Vanderbilt Journal of Entertainment & Technology Law

Volume 23 | Issue 3 Article 4

2021

Current Regulatory Challenges in Consumer Credit Scoring Using Alternative Data-Driven Methodologies

Sahiba Chopra

Current Regulatory Challenges in Consumer Credit Scoring Using Alternative Data-Driven Methodologies

ABSTRACT

Credit is a crucial determinant of financial success for most US consumers, but not all consumers can access it. This financial exclusion is partially due to traditional credit-risk scoring and approval processes that cannot assess the creditworthiness of "credit invisible" or "thin file" consumers—that is, consumers who do not have enough traditional data depicting their financial payment history. Consequently, some consumer-reporting agencies and lenders turn to alternative data credit-scoring systems as a way to increase financial inclusion. The enormous complexity of these alternative consumer credit-scoring systems, however, raises significant accuracy and transparency issues—most of which stem from their secret, legally protected status—as well as heightened concerns over the use of discriminatory and biased scoring practices using nontraditional behavioral data. If these issues are not addressed, alternative data-driven credit-scoring systems can potentially amplify transparency and discrimination issues, preventing consumers from understanding the factors that impact their credit scores. At the same time, they can position underprivileged groups to face increased discrimination in terms of both accessing credit and receiving favorable interest rates.

This Note proposes four regulatory solutions and suggests enhancements to the Model Fairness and Transparency in Credit Scoring Act developed by legal and technology scholars Hurley and Adebayo. The current regulatory framework can better address discrimination by requiring lenders to disclose how they define "creditworthiness" so that consumers can gain a better understanding of the standards to which they are being held. It can also push lenders to foster more appropriate credit standards. Moreover, federal legislation is needed to curtail or prohibit the use of nontraditional behavioral data, especially data derived from a consumer's social networks, which can unfairly penalize consumers for their social or cultural associations. If

this type of legislation is not likely to pass at a federal level, then regulatory agencies should regulate these firms under the presumption that behavioral data is inherently discriminatory until proven otherwise. Finally, regulators should seek to incentivize firms using alternative credit scoring methodologies to seek no-action letters.

TABLE OF CONTENTS

Imagine you are eighteen, on the cusp of many exciting life changes—going off to college, buying your first car so you can actually get there, applying for a summer job to pay for gas, and maybe even putting in an application for a nice first apartment. Now imagine another scenario—you are brand new to this country, eager to start chasing your version of the American dream. You have never held a job or owned anything in this country, nor have you paid any bills or opened any bank accounts. In both of these scenarios, you have big dreams that are nearly impossible to achieve unless you can access credit.

Access to credit can be a key determinant of financial success for a majority of Americans and is typically determined by an individual's consumer credit rating.[1] A consumer's credit rating can impact her in many important ways, such as gaining access to employment opportunities, obtaining higher education, and purchasing assets, such as a home and a car—assets that are traditionally considered crucial to building individual financial wealth.[2] In some instances, such as when families struggle financially, credit is necessary for survival and is used to pay for nondiscretionary, essential goods, such as food and housing.[3] Unfortunately, there are approximately forty-five million people, primarily from Black and Hispanic backgrounds, who are considered "unscorable" because credit-scoring firms are unable to provide an assessment of their credit risk using traditional scoring tools.[4]

Credit risk, which is summarized in a credit score, is simply defined as the "potential that a borrower or counterparty will fail to perform on an obligation."[5] However, with recent advances in machine learning and the proliferation of credit firms that utilize new types of data ("alternative data") and methodologies, the way credit risk, particularly consumer credit risk, is assessed will continue to change significantly.[6]

This Note examines how consumer credit risk scoring works, how it is changing, and potential solutions to fill the gaps in the existing regulatory framework. Part I discusses the differences between using traditional and alternative data in credit risk scoring. Next, Part II introduces the existing regulatory scheme for credit reporting and identifies areas where it falls short in protecting consumers. Part III proposes potential solutions to improve the consumer credit scoring process to address the risk of discrimination posed by using alternative data and concludes that lenders using

---

1. *See* Mikella Hurley & Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE J.L. & TECH. 148, 153–54 (2016).

2. *See id.* at 202.

