regulators. Clarifying regulatory expectations in this way would encourage firms to develop new innovations and apply for NALs with a better idea of their business will be impacted. It is precisely the potential of innovation to address longstanding shortcomings of traditional credit scoring that requires that it be evaluated based on how it improves outcome equality *and* validity.

The utilization of NAL could still include current safeguards against its abuse as a way to avoid regulation. Every NAL could still be for a limited, specified duration and subject to continuous review.¹⁴⁴ These safeguards help to mitigate risk from unequal outcomes, as exemplified in a 2020 Senate letter arguing that “the CFPB should not issue ‘No Action’ letters related to ECOA to any lender or company . . . where Black and brown borrowers unwillingly serve as the test subjects.”¹⁴⁵ Granting NALs to one firm at a time allows for an incremental, piecemeal rollout of temporary new credit scoring techniques. This would, in turn, minimize “unwilling service” as test subjects by mostly preserving the status quo. Ideally, an experiment benefits everyone by more efficiently matching lenders with qualified applicants.

Another tool that lenders could potentially use for balancing equal outcomes and validity is by creating a Special Purpose Credit Program (SPCP). SPCPs are initiatives designed to provide credit to economically or socially disadvantaged groups that may otherwise face barriers in accessing credit.¹⁴⁶ These programs are authorized under the Equal Credit Opportunity Act (ECOA) and allow creditors to offer more favorable terms to eligible borrowers to promote fair lending and financial inclusion.¹⁴⁷ SPCPs are tailored to address specific needs, such as increasing homeownership among low-income communities or supporting minority-owned businesses, thereby helping to reduce disparities in credit access and promoting economic equality. SPCPs can be motivated by the need to target lending for certain groups in ways that might be in tension with equal outcomes but that promotes validity. For

¹⁴³ 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 (The Model Risk Assessment Plan requires Upstart to “Test 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 Bureau.”).

¹⁴⁴ CFPB, POLICY ON NO-ACTION LETTERS, 84 FR 48229-01 (2019).

¹⁴⁵ Sen. Sherrod Brown, Sen. Elizabeth Warren, & Sen. Kamala Harris, Letter to CFPB (July 30, 2020), https://www.banking.senate.gov/imo/media/doc/2020-07-30_Letter%20to%20CFPB%20re%20use%20of%20educational%20data.pdf.

¹⁴⁶ The CFPB reiterated in 2020 that lenders can establish SPCPs without formal regulatory approval so long as the program “is not administered with the purpose of evading the requirements of [ECOA].” 12 C.F.R. §1002.8.

¹⁴⁷ Reg B instructs that ECOA permits lenders to implement SPCPs that explicitly favor members of historically disadvantaged protected classes. 12 C.F.R. §1002.8.

example, varying the credit score threshold for protected groups may violate outcome equality, such as conditional AIR, but may reduce validity inequality.¹⁴⁸

The permissibility of the explicit use of protected group status in credit decisions using SPCP has been called in to question in light of the Supreme Court’s recent holding in *Students for Fair Admissions, Inc.* The decision stuck down affirmative action in the context of university admissions as unconstitutional and a violation of Title VI on grounds that the challenged programs explicitly favor protected class membership, such as race.¹⁴⁹ However, the SCPC differs from university admissions in several crucial ways.¹⁵⁰ Congress outlined safeguards for SPCPs over half a century ago, specifically designed to satisfy a strict scrutiny analysis under Equal Protection, such as one implicating race.

SCPCs include several safeguards to make sure they are narrowly tailored to their legislative purpose. For example, an SPCP administrator must “identif[y] the class of persons that the program is designed to benefit and sets forth the procedures and standards for extending credit pursuant to the program.” Importantly, the goal of any SPCP is to provide credit to people “who, under the organization’s customary standards of creditworthiness, probably would not receive such credit or would receive it on less favorable terms than are ordinarily available to other applicants[.]”¹⁵¹ These goals are consistent with a SPCP considering validity.

Using a framework, such as the one in Appendix B, would allow lenders greater certainty in establishing whether a SPCP meets fair lending goals. The current standard-like¹⁵² nature of SCPC mandates might be deterring firms from establishing SPCPs, who may seek to address one lending challenge that could impact another. For example, remedying the issue of unbanked consumers for which there is too much negative data on a credit report may hurt those with too little positive data. Quantifying the trade-off *ex ante* using our LDA constraint will allow both firms and regulators to benefit from a more predictable and streamlined assessment of fairness in lending.

In sum, trying different balances in a streamlined, quantifiable, and flexible way is a promising path forward to addressing systemic inequalities in access to credit. In what follows, we discuss some challenges of expanding disparity metrics to include validity.

¹⁴⁸ See Meursault et al., *supra* note 108.

¹⁴⁹ *Students for Fair Admissions, Inc.*, 600 U.S. 181 (2023).

¹⁵⁰ For example, SCPCs are meant to address historical discrimination and not to create diversity. See generally Brad Blower, *Special Purpose Credit Programs Remain on Solid Legal Ground Despite Supreme Court’s Affirmative Action Decision*, NCRC (Sept. 11, 2023), <https://ncrc.org/special-purpose-credit-programs-remain-on-solid-legal-ground-despite-supreme-courts-affirmative-action-decision/>.

¹⁵¹ 12 C.F.R. §1002.8.

¹⁵² Legal commands are often characterized as either bright-line rules (e.g., “do not exceed 70 miles per hour on the highway”) or relatively more ambiguous standards (e.g., “do not drive recklessly on the highway”). See generally Louis Kaplow, *Rules Versus Standards: An Economic Analysis*, 42 DUKE L.J. 557-629 (1992). Our proposal can be interpreted as a procedure for ‘rulifying’ LDA analysis under ECOA, where such analysis operates more like a standard in current practice, without enumerated thresholds for making the relevant trade-offs. However, the LDA constraint does not necessarily yield one particular rule—for example, every point in the shaded area of Figure 1 constitutes a different rule for achieving an LDA that satisfies our constraint.

B. Challenges in Expanding Disparities Metric

In this section, we address several challenges in implementing a disparities approach that balances outcome equality and validity. First, we discuss a well-known issue in the algorithmic fairness literature: estimating error rates with missing outcomes or labels. When models are used for decisions, such as lending, the loan outcome—whether the borrower defaults—is only observed if the loan is approved. We lack information on the loan outcomes for denied applicants, creating a significant challenge in estimating the accuracy for these individuals. We begin by exploring this challenge and potential solutions.

Next, we examine two implementation challenges that, while not unique to our approach, are relevant for the explicit tradeoff of our framework. The first challenge is estimating group disparities when lenders do not collect or report demographic information. The second challenge involves explicitly considering and adopting models based on group characteristics, which may raise issues related to other discrimination doctrines, such as disparate treatment.

1. Missing Outcome Data Loan Outcomes

To implement our balancing approach that considers 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 how rejected loan applicants would have performed if granted a loan.¹⁵³ We propose four approaches for assessing these counterfactual if-not-rejected loan outcomes: a surrogate-outcomes strategy from other loan products;¹⁵⁴ outcomes observed with competing lenders, an approach which incentivizes competitors to lend to each other's rejected applicants; an approval lottery implemented as part of fair-lending compliance exams; and a “stress testing” approach. We illustrate these ideas with a concrete example using simulated data on consumer lending inputs and outcomes.