3. *Examining the Use of Alternative Data in Underwriting and Credit Scoring to Expand Access to Credit: Hearing Before the Task Force on Fin. Tech. of the H. Comm. on Fin. Servs.*, 116th Cong. 7–8 (2019) [hereinafter *Hearing*] (statement of Kristin N. Johnson, McGlinchey Stafford Professor of Law and Associate Dean of Faculty Research, Tulane University Law School), https://www.congress.gov/116/meeting/house/109867/witnesses/HHRG-116-BA00-Wstate-JohnsonK-20190725.pdf [https://perma.cc/T7FY-S8YP].

4. *See* Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process, 82 Fed. Reg. 11183, 11184 (2017) [hereinafter Request for Information]; BD. OF GOVERNORS OF THE FED. RSRV. SYS., REPORT ON THE ECONOMIC WELL-BEING OF U.S. HOUSEHOLDS IN 2018, at 26 (2019).

5. *Supervisory Policy and Guidance Topics*, BD. OF GOVERNORS OF THE FED. RSRV. SYS., https://www.federalreserve.gov/supervisionreg/topics/credit_risk.htm [https://perma.cc/2LVY-Z5VE] (Dec. 22, 2020).

6. *See generally* Hurley & Adebayo, *supra* note 1.

alternative credit scoring methodologies should be held to more stringent regulatory standards. Finally, Part IV provides a brief future outlook on the use of alternative data in consumer credit scoring.

I. TRADITIONAL VERSUS ALTERNATIVE DATA IN CREDIT RISK SCORING

Access to household debt is typically determined by the automated scoring criteria adopted by a specific lender.[7] Lenders can evaluate the creditworthiness of a consumer based on scores derived from "traditional data," "alternative data," or a combination of both.[8] Traditional data, according to the Consumer Financial Protection Bureau (CFPB), includes information relating to loans or credit limits, repayment of debt, inquiries into credit history, and other relevant information from publicly available records.[9]

In contrast, alternative data consists of all data that falls outside of the scope of traditional data, though there is no bright-line rule to differentiate the two.[10] The CFPB mentioned several forms of alternative data in its 2017 request for information, including periodic payments data for non-loan products such as phone payments; rent, insurance, and utility bill payments; checking account transaction-level data; data related to a consumer's educational and occupational history; consumer behavioral data; and data derived from a consumer's social media network.[11] Alternative data can also capture the consumer's history of using alternative credit products, such as "payday loans, cash advances, short-term

---

7. *See* Robert B. Avery, Paul S. Calem, Glenn B. Canner & Raphael W. Bostic, *An Overview of Consumer Data and Credit Reporting*, 89 FED. RSRV. BULL. 47, 48 (2003); Request for Information, *supra* note 4; FED. RSRV. BANK OF N.Y., QUARTERLY REPORT ON HOUSEHOLD DEBT AND CREDIT 2019: Q3, at 3 (2019), https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/hhdc_2019q3.pdf?fbclid=IwAR2yyO5tgmW-5ZawcKGLKNpemcBrgxOZ7SQCwIy-l3pr5Na_MLEDLVBMdmc [https://perma.cc/H5TU-XP5B] (reporting that aggregate household debt totaled $13.95 trillion in the third quarter of 2019).

8. Request for Information, *supra* note 4, at 11184.

9. *Id.* ("[D]ata assembled and managed in the core credit files of the nationwide consumer reporting agencies, which includes tradeline information (including certain loan or credit limit information, debt repayment history, and account status), and credit inquiries, as well as information from public records relating to civil judgments, tax liens, and bankruptcies. It also refers to data customarily provided by consumers as part of applications for credit, such as income or length of time in residence.").

10. *Id.*

11. *Id.* at 11185.

installment loans, rent-to-own and title loans."[12] Many forms of alternative data are also considered "Big Data," which is a distinct concept.[13] Big Data is defined as "high-volume, high velocity, and high-variety" information and extends all the way from data related to consumer payment history to digital footprint data from users of smartphones.[14]

A. How Consumer Credit Scores Are Used

In evaluating the creditworthiness of a consumer, lenders can use their own proprietary scoring models, refer to well-known third-party models such as FICO or VantageScore, or utilize some combination of both.[15] The creditworthiness of a consumer is summarized in the credit score assigned to her, and it is used by lenders to evaluate the consumer's likelihood of defaulting, making significantly delinquent payments, or triggering other negative financial shock.[16] For over thirty years, the third-party models developed by FICO and VantageScore have been the primary ways of scoring consumers seeking credit.[17] In addition to lenders, potential employers and landlords also frequently use credit scores as a way to evaluate potential employees or tenants.[18]