The most easily implementable approach is to use default on other loan products to assess default risk on denied loans. For example, an applicant rejected for a credit card might have an existing student loan on which subsequent default is straightforwardly observed in existing U.S. consumer credit report data. This approach does not require that *every* borrower who would have defaulted on, for example, their rejected credit card loan will also default on their student loan. Rather, this only requires that, on average, group-level default rates on one type of loan are informative about average default rates on other loans in a known way. This approach can be particularly successful in contexts like mortgage lending, where the vast majority of rejected applicants have other, non-mortgage loans on which subsequent default can be observed.¹⁵⁵ However, this approach may also be vulnerable to confounds from selection effects, as rejected borrowers from certain lenders may select endogenously

¹⁵³ 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).

¹⁵⁴ *See also* Blattner & Nelson, *supra* note 16.

¹⁵⁵ *See id.* for one such implementation.

into different alternative credit products than rejected borrowers from other lenders, such that differences in prevailing default rates across alternative products would imply different true-negative rates and false-negative rates in ways that make the LDA constraint spuriously tighter for some lenders than for others within a given market.¹⁵⁶ Each of our other proposals below uses different strategies to address such endogeneity concerns.

A second approach is to assess whether one lender's rejected loans are true negatives or false negatives using loans from competing lenders who provided the same loan product. Like the first approach, this is also readily implementable using existing U.S. consumer credit report data. However, this approach would have the additional advantage of incentivizing competitors to lend to each other's rejected applicants, in order to reveal to the regulator cases where a competitor's rejected loans were false negatives. This approach would not require an unbiased estimate of true-negative and false-negative rates, and in that way, this approach is robust to the selection concern we noted in our first approach: in our second approach each lender would face similar efforts by its competitors in a given market to lend to its rejected applicants, and the effects of these competitive pressures would imply that estimates of true-negative and false-negative rates would have similar statistical properties across different lenders in a market. Even in the absence of an unbiased estimates of true-negative and false-negative rates, the constant term " \mathcal{A} " in the LDA constraint would absorb such effect, making the LDA constraint comparable across lenders within market.

A third approach is more costly to implement but uses gold-standard methodology for classifying a lender's rejected applicants: fair lending exams can require small-scale randomized trials in which rejected applicants are randomly granted loans. By construction, these trials are robust to selection concerns and provide unbiased estimates of true-negative and false-negative rates. Regulators could require that defaults on these loans not be furnished to credit bureaus and that defaulted loan balances not be pursued in collections, such that any consumers affected by a loan default in these randomized trials to not endure other hardship as a result of that default. And to be clear, these loans would only be provided to consumers who had *applied* for them, assuaging concerns about unwilling participation in the evaluation. The advantage of this approach would be an unbiased estimate of rejected applicants' positive or negative class membership that requires no extrapolation from other loans' performance. Moreover, while defaults on the randomly granted loans would generate some costs for lenders, this kind of randomized testing is already used at many lenders to assess underwriting criteria, and any costs net of testing the lender already implements could either be subsidized directly or could be deferred through other revenue streams. Considering the ample resources that lenders evidently devote to fair-lending compliance, this approach may in fact lead to cost *savings* at many lenders even after accounting for loan defaults in the randomized acceptances.

A fourth approach would build on Gillis and Spiess's proposal of "stress testing" an underwriting model.¹⁵⁷ Rather than using real-world data to evaluate factors like the outcome equality and differential accuracy induced by a given model, a regulator could deploy a lender's underwriting

¹⁵⁶ For example, suppose one mortgage lender's rejected applicants' other credit products are predominantly auto loans, and another mortgage lender's rejected applicants' other credit products are predominantly credit cards. For any given borrower, default rates on credit cards are typically higher than default rates on auto loans, so even if these two lenders' applicant pool and underwriting decisions were otherwise identical, the former lender would erroneously be inferred to have more true negatives and fewer false negatives than the latter lender. See Sewin Chan et al., *Determinants of Mortgage Default and Consumer Credit Use: The Effects of Foreclosure Laws and Foreclosure Delays*, 732 FED. RESERVE BANK N.Y. STAFF REP. (2015).

¹⁵⁷ Gillis & Spiess 2019, *supra* note 4.

practices on a pre-determined dataset or a range of hypothetical applicant scenarios. This approach is similar to how banks' safety-and-soundness regulators use stress tests to evaluate how bank assets and capital ratios respond to various hypotheticals. Relative to our three other approaches, this "stress test" approach would entail a different set of costs and benefits for both the regulator and the lender; these costs and benefits include the need for the underwriting to be sufficiently algorithmic that it could be deployed on *hypothetical* loan applicants, the simplicity of not using data from many different banks' actual applicant pools, and the costs for the regulator to develop a battery of stress tests that are sufficiently representative of the range of real-world applicants a lender might face.

2. Missing Data on Demographics

The estimation of racial and other disparities relies on the availability of demographic information. However, in many settings there is a lack of demographic information at the individual level, often referred to as the "missing data problem."¹⁵⁸ This problem arises when lenders do not collect or report information about customers' race or other protected characteristics. When there is no information about the protected characteristics of customers, it is difficult to consider the disparate impact of policies.¹⁵⁹ This means that discrimination can go undetected and discrimination law unenforced.

The missing data problem is prevalent for most types of consumer financial products, with one important exception. The Home Mortgage Disclosure Act (HMDA) requires lenders to collect data on mortgage applicants' race and to disclose this information,¹⁶⁰ facilitating analysis of mortgage market disparities and administrative action and litigation.¹⁶¹ For other types of lending there is a general prohibition on asking applicants about their race, color, religion, national origin or sex,¹⁶² although Regulation B does allow for the collection of demographic information for self-testing and compliance,¹⁶³ or for purpose of determining SPCP eligibility.¹⁶⁴ Lenders themselves may be hesitant to collect this information, either because they might try to avoid liability by claiming that it is not

¹⁵⁸ See Da Lin, *Missing Data and Anti-Discrimination Laws*, HARV. L. REV. BLOG (Apr. 2, 2018), <https://harvardlawreview.org/blog/2018/04/missing-data-and-anti-discrimination-laws/> (discussing the missing data problem in insurance and lending).

¹⁵⁹ This issue was also raised in the context of equal pay. According to a Presidential Memorandum from 2014, "[e]ffective enforcement of this mandate, however, is impeded by a lack of sufficiently robust and reliable data on employee compensation, including data by sex and race." The White House, Presidential Memorandum—Advancing Pay Equality Through Compensation Data Collection (April 8, 2014), available at: <https://obamawhitehouse.archives.gov/realitycheck/the-press-office/2014/04/08/presidential-memorandum-advancing-pay-equality-through-compensation-data>.

¹⁶⁰ 12 CFR § 1002.13(a) ("Information to be requested. A creditor that receives an application for credit primarily for the purchase or refinancing of a dwelling occupied or to be occupied by the applicant as a principal residence, where the extension of credit will be secured by the dwelling, shall request as part of the application the information regarding the applicant's ethnicity, race, sex, marital status, and age required to be collected and reported")

¹⁶¹ See Ian Ayres, Gary Klein & Jeffrey West, *The Rise and (Potential) Fall of Disparate Impact Lending Litigation*, in EVIDENCE AND INNOVATION IN HOUSING LAW AND POLICY 231 (Lee Anne Fennell & Benjamin J. Keys eds., 2017).

¹⁶² 12 CFR § 1002.5(b) ("Except as otherwise provided in this paragraph, a creditor shall not inquire about or consider the race, color, religion, national origin, or sex of an applicant or any other person in connection with a credit transaction.")

¹⁶³ 12 CFR § 1002.15(b)(1) ("The results or report of a self-test...shall be privileged unless it is voluntarily disclosed by the creditor or the self-test is required by any governmental agency, or the creditor has failed to take corrective action.")

¹⁶⁴ 12 CFR § 1002.8(b) ("A creditor shall inform an applicant that it is requesting the information in connection with a special purpose credit program, and that it is required to collect the information to comply with the program.")

even aware of customers' protected characteristics,¹⁶⁵ or because of sensitivity regarding compliance with disparate treatment restriction as reflected in the Supreme Court's *Ricci* decision.¹⁶⁶

Regulators and analysts have developed methods to infer demographic information such as a race,¹⁶⁷ the most prevalent method being the Bayesian Improved Surname Geocoding (BISG). BISG combines individual names and addresses to predict race and has been used repeatedly in enforcement actions and supervisory activity.¹⁶⁸ Despite their prevalence, BISG methods have been criticized for the degree of inaccuracy of the prediction and their bias in estimating disparities.¹⁶⁹ As a result, several methods have been proposed to improve on BISG methods.¹⁷⁰

The challenges that arise from missing demographic information are not unique to our framework, but given that our method explicitly balances different disparity measures, an unbiased method of inferring demographic information is crucial. Improvements in prediction methods for demographic characteristics and pressures to require lenders to collect demographic information are likely to alleviate at least some aspects of the missing data problem. The CFPB has taken steps to expand the collection and reporting demographic information for certain lenders,¹⁷¹ and supervisory efforts create informal pressure on lenders to reliably demonstrate that their policies are not

¹⁶⁵ For example, Goldman Sachs, when accused that its Apple Card discriminates against women, relied on the defense that this is not possible because it does not collect information on gender. See Shahien Nasiripour, Jennifer Surance, and Sridhar Natarajan, *Apple Card's Gender-Bias Claims Look Familiar to Old-School Banks*, BLOOMBERG BUSINESSWEEK (Nov. 11, 2019, 3:01 PM), <https://www.bloomberg.com/news/articles/2019-11-11/apple-card-s-ai-stumble-looks-familiar-to-old-school-banks>.

¹⁶⁶ Although *Ricci* suggested that there might be tension between being race aware to avoid liability under the disparate impact doctrine, and Equal Protection (and likely disparate treatment), it is unclear how exactly *Ricci* should be interpreted. See Richard Primus, *The Future of Disparate Impact*, MICH. L. REV. 1341 (2009–2010).

¹⁶⁷ There have also been significant difficulties in using proxies for race. For further discussion of the use of proxies, see CFPB, *Using Publicly Available Information to Proxy for Unidentified Race and Ethnicity* (Summer 2014), https://files.consumerfinance.gov/f/201409_cfpb_report_proxy-methodology.pdf.

¹⁶⁸ Kosuke Imai & Kabir Khanna, *Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records*, 24 POLITICAL ANALYSIS 263–272 (2016); Nathan Kallus, Xiaojie Mao, & Angela Zhou, *Assessing Algorithmic Fairness with Unobserved Protected Class Using Data Combination*, 68 MANAGEMENT SCIENCE 1959–1981 (2021).

¹⁶⁹ See Cory McCartan et al., *Estimating Racial Disparities When Race is Not Observed* (Apr. 17, 2024), <https://arxiv.org/html/2303.02580v2> (showing that the race is correlated with the outcome of interest even after controlling for name and location). In other words, residuals are often correlated with outcomes of interest, which could lead to an under estimation of disparities.

¹⁷⁰ For example, the BIRDIE method that uses BISG probabilities as inputs to produce racial disparity estimates. See McCartan et al., *Estimating Racial Disparities When Race is Not Observed*, ARXIV:2303.02580 (2024), <https://arxiv.org/pdf/2303.02580>. See also REPUBLICAN STAFF OF THE COMM. ON FIN. SERVS., U.S. HOUSE OF REPRESENTATIVES, UNSAFE AT ANY BUREAUCRACY: CFPB JUNK SCIENCE AND INDIRECT AUTO LENDING 3 (Nov. 24, 2015), https://financialservices.house.gov/uploadedfiles/11-24-15_cfpb_indirect_auto_staff_report.pdf (“[T]he controversial statistical method the Bureau employed to measure racial disparities is less accurate than other available methods and prone to significant error . . .”). See also Charles River Associates (CRA) Report for Prepared for American Financial Services Association (AFSA) [hereinafter the CRA/AFSA Report], available at <http://www.crai.com/sites/default/files/publications/Fair-Lending-Implications-for-the-Indirect-Auto-Finance-Market.pdf>.

¹⁷¹ The CFPB has been working to implement Section 1071 of the Dodd-Frank Act, which requires financial institutions to collect and report demographic data on small business credit applications.

discriminatory.¹⁷² Other stakeholders also advocate for collection of demographic data for compliance purposes.¹⁷³ Lastly, some of the methods discussed in the previous section also allow for testing of disparities even when lenders do not collect this information on an ongoing basis. For example, discrimination stress-testing allows for lender trained models to be analyzed using a hold-out set that contains demographic information.

3. The Use of Protected Characteristics

An open question is the extent to which protected characteristics can be used in model training, debiasing, and deployment to mitigate disparities and when the use of these features can trigger other doctrines like disparate treatment. This tension arises from the need to balance fairness and anti-discrimination principles with effective model performance and regulatory compliance. In our setting demographic information is used directly to quantify disparities and select a model based on the balancing of different disparity measures. Moreover, a direct attempt to address outcome equality or validity through setting different lending thresholds at deployment may also be in tension with disparate treatment.¹⁷⁴

There have been differing views as to whether model the search for a LDA model explicitly considering demographics could violate fair lending laws. One view is that as long as the model's at deployment themselves are "blind" to protected characteristics, methods that train models to reduce disparities through the use of protected characteristics, such as adversarial debiasing, are permissible.¹⁷⁵ A differing view is that any explicit consideration of protected characteristics during model development is a violation of fair lending and that models should be trained in a way that does not consider protected characteristics; then, separately, the disparities of each model are calculated (likely by a separate team with access to protected characteristics) to select the least discriminatory model.¹⁷⁶ Both views would have an impact on the precise way in which our framework would be implemented

¹⁷² By requiring lenders to submit detailed data during these examinations, the CFPB can assess and address discriminatory practices, pushing for broader demographic data collection in these markets.

¹⁷³ For example, the National Community Reinvestment Coalition have pushed for legislation and policies that require lenders to collect and report demographic data across all types of loans. See Jason Richardson, *The Critical Need to Address Missing Data in HMDA*, NCRC (Aug. 25, 2022), <https://www.ncrc.org/the-critical-need-to-address-missing-data-in-hmda>. See also Cara V. James et al., *Using Race and Ethnicity Data to Advance Health Equity: Examples, Promising Practices, Remaining Challenges, and Next Steps*, URBAN INST. HEALTH POL'Y CTR. (Oct. 2023), <https://www.urban.org/sites/default/files/2023-10/Using%20Race%20and%20Ethnicity%20Data%20to%20Advance%20Health%20Equity.pdf>.