Credit scores and the underlying data are compiled into credit reports (also referred to as "consumer reports") by consumer reporting agencies (CRAs), such as TransUnion, Experian, and Equifax.[19] The data in the credit reports created by CRAs informs the traditional consumer credit score (e.g., FICO or VantageScore) that is reported, and, if lenders choose to use their own proprietary models, they can evaluate consumers using the data captured in the report.[20] For the purposes of reporting traditional consumer credit scores, the

---

12. *The State of Alternative Credit Data*, EXPERIAN 18 (2018), https://www.experian.com/assets/consumer-information/white-papers/alternative-credit-data-paper.pdf [https://perma.cc/6EGA-MGHR].

13. Request for Information, *supra* note 4, at 11184 n.4.

14. Ceylan Onay & Elif Öztürk, *A Review of Credit Scoring Research in the Age of Big Data*, 26 J. FIN. REGUL. & COMPLIANCE 382, 382–83 (2018).

15. Request for Information, *supra* note 4, at 11184.

16. Hurley & Adebayo, *supra* note 1, at 153–54; Request for Information, *supra* note 4, at 11184.

17. *See* Hurley & Adebayo, *supra* note 1, at 154; Latoya Irby, *VantageScore Credit Score Overview*, THE BALANCE, https://www.thebalance.com/vantagescore-credit-score-overview-961140 [https://perma.cc/5DL6-7HG9] (May 1, 2020).

18. Hurley & Adebayo, *supra* note 1, at 154.

19. *Id.*

20. Request for Information, *supra* note 4, at 11184.

CRAs maintain and utilize traditional data that falls into four categories: header data (data that helps identify the consumer), public record data, tradeline data on each loan or line of credit the consumer has obtained, and inquiry data that depicts the number of inquiries made into the consumer's credit files.[21]

The credit score reported in a consumer's credit report is not entirely a reflection of the individual consumer's likelihood of default; rather, it reflects the historical rates of default within a group of borrowers who share the same credit score.[22] In other words, borrowers are segmented into various score bands, and each score band has a corresponding predicted rate of default, where consumers in the higher score bands have lower historical rates of default and consumers in the lower score bands are viewed as high-risk borrowers due to higher historical rates of default.[23] Lenders use these scores to determine which consumers they view as creditworthy; to this end, lenders establish a cutoff score below which they will not extend credit.[24]

B. The Purpose of Using Alternative Data for Consumer Credit Scoring

Given how important a credit report with a reportable score is to accessing consumer credit, consumers who either do not have any credit history on file with a CRA or have not yet generated a sufficient credit history for a traditional credit score have a very difficult time accessing consumer credit, and they may need to resort to high-interest substitutes that can further disadvantage them financially.[25] As of 2015, the Bureau has estimated that twenty-six million Americans are "credit invisible" and have *no file* at the three major CRAs, while another nineteen million do not have sufficient data on file to develop a traditional credit score.[26] This population of forty-five million unscorable consumers is primarily comprised of individuals from historically underprivileged communities, including consumers from Black, Hispanic, and low-income backgrounds.[27] Of the adults who applied for credit in 2018, nearly one-third were

21. Peter Carroll & Saba Rehmani, *Alternative Data and the Unbanked*, OLIVER WYMAN 14 (2017), https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2017/may/Alternative_Data_And_The_%20Unbanked.pdf [https://perma.cc/GH5K-HYZ8].

22. *Id.* at 4.

23. *Id.* at 5.

24. *Id.* at 4.

25. Request for Information, *supra* note 4, at 11184.

26. *Id.*

27. *Id.*

denied credit or offered less than what they applied for.[28] Of that sample, 76 percent of those denied credit were Black and Hispanic.[29]

Alternative data and its corresponding algorithms have been recognized by both lenders and regulators alike as potential tools to increase inclusion of consumers from historically disadvantaged communities into the financial system.[30] Additionally, the use of alternative data may allow lenders to identify creditworthy consumers who would otherwise fall into score bands below the cutoff in traditional credit scoring systems; in other words, alternative data may be able to improve the granularity of the score bands used to compute credit scores.[31] This added granularity could help borrowers in lower score bands access credit at lower interest rates, as there would be a way to differentiate between borrowers who are near the cutoff but still creditworthy.[32] Finally, in addition to increasing financial inclusion and enhancing estimates of creditworthiness, alternative data also has the potential to improve the timeliness, or temporal relevance, of assessments and decrease transaction costs for lenders by improving the accuracy of decision-making.[33]