¹⁷⁴ See, e.g., the Joint Statement: "Generally, a lender that applies different lending standards or offers different levels of assistance on a prohibited basis, regardless of its motivation, would be violating both the FH Act and the ECOA." The Statement qualifies this by saying- "In addition, providing different treatment to applicants to address past discrimination would be permissible if done in response to a court order or otherwise in accord with applicable legal precedent."

¹⁷⁵ This is what Ric Pace calls "fairness through unawareness." Pace, *supra* note 50. As pointed out by Pace, the recent settlement with Meta seems to endorse a system that explicitly uses demographic data to reduce disparities in ad audiences. See U.S. DEP'T OF JUSTICE, *Justice Department and Meta Platforms Inc. Reach Key Agreement as They Implement Groundbreaking Resolution to Address Discriminatory Delivery of Housing Advertisements* (Jan. 9, 2023), <https://www.justice.gov/opa/pr/justice-department-and-meta-platforms-inc-reach-key-agreement-they-implement-groundbreaking>.

¹⁷⁶ Ric Pace argues that even this second type of model selection could trigger disparate treatment, as the model selected is based on an explicit use of demographics. Pace, *supra* note 50. This concern is consistent with a read of the Ricci decision as any process that is selected based on the demographic outcomes could trigger disparate treatment.

so that future regulatory guidance can clarify the exact extent of permissible use of protected characteristics for the purpose of reducing disparities.

C. Implications for Multiple Protected Groups

Fair lending enforcement typically focuses on assessing a lender’s model based on its impact on a single protected group, such as race.¹⁷⁷ However, when deciding whether to adopt a new model—such as when another model is a less discriminatory alternative—the implications extend beyond one protected group and can affect additional categories like sex and marital status. Sometimes, discrimination *overlaps* protected groups, such as a policy to not extend credit to women without a husband’s co-signature. Such a practice yields discriminatory outcomes against several groups: those without husbands (marital status) and all women in general if men are not subject to the same requirement (sex). Such a policy seems primarily *sex* discrimination, because “women” is a larger group than “married women,” the latter being a subset of the former.

Sometimes, though, there are conflicts: discriminatory practices can also *antagonize* protected groups, in the sense that one protected class may receive improved equal outcomes, while another might experience greater outcome inequality (i.e., equality of outcome might involve tradeoffs across protected groups). Under traditional fair lending practice, such trade-offs among groups was not particularly salient. Although, legally, any unequal outcomes could trigger a disparate impact claim, regardless of whether the disparity arose from ‘well-meaning’ intentions to improve outcomes for another protected group, in practice, fair lending enforcement has typically targeted one protected group at a time without holistically considering the impacts on all protected groups.

Discrimination that antagonizes protected groups becomes more salient under our framework, as outcome inequality and differential validity are quantified and compared explicitly across models. A lender considering a new model that enhances overall accuracy and improves outcomes for one group (e.g., race) but potentially worsens them for another group (e.g., sex) faces a complex decision. This scenario highlights a significant gap in current fair lending frameworks, which lack clear guidance on balancing such trade-offs and instead concentrate on addressing disparities for one group. Existing regulations and fairness metrics are typically designed to minimize disparities for a specific protected group, but they do not sufficiently address situations where improvements for one group might lead to detriments for another.

Such concerns are stylized in cases of “fairness gerrymandering” where outcomes are equal for each legally recognized protected class but not for combinations of such classes.¹⁷⁸ To take an oversimplified example, suppose a population has two binary classes: one of race (e.g., Black and white) and one of gender (e.g., men and women), where both features are distributed independently and equally, and consider a model that yields positive outcomes only for Black men or white women.¹⁷⁹ This model provides equal outcomes across the protected groups of race (positive 50% of the time

¹⁷⁷ See, e.g., *U.S. v. First Nat’l Bank of Pa.*, No. 1:24-cv-00088-CCE-LPA (M.D.N.C. 2024), https://www.justice.gov/d9/2024-02/settle_first_national_bank_of_pa-final.pdf; *U.S. v. Patriot Bank.*, No. 2:24-cv-02029-TLP-tmp (W.D. Tenn. 2024), https://www.justice.gov/d9/2024-01/1_settle_patriot-final.pdf.

¹⁷⁸ See Michael Kearns et al., *Preventing Fairness Gerrymandering: Auditing and Learning for Subgroup Fairness*, PROC. OF THE 35TH INT’L CONF. ON MACHINE LEARNING (2018).

¹⁷⁹ *Id.*

for both Black and white people) and sex (positive 50% of the time for both men and women) but it obviously provides unequal outcomes for Black women and white men (positive 0% of the time for either).

Our proposed framework emphasizes considering multiple disparity goals but can also provide a conceptual approach for evaluating disparities across multiple protected groups. To address such gerrymandering, “intersectional” measures of fairness, as proffered by Foulds et al. (2019), examine equality of outcome across all protected classes *and all combinations of such classes*.¹⁸⁰ Once we recognize that fair lending involves competing goals, it is also natural that trade-offs between different protected groups may arise in certain circumstances.

To the extent that equal-outcomes and validity cannot be achieved simultaneously, then even intersectional measures of equal outcomes necessarily trade away validity. Given our argument that validity can be just as important as outcome equality, these trade-offs should be explicitly acknowledged, similar to how we approach disparity metrics. Once the trade-off is acknowledged, it can be optimized. A lender’s LDA search could be guided by a weighted measure of disparities across different protected groups with the weights for each group determined by regulatory guidelines.¹⁸¹

D. Implications for Disparity-Performance Tradeoff

Disparate impact has been traditionally interpreted as paying little attention to the magnitude of disparities, if the policy has a business justification that cannot be achieved in a less discriminatory

¹⁸⁰ James R. Foulds et al., *An Intersectional Definition of Fairness*, ARXIV:1807.08362 (2019), <https://arxiv.org/abs/1807.08362> (“For example, [under an intersectional definition of fairness] the probability of being given a loan would be similar regardless of a protected group’s intersecting combination of gender, race, and nationality . . .”).

¹⁸¹ Focusing on the ex-ante approach to LDA alleviates the concern that has been discussed previously in the context of employment discrimination litigation of successive suits. *See Civil Rights Act of 1991 and Less Discriminatory Alternatives in Disparate Impact Litigation Impact*, 106 HARV. L. REV. 1621, 1627 (1993). What if, after a court deems a policy to be ‘the’ LDA at the final stage of disparate impact litigation, technological advancements allow for a new policy that performs even better with respect to equal outcomes? Presumably, the plaintiff is precluded from relitigating to argue the defendant should now adopt the ‘new’ LDA. *See Civil Rights Act of 1991* § 108, 42 U.S.C.A. § 2000e-2(n) (precluding successive suits by members of the same protected group under Title VI).