The characteristics of good alternative data include timeliness and accuracy, relevance to the intended behavioral prediction, regulatory compliance, "broad and consistent coverage" across consumers, "consumer-specific" elements (rather than elements based on consumer segments), and "orthogonality"—the notion that the alternative data can work in conjunction with traditional data to improve the "predictive accuracy" of the credit score.[34] Logically, these characteristics are relevant to any data used in predictive modeling: the data should be related to the purpose of the model, comply with existing regulations, and be as individualized as possible, given that the model's ultimate prediction will be individualized. Moreover, alternative data should be thought of as a way to enhance or improve existing data.[35]

---

28. BD. OF GOVERNORS OF THE FED. RSRV. SYS., *supra* note 4, at 26.

29. *Id.*

30. *See* Press Release, Bd. of Governors of the Fed. Rsrv. Sys., Consumer Fin. Prot. Bureau, Fed. Deposit Ins. Corp., Nat'l Credit Union Admin. & Off. of the Comptroller of the Currency, Interagency Statement on the Use of Alternative Data in Credit Underwriting (Dec. 3, 2019), https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20191203b1.pdf [https://perma.cc/46NW-KSNB].

31. Carroll & Rehmani, *supra* note 21, at 2.

32. *Id.* at 10.

33. Request for Information, *supra* note 4, at 11186.

34. Carroll & Rehmani, *supra* note 21, at 9.

35. *See id.* at 8–9.

January 2019 research from the Federal Reserve Bank showed optimistic results for using alternative data in consumer lending scoring models.[36] Specifically, LendingClub consumer ratings were more strongly correlated with loan performance and interest rates than ratings created by traditional lenders (i.e., banks).[37] Additionally, financial inclusion was increased, and borrowers who would have otherwise been in a subprime score band were able to gain access to credit.[38] While this may be very promising news, there are several regulatory issues that must be addressed before similar results can be seen across the credit-scoring industry.

C. Current Issues with Using Alternative Data in Consumer Credit Risk Scoring

One of the primary issues stemming from the use of alternative data is the ability to use nontraditional data—especially behavioral data unrelated to a consumer's financial status or history.[39] Internet browsing-related data (e.g., search history) and social network data—including an analysis of where the consumer is perceived to fall within the hierarchy of her social network—fall within this category of alternative data.[40] One fintech firm, ZestFinance, collects behavioral data from its own website, including how quickly a consumer scrolls through the firm's consumer disclosures to represent how carefully the consumer arrives at a decision.[41] Other examples include club memberships, online shopping behavior, and online profile data.[42] In addition to the potential regulatory violations (discussed in Part II), this type of data, on its face, is not intuitively relevant to the intended behavioral prediction: the consumer's propensity to default or make delinquent payments.[43] Indeed, experts

36. *See* Julapa Jagtiani & Catharine Lemieux, *The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform* 18 (Fed. Rsrv. Bank of Phila., Working Paper No. 18-15, 2019), https://www.philadelphiafed.org/-/media/frbp/assets/working-papers/2018/wp18-15r.pdf?la=en [https://perma.cc/FR8U-3PQC].

37. *See id.*

38. *See id.*

39. Hurley & Adebayo, *supra* note 1, at 164–68.

40. *See Hearing*, *supra* note 3, at 5.

41. *See* Hurley & Adebayo, *supra* note 1, at 164–65 (describing the scoring model used by ZestFinance, "one of the most prominent players in the alternative credit-scoring and underwriting industry").

42. *Id.* at 165 (providing examples of other alternative data inputs used by firms in consumer credit scoring).

43. *Id.* at 164–65.

criticize the use of data elements that are not inherently tied to creditworthiness in the "all data is credit data" approach.[44]

1. Data Quality and Accuracy Issues with Alternative Data

The use of alternative data raises additional challenges—data quality and accuracy issues.[45] CRAs run into accuracy issues even with traditional data, which uses a smaller set of data relative to most forms of alternative data.[46] It makes sense that with alternative data—where there is a much higher volume of data generated—size itself becomes an issue when it comes to ensuring that the collected data is accurate.[47]