The problem of successive suits is magnified in the case of discrimination that antagonizes protected groups—one plaintiff’s LDA could be another’s disparity. *See Civil Rights Act of 1991 and Less Discriminatory Alternatives in Disparate Impact Litigation Impact*, 106 HARV. L. REV. 1621, 1627 (1993) (“LDAs may themselves have disparate impacts on different identifiable groups.”). If a policy deemed to be an LDA with respect to one protected group (Group 1) is later discovered to have an adverse impact on equal outcomes for another protected group (Group 2), then procedural issues arise. For example, a member of Group 2 may not be able to challenge the policy if they were on notice of Group 1’s suit and given an opportunity to object to LDA proposals. § 108, 42 U.S.C.A. § 2000e-2(n) (precluding claims under Title VI “by a person who, prior to the entry of the judgment or order had . . . actual notice of the proposed judgment or order sufficient to apprise such person that such judgment or order might adversely affect the interests and legal rights of such person and that an opportunity was available to present objections to such judgment or order by a future date certain. . .”). An alternative strategy for Group 2 to litigate could be to argue that the later discovery of adverse impact on Group 2 is “an intervening change in law or fact,” in which case Group 2’s discrimination claim might not be precluded. § 108, 42 U.S.C.A. § 2000e-2(n) (precluding claims under Title VI “by a person whose interests were adequately represented by another person who had previously challenged the judgment or order on the same legal grounds and with a similar factual situation, *unless there has been an intervening change in law or fact*”) (emphasis added).

The ex-ante approach avoids these thorny procedural issues. Rather than encumbering the LDA search with questions about when plaintiffs can enter or re-enter court, proactively searching for LDA’s under an ongoing compliance regime keeps LDA analysis fresh and dynamic, sensitive to ongoing technological advancements.

way. Although the magnitude of disparities may play a limited role as a threshold for satisfying the first stage of disparate impact,¹⁸² the degree of disparity arising from a pricing policy still does not influence whether there is a business justification for the pricing policy and has not explicitly played a role in evaluating alternatives. For instance, consider a lender transitioning to a newer pricing algorithm, leading to higher interest rates for women. If improved accuracy, even slightly, is achieved by the new algorithm, it can validate any arising disparities, which is a troubling proposition.¹⁸³

This binary interpretation of disparate impact undermines the goals of fair lending, as it does not consider the performance gains relative to the disparity harm. Our proposed framework can also provide a conceptual approach for balancing model performance of a new model with the disparities it creates. This would mean, for example, that when a lender considers the inclusion of alternative data for assessing creditworthiness, such as utility payments and education information, the relevant assessment is not only whether this information increases the overall performance of the model in a way that is not achieved by a less discriminatory model but also whether the performance gains are substantial enough to offset and disparities created.

Although traditionally fair lending has not typically weighed improved performance against disparity harm, a closer reading of existing regulation suggests balancing of the benefits and harms could be considered. In particular, the less discriminatory alternative may provide an alternative model that does not strictly have the same level of performance of existing models, such that there may be some acceptable reduction in performance for a given change in disparities.¹⁸⁴ The possibility that a worse performing model can still be a LDA is suggested by the 2023 HUD Disparate Impact Rule, reinstating the 2013 Disparate Rule, which explicitly rejects the interim 2020 Rule's requirement that a LDA be "equally effective."¹⁸⁵ The preamble to the 2023 Disparate Impact Rule explains that HUD

¹⁸² The first step of the burden shifting framework does not make reference to required magnitude of disparities. *See* HUD 2023 Disparate Impact Rule, at 19,486 ("HUD further declines to set statistical standards, including statistical thresholds"); *see also* Kevin Tobia, *Disparate Statistics*, 126 YALE L.J. 2382 (2017). In employment disparate impact claims typically follow a showing of a magnitude of disparities of at least 4/5 to survive the first step of the burden shifting framework. *See* Uniform Guidelines on Employment Selection Procedures, 29 C.F.R. § 1607.15 (1978). That said, this ratio can be regarded as more of a rule of thumb than a definitive test for identifying disparate impact. *See* Manish Raghavan & Pauline T. Kim, *Limitations of the 'Four-Fifths Rule' and Statistical Parity Tests for Measuring Fairness*, 8 GEO. L. TECH. REV. (forthcoming 2023). Moreover, the third step, which allows the plaintiff to show that there is a less discriminatory alternative, also limits the ability of lenders to adopt policies that create unnecessarily large disparities.

¹⁸³ *See generally* Fuster et al., *supra* note 20 (explaining how the statistical technology itself could change the disparities created by a pricing rule).

¹⁸⁴ In the third Upstart monitorship report, the monitor proposed a methodology, to assess whether the reduction in predictive accuracy of an LDA model would be acceptable to the lender in exchange for an improved AIR, concluding that this trade-off was acceptable. RELMAN COLFAX PLLC, FAIR LENDING MONITORSHIP OF UPSTART NETWORK'S LENDING MODEL, THIRD REPORT OF THE INDEPENDENT MONITOR at 29-30 (2022), <https://www.relmanlaw.com/assets/htmldocuments/PUBLIC%20Upstart%20Monitorship%203rd%20Report%20FINAL.pdf>. However, the Fourth and Final Report of the monitorship noted that Upstart disagreed with this methodology and chose not to adopt it. RELMAN COLFAX PLLC, FAIR LENDING MONITORSHIP OF UPSTART NETWORK'S LENDING MODEL, FOURTH REPORT OF THE INDEPENDENT MONITOR at 4 (2024), <https://www.relmanlaw.com/assets/htmldocuments/PUBLIC%20Upstart%20Monitorship%203rd%20Report%20FINAL.pdf>.

¹⁸⁵ 24 C.F.R. § 100.500(c)(3) ("If a defendant rebuts a plaintiff's assertion under paragraph (c)(1) of this section, the plaintiff must prove by the preponderance of the evidence either that the interest (or interests) advanced by the defendant are not valid or that a less discriminatory policy or practice exists that would serve the defendant's identified interest (or interests) in an equally effective manner without imposing materially greater costs on, or creating other material burdens for, the defendant.").

does not hold that “a proposed less discriminatory alternative fail simply because there will be some amount of increased cost associated with the alternative policy.”¹⁸⁶ The HUD Rule does not explain what increase in costs (in the form of performance loss) may be tolerated for a decrease in disparities, but our proposed framework could provide a method for trading off performance and disparities.¹⁸⁷

CONCLUSION

At the core of this Article is the recognition that credit market disparities, which fair lending aims to address, are complex and involve multiple, often conflicting dimensions. These tensions arise due to 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 reflect 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 inherently creates a tension between 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 two critical ways. First, it should expand the dimensions of disparity to be measured and addressed by lenders. The traditionally narrow focus 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. Second, fair lending should acknowledge that equal outcomes and equal validity are goals that cannot be fully achieved simultaneously. Thus, regulatory oversight and compliance must adopt a framework for measuring and balancing these competing goals, as we propose.

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, creating tensions between various notions of disparity. More broadly, our analysis charts a path for confronting conflicting goals in discrimination doctrines, such as when policies impact multiple protected groups who are differentially impacted by policies, or when balancing the benefits of improved model performance against disparities.

By explicitly recognizing these competing goals and challenges, and creating frameworks to evaluate and oversee models, discrimination law can be better equipped to address contemporary challenges. Modernizing fair lending in the ways we propose will prepare it to extend credit more equitably and validly, ultimately fostering fairer access to credit.

¹⁸⁶ Reinstatement of HUD’s Discriminatory Effects Standard, 24 C.F.R. § 100.