In 2013, fifteen National Consumer Law Center employees conducted a survey to view the consumer data collected on each of them by four Big Data brokers: eBureau, ID Analytics, Intelius, and Spokeo.[48] Errors were found in nearly two-thirds of the sixty reports generated.[49] The survey participants had to take several steps to request the reports, verify their identifies, and sometimes pay to receive the individual consumer reports.[50] There was a broad range of types of information errors; most errors were observed in data related to address and residence information, education, family members, social profiles, and income.[51]

Even more recently, research has shown that some of the Big Data "continuously mined" from consumer activities "may incorporate a high degree of inaccurate information."[52] The CFPB, in its 2017 request for information, indicated that the data accuracy issues in alternative data are greater than those observed in traditional data. This inaccuracy is attributable to either the nature of the data itself or lower standards for data quality and accuracy in uses of the alternative data outside of credit scoring.[53]

---

44. *Id.* at 158, 164.
45. *Id.* at 152.
46. *Id.* at 152–53.
47. *Id.* at 153.
48. Persis Yu, Jillian McLaughlin & Marina Levy, *Big Data: A Big Disappointment for Scoring Consumer Creditworthiness*, NAT'L CONSUMER L. CTR. 15 (Mar. 3, 2014), https://www.nclc.org/images/pdf/pr-reports/report-big-data.pdf [https://perma.cc/VN4Q-TSF6].
49. *Id.* at 18.
50. *Id.* at 16–17.
51. *Id.* at 18.
52. Hurley & Adebayo, *supra* note 1, at 153.
53. Request for Information, *supra* note 4, at 11187.

2. The Role of Algorithms in Consumer Credit Scoring

Before addressing the issues stemming from the algorithms and tools used in conjunction with alternative data, it is important to understand the relationship between these algorithms and Big Data used to analyze or identify alternative data for credit scoring. An algorithm is a model that uses a computational process to analyze input data and generate output data.[54] Firms can develop algorithms to identify relationships between various types of input data. These firms can identify relationships either through supervised machine learning, where a researcher assesses how different data elements impact the desired output (e.g., predicted default rate), or through unsupervised learning, where the algorithm identifies relationships between data inputs and identifies patterns in data regardless of how they relate to the specific desired output variable, if there is one.[55]

The extremely complex credit-scoring algorithms are designed to analyze and identify relationships within a high volume and broad variety of Big Data, especially nontraditional data (e.g., behavioral data) generated from consumer social media and spending records.[56] At times—and especially in the context of supervised algorithms—raw data must be transformed into data sets.[57] Data transformation can involve complex steps that incorporate multiple layers of information, sometimes generating "metavariables" that summarize relationships between multiple data points.[58] Data transformation is just one component that complicates the use of alternative credit scoring systems. Once the data is actually translated, it is analyzed through the many complex models that comprise the algorithm.[59] The complexity of this process makes it highly unlikely that consumers could identify data quality issues in the input data feeding the models.[60] This complexity demands investigation of the gaps and recent developments in the existing regulatory scheme, which must adapt in a timely manner to address the risks stemming from such a complex process.

54. *See* Hurley & Adebayo, *supra* note 1, at 159 (defining algorithms as "any well-defined computational procedure that takes some value, or set of values, as input and produces some value, or set of values, as an output.") (citation omitted).

55. *Id.* at 161–62.

56. *Id.* at 152, 163.

57. *See id.* at 174, 176.

58. *Id.* at 176.

59. *Id.* at 181.

60. *Id.* at 182.

II. THE EXISTING REGULATORY SCHEME AND ITS CURRENT SHORTFALLS

A. The Fair Credit Reporting Act

The purpose of the Fair Credit Reporting Act (FCRA) is to ensure "[a]ccuracy and fairness of credit reporting" and require CRAs to "adopt reasonable procedures" to protect the "confidentiality, accuracy, relevancy, and proper utilization" of sensitive consumer information "in a manner which is fair and equitable to the consumer."[61] The FCRA protects consumer privacy by limiting how consumer credit information can be communicated or used, and it gives consumers ways to access the data underlying their credit scores, along with an understanding of how third parties use consumer data in relation to credit, employment, and insurance decisions.[62]

The FCRA primarily regulates CRAs, though it also places obligations on third-party users of consumer reports and third-party furnishers of data who provide CRAs with consumer information.[63] The statute defines a CRA as "any person which . . . regularly engages in whole or in part in the practice of assembling or evaluating consumer credit information or other information . . . for the purpose of furnishing consumer reports to third parties."[64] The last qualifier in this definition—the requirement that consumer credit information is furnished by CRAs *to third parties*—can help firms employing alternative credit data avoid governance by the FCRA.[65] Firms that do not resell the data to third parties but still use the data for credit-scoring purposes are not currently within the scope of the FCRA because of the limitation in the statutory definition of CRA.[66]