¹⁸⁷ See John K. Lucey, *The Redlining Battle Continues: Discriminatory Effect v. Business Necessity under the Fair Housing Act*, 8 ENV’T L. AFF. 357, 390 (1979) (“[T]he viability of less discriminatory alternatives will be a function of both the increase in costs borne by the defendant and the amount by which the disproportionate impact is reduced.”).

MODERNIZING FAIR LENDING - APPENDIX

Spencer Caro, Talia Gillis & Scott Nelson

APPENDIX A

This Appendix provides further details on the simulation exercise in II.C.1 of the main body of this Article.

We simulate 500,000 individuals who belong to one of two protected groups. One of the groups is a *statistical* minority, in the sense that it comprises only 10% of the population (50,000 individuals). The other group is a statistical majority and comprises the remaining 90% of the population. These statistical minority and majority groups can be, but do not need to be, also thought of as racial and ethnic minorities and non-minorities, or other protected classes that in a particular setting may comprise unequal shares of the overall population of applicants.¹

Default outcomes are generated from individual-level Bernoulli default probabilities. These default probabilities are computed from an individual-level log odds of default that is the weighted sum of two “factors,” each of which is normally distributed by group. In other words, there is one normal distribution of the first factor for minority individuals, another normal distribution of the second factor for non-minority individuals, and so on.

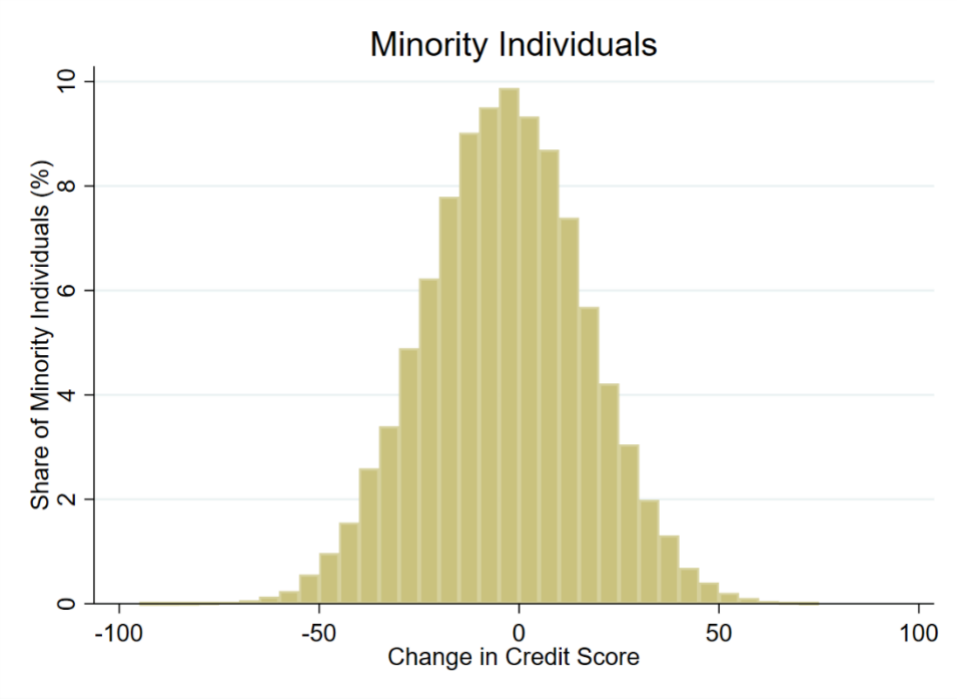
These factors are observed with noise in the lender’s data. Specifically, each datapoint has standard normal noise centered around the unobserved value of a given factor. In the simulation, we vary the characteristics of these factors by changing the extent to which they are informative about default outcomes and the extent to which they are correlated with group membership. We simulate 81 different versions of this environment. In each version, the lender has a baseline credit scoring model based only on data that partially reveals the first factor, and the lender is considering a novel credit scoring model that uses data that partially reveals the second factor. In other words, in each of the 81 simulations, the lender gains access to a signal of the second factor.

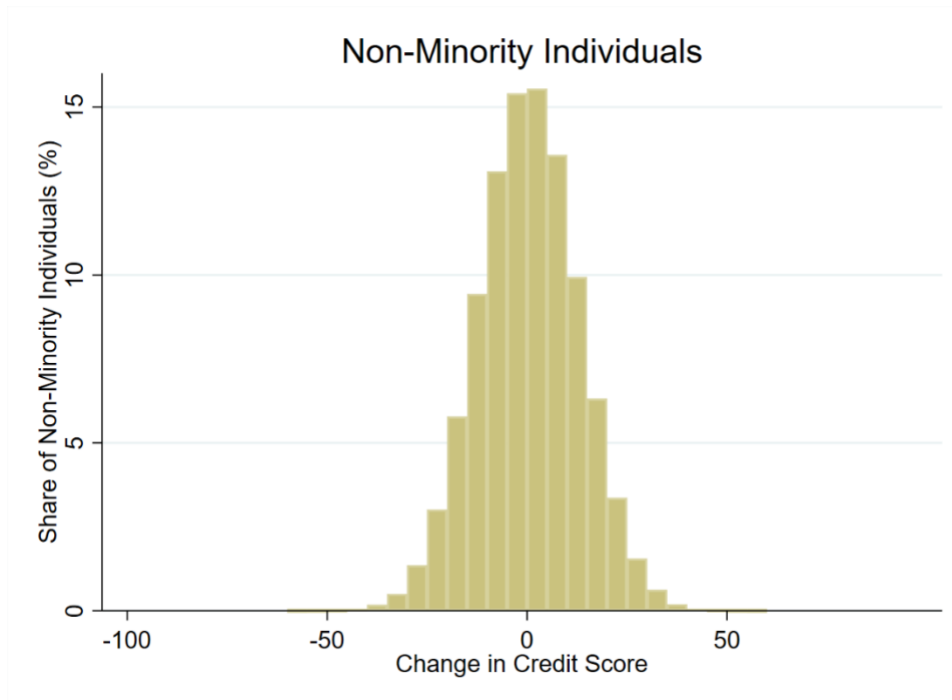
For simplicity, credit scoring models in each simulation are based on OLS regressions of default on the observed signal(s), plus a constant. These regressions are estimated on a training sample that comprises half of the simulated individuals, and then are deployed on a holdout sample that comprises the other half. Approval rates are defined by assuming an approval threshold of a credit score of 660, similar to some real-world lending policies, after applying a realistic mapping from OLS-predicted default rates to credit scores.

An example of the distribution of changes in credit score for minority and non-minority individuals in one of these 81 versions of the simulation is shown in Appendix Figure 1. As the figure

¹ This feature of the simulation is not crucial for our results but incorporates the real-world feature that credit scoring models may have less validity for some protected groups when those groups, by virtue of being a statistical minority, receive less weight in a model training process that optimizes model performance in the population overall.

emphasizes, gaining access to the additional underwriting variable in this version of the simulation provides substantially more additional information about minority individuals, in the sense of inducing larger-magnitude changes in the implied credit score for these individuals. This additional information facilitates a reduction in differential validity, by better revealing which individuals are least likely to default. However, this additional information does not uniformly benefit all individuals in the minority group (indeed, if information were the same for all individuals it would hardly be information at all). Some individuals are predicted to have higher default risk when this new underwriting variable is made available, and on average, approval rates fall in the minority group given this new information.





Appendix Figure 1: Simulated Change in Credit Score by Group

Concretely, in the example shown in Appendix Figure 1, simulated minority and non-minority individuals have approval rates of 49.4% and 55.2% respectively before the introduction of the new underwriting variable, and have and false negative rates of 50.5% and 43.8% respectively. The new model decreases minority approval rates by 6.1 percentage points, while decreasing non-minority approval rates by only 1.8 percentage points – exacerbating outcome inequality. However, the new model also decreases the minority false negative rate by 37.8 percentage points, and the non-minority false negative rate by only 2.7 percentage points – implying substantial gains in validity for the group that previously faced a disparity in this dimension. Table 1 summarizes the disparities in this example under both the baseline policy (“Model 1”) and new underwriting variable (“Model 2”).

Model	Non-Minority Group	Minority Group	Difference
Approval Rates			
Model 1	55.2%	49.4%	5.8%
Model 2	53.4%	43.3%	10.1%
False Negative Rates			
Model 1	43.8%	50.5%	6.7%
Model 2	41.1%	12.7%	28.4%

Appendix Table 1: Approval Rates and False Negative Rates Differences

In general, the trade-off between differential validity and unequal accuracy emerges when the new underwriting variable is strongly correlated with protected group membership, but also highly informative about default risk *within* group. The example shown in Appendix Figure 1 is one such case.

The remainder of this appendix provides the Stata 17.0 code used to run the simulation:

```
clear all
set seed 1896

save MFLmodelruns, emptyok replace

local o = 500000
local muW = 3.7
local muB = -2.8

set obs `o'
gen holdout = _n / `o' < .5
gen black = round(runiform())-.4
gen white = 1-black

compress
tempfile loopstart
save `loopstart', replace

local counter = 0
foreach m2white in 0 1 2 {
  foreach m2black in 0 2 4 {
    foreach s2white in 1 3 5 {
      foreach s2black in 3 5 7 {

local counter = `counter'+1
use `loopstart', clear

gen f1 = .
replace f1 = rnormal(-3.7, 1) if white
replace f1 = rnormal(-3.5, .2) if black

gen f2 = .
replace f2 = rnormal(`m2white', `s2white') if white
replace f2 = rnormal(`m2black', `s2black') if black

gen odds = .
replace odds = exp((f1 + .01*f2) / 1.01) if white
replace odds = exp((f1 + 10*f2)/11) if black
gen delta = odds / (1+odds)
gen default = runiform()<delta

gen s1 = .
replace s1 = rnormal(f1, 1/sqrt(1.0)) if white
replace s1 = rnormal(f1, 1/sqrt(1.0)) if black

gen s2 = .
```

```

replace s2 = rnormal(f2, 1/sqrt(1.0)) if white
replace s2 = rnormal(f2, 1/sqrt(1.0)) if black

reg default s1 if holdout==0
predict fico1 if holdout==1
replace fico1 = 680 - 200*fico1

reg default s1 s2 if holdout==0
predict fico2 if holdout==1
replace fico2 = 680 - 200*fico2

keep if holdout==1

gen approved1 = fico1 > 660
gen approved2 = fico2 > 660
gen rejected1 = 1-approved1
gen rejected2 = 1-approved2

sum rejected1 if black==0 & default==0, meanonly
local mean1 = `r(mean)´
sum rejected2 if black==0 & default==0, meanonly
local mean2 = `r(mean)´
local d_FNR_white = `mean2´ - `mean1´
local FNR_white = `mean1´

sum rejected1 if black==1 & default==0, meanonly
local mean1 = `r(mean)´
sum rejected2 if black==1 & default==0, meanonly
local mean2 = `r(mean)´
local d_FNR_black = `mean2´ - `mean1´
local FNR_black = `mean1´

sum approved1 if black==0, meanonly
local mean1 = `r(mean)´
sum approved2 if black==0, meanonly
local mean2 = `r(mean)´
local d_AR_white = `mean2´ - `mean1´
local AR_white = `mean1´

sum approved1 if black==1, meanonly
local mean1 = `r(mean)´
sum approved2 if black==1, meanonly
local mean2 = `r(mean)´
local d_AR_black = `mean2´ - `mean1´
local AR_black = `mean1´

clear
set obs 1

gen m2white = `m2white´
gen m2black = `m2black´
gen s2white = `s2white´
gen s2black = `s2black´
gen AR_white = `AR_white´
gen FNR_white = `FNR_white´

```

```

gen AR_black = `AR_black'
gen FNR_black = `FNR_black'
gen d_AR_white = `d_AR_white'
gen d_FNR_white = `d_FNR_white'
gen d_AR_black = `d_AR_black'
gen d_FNR_black = `d_FNR_black'

append using MFLmodelruns
save MFLmodelruns, replace

}
}
}
}

gen d_gap_validity = (-d_FNR_black) - (-d_FNR_white)
gen d_gap_parity = d_AR_black - d_AR_white

cap drop area*
gen area_x = .
gen area_y = .
replace area_x = 0 in 1
replace area_x = .53 in 2
replace area_y = 0 in 1
replace area_y = -.49*.53 in 2
replace area_x = 0 in 3
replace area_x = .53 in 4
replace area_y = .1 in 3
replace area_y = .1 in 4

twoway area area_y area_x in 1/2, color(gs14) || ///
      area area_y area_x in 5/6, color(gs14) || ///
      scatter d_gap_parity d_gap_validity, mcolor(navy) || ///
      function y = 0*x, lpattern(shortdash_dot) lcolor(maroon) range(0 .53) || ///
      function y = 0-.49*x, lpattern(dash) lcolor(forest_green) range(.002 .53) ///
      xtitle("Change in Validity Disparity") ///
      ytitle("Change in Outcome Disparity") ///
      graphregion(fcolor(white) color(white) icolor(white)) plotregion() ///
      legend(label(5 "Balanced Regulation") label(4 "Status-Quo Regulation") ///
      label(2 "Status-Quo Compliant") label(1 "Balanced Compliant") ///
      order (4 5) row(1))
      graph export scatter.png, replace

```

APPENDIX B

This Appendix develops more formally our proposal for a fair-lending regulatory framework that balances the competing goals of equal outcomes and equal model validity across protected groups.

To write our proposal formally, consider a lender that chooses a model f from a set of possible models F , and data inputs x from a set X of possible groups of inputs, in order to maximize a private benefit Π . The proposed regulatory framework can then be written as a constraint on the lender’s optimization problem,

$$\max_{f \in F, x \in X} \Pi(f(x))$$

$$\text{s.t. } x \in \tilde{X} \text{ and } \alpha \text{OutcomeInequality}(f(x)) + \beta \text{DifferentialValidity}(f(x)) \leq A$$

In this mathematical constraint, *OutcomeInequality* refers to inequalities in average outcomes across protected classes, and *DifferentialValidity* refers to inequalities in the rate of true positives across protected classes. Intuitively, a regulator may tolerate some quantum of disparity in outcomes in order to reduce differential validity, while the terms α and β quantify this tolerance. The endowment term A can be chosen in view of what outcomes are currently feasible for a particular protected class, line of business, and lender, or equivalently to “grandfather-in” the status quo for lenders seen as in compliance with current fair-lending regulation.²

As described in the main body of this Article, we particularize our more general definition of differential validity to focus on the approval rate *among* the applicants who would later repay a loan if approved, known as the “true positive rate”.³ To write this particularized definition formally, note that lending to a loan repayer can be termed a *true positive* (“TP”). In contrast, making a loan to an applicant who will ultimately default on the loan can be termed a *false positive* (“FP”). Likewise, denying a loan to someone who would have defaulted is a *true negative* (“TN”), and denying a loan to someone who would have repaid it is a *false negative* (“FN”). In situations where credit risk is described by later repayment, and where the underwriting decision can be viewed as a binary decision, these four categories exhaustively describe the outcomes and potential outcomes of a lending decision. Using these definitions, we then can write differential validity formally as,

² C.f. Muhammad Zafar et al., *Fairness Beyond Disparate Treatment & Disparate Impact: Learning Classification without Disparate Mistreatment* (2017), <https://arxiv.org/abs/1610.08452> (introducing a notion of “disparate mistreatment” that grounds another optimization problem subject to a differential accuracy constraint). In contrast to Zafar et al., our approach formalizes a “budget” that allows efficient trade-offs between parity, accuracy, and performance goals, all of which can differ from the lender’s own objective. The endowment term A in this budget allows for the fact that, in some markets or contexts where the LDA constraint is deployed, predicting outcome class membership may be either sufficiently difficult, or sufficiently different for different protected classes (e.g., due to heteroskedastic risk that varies with class membership), such that the constraint can only be slack when A takes a positive nonzero value. The “grandfathering-in” approach described in the text guarantees that A takes an appropriate value to enable the appropriate slackness in the constraint.

³ This notion of accuracy is related to, but different from, the concepts of balance for the positive class and balance for the negative class, as in Kleinberg et al, *supra* note ____.

$$\text{DifferentialValidity}(f(x)) = \left| \frac{TP_g}{TP_g + FN_g} - \frac{TP_{g'}}{TP_{g'} + FN_{g'}} \right|$$

where g and g' denote two protected groups. Similarly, given our focus on (unconditional) approval rates in defining outcome disparities, we can define outcome inequality formally as,

$$\text{OutcomeInequality}(f(x)) = \left| \frac{TP_g + FP_g}{N_g} - \frac{TP_{g'} + FP_{g'}}{N_{g'}} \right|$$

The denominators N_g and $N_{g'}$ denote the total number of applicants in each protected group, while the numerators are counts of true positives and false positives in each class. Intuitively, this measure is simply the absolute difference in approval rates across groups. Equality of outcomes is violated if and only if this difference is nonzero, while greater differences in approval rates imply a greater value for the *OutcomeInequality*($f(x)$) term in our constraint.

Intuitively, this formal regulatory constraint is equivalent to a “scorecard” for lenders, similar to the scoring approach used by regulators or other evaluators in various contexts such as restaurants’ health code compliance, hospital safety, and college rankings. To be fair-lending compliant, a lender needs to achieve a minimum score, related to the term \mathcal{A} in the formal constraint. The lender loses “points” on this score for each quantum of outcome inequality or differential validity, and the regulator’s calibration of the terms α and β determine how many points these quanta imply for the lender.

In some contexts, a regulator may also be interested in allowing the term \mathcal{A} to vary based on how well a model performs overall. Such an approach would capture some of the value ascribed to business necessity under status-quo fair lending regulation: a model that performs better may better satisfy a lender’s business necessity, and thus an increase in one or more disparity measures may be seen as permissible when the increase is accompanied by greater model performance. To implement such an approach, which we describe in Section III.D. of the main Article, a regulator would update the term \mathcal{A} in the constraint above to instead take the form,

$$\mathcal{A} + \gamma \text{Performance}(f(x))$$

where we define,

$$\text{Performance}(f(x)) = \left[\left| \frac{TP_g + TP_{g'}}{TP_g + FN_g + TP_{g'} + FN_{g'}} \right| + \left| \frac{TN_g + TN_{g'}}{TN_g + FP_g + TN_{g'} + FP_{g'}} \right| \right]$$

This expression is similar to that for differential validity, but rather than emphasizing a difference between protected classes in approval rates for borrowers who would repay, it sums both these approval rates and also the rate at which future *defaulters* are rejected, across both protected classes. These two terms could, if desired, also be weighted to reflect different social preferences for true positives versus true negatives, based on whether it is more costly to erroneously deny a loan or erroneously grant a loan.

By way of analogy, the regulator’s chosen balance between false positives and false negatives can be seen as a balance between *Type I* and *Type II* errors at the level of each loan application: each applicant presents a risk of either falsely rejecting an applicant who should be approved, or falsely

approving an applicant who should be rejected. Each type of error has costs. And while the appropriate balance between these two error types has been prominent in other regulatory frameworks, for example FDA trials of drug efficacy,⁴ this balance has been largely absent from DI analyses.

One last aspect of our proposed regulatory constraint deserves clarification. As can be seen above, in addition to the constraint on disparities we also impose the constraint $x \in \tilde{X}$. This auxiliary constraint can be thought of as capturing any input variables that are deemed socially objectionable to use in underwriting. For example, a regulator may wish to exclude protected class membership itself from the set of permissible inputs, in which case the set \tilde{X} could be specified to exclude these. In this way, our proposed constraint could also achieve some of the goals of disparate *treatment* regulation.⁵

In fact, the constraint may prevent other forms of disparate treatment that arise through use of proxy variables, in addition to blocking explicit disparate treatment through the design of the set \tilde{X} . To see this, consider a proxy variable that a lender uses with the goal of discriminating against a protected class, such that this proxy does not enhance model performance. Any use of this proxy would increase the unequal outcomes term on the left-hand-side of the constraint without increasing the Performance term that potentially appears on the right-hand-side, leading to a violation of the constraint. Because the focus of our work is on disparate impact, we leave a fuller treatment of these disparate treatment concerns to future work.

⁴ See Michael Intriligator, *Drug Evaluations: Type I vs. Type II Errors* (1996) (unpublished manuscript), <https://escholarship.org/uc/item/5fg9n284>.

⁵ See *Price Waterhouse v. Hopkins*, 490 U.S. 228, 243 (1989) (quoting congressional records that said Title VII “expressly protects the employer’s right to insist that any prospective applicant, Negro or white, must meet the applicable job qualifications. Indeed, the very purpose of Title VII is to promote hiring on the basis of job qualifications, rather than on the basis of race or color.”). Cf. *Int’l Bhd. of Teamsters v. United States*, 431 U.S. 324, 335 (1977) (“‘Disparate treatment’ ... is the most easily understood type of discrimination. The employer simply treats some people less favorably than others because of their race, color, religion, sex, or national origin. Undoubtedly disparate treatment was the most obvious evil Congress had in mind when it enacted Title VII. ... Claims of disparate treatment may be distinguished from claims that stress ‘disparate impact.’ The latter involve employment practices that are facially neutral in their treatment of different groups but that in fact fall more harshly on one group than another and cannot be justified by business necessity.”).

More generally, our proposed constraint can accommodate any input restrictions, such as the “Input Accountability Test” proposed in Robert P. Bartlett et al., *Algorithmic Discrimination and Input Accountability under the Civil Rights Acts*, 36 BERKELEY TECH. L. J. 675-736 (2021).