While the FCRA does not promulgate definitions for users and furnishers of consumer information, it does outline duties for entities who use and provide consumer reports.[67] Users of consumer reports can only obtain the report if they have a statutorily permissible purpose, such as credit transactions pertaining to a consumer seeking credit, employment, or insurance underwriting.[68] Furnishers of credit

61. Fair Credit Reporting Act, Pub. L. No. 91-508, § 602, 84 Stat. 1114, 1128 (1970) (codified as amended at 15 U.S.C. §§ 1681–81x (2018)).

62. Hurley & Adebayo, *supra* note 1, at 184.

63. *See* 15 U.S.C. §§ 1681e, 1681m, 1681s-2.

64. *Id.* § 1681a(f).

65. Hurley & Adebayo, *supra* note 1, at 187.

66. *See id.*

67. *See* 15 U.S.C. §§ 1681a, 1681m, 1681s-2.

68. *Id.* § 1681b(a)(3)(A)–(C).

reports have statutorily defined duties including providing accurate information to CRAs, and in cases where information is inaccurate, taking timely steps to provide notice to the CRA that the information is inaccurate or may be in dispute.[69]

Users of consumer reports must also notify a consumer when they use a CRA-provided report as the basis for an adverse action against a consumer. An adverse action includes the denial of credit, insurance, or employment opportunities or an increase in rates charged for credit or insurance.[70] The user of the consumer report is also required to provide the relevant numerical credit score used in the determination, all of the key factors (or the top four) that adversely affected the consumer's credit score, the date the relevant credit score was created, and the name of the entity that provided the credit score.[71]

Notably, there is no statutory requirement for a detailed explanation of factors that are included as part of adverse action notice.[72] For example, users (and CRAs, when a consumer requests her credit score) do not have to explain how the factors are weighted or what the factors even mean; vague "phrases like 'type of bank accounts' and 'type of credit references'" are acceptable, even though they do not help a consumer reliably understand how her individual actions impact her credit score.[73] This type of phrasing will be even less useful for consumers when firms employ extremely complex alternative scoring methodologies to provide adverse action notices.

Whether firms using alternative data generate reports that fall into the statutory definition of consumer reports indicates how these firms will be regulated.[74] The definition of a consumer report is broad and includes any information from a CRA that influences the creditworthiness or reputation of a consumer.[75] However, the definition is limited by the requirement that information must relate to "an identifiable person," and not a subset of individuals in the aggregate, such as a household or all of the individuals who live in the same neighborhood.[76] The Federal Trade Commission (FTC) has adopted the view that even if information is not linked to an

69. *Id.* § 1681s-2.

70. *Id.* § 1681a(k).

71. *Id.* §§ 1681m(a)(2), 1681g(f)(1)(B)–(E).

72. Danielle Keats Citron & Frank Pasquale, *The Scored Society: Due Process for Automated Predictions*, 89 WASH. L. REV. 1, 17 (2014).

73. *Id.*

74. *Hearing*, *supra* note 3, at 10–11.

75. *See* Hurley & Adebayo, *supra* note 1, at 185.

76. *Id.*

identifiable person (e.g., by name), it qualifies as a consumer report "if it could be reasonably linked [back] to the consumer."[77] The FTC's view is important given the expansion in the types of alternative Big Data that are used, some forms of which may be traceable back to individual consumers.[78]

B. The Equal Credit Opportunity Act

The Equal Credit Opportunity Act (ECOA) prohibits discrimination by a creditor against an applicant on the basis of certain protected characteristics during any aspect of a credit transaction.[79] The following characteristics are protected against discrimination: race, color, religion, national origin, marital status, sex, age, public assistance status, and exercise of rights under the Consumer Credit Protection Act.[80] While "credit transaction" is not explicitly defined in the ECOA, its enacting regulations define the term broadly and include all aspects of an applicant's interaction with a creditor related to an application for new or existing credit.[81] Experts interpret the breadth of these definitions to mean that in addition to CRAs, fintech firms who provide consumer credit scores or credit assessment tools are within the scope of the ECOA, even if they do not make the ultimate lending decisions.[82]

There are two ways for a plaintiff to allege discrimination under the ECOA: she can allege disparate treatment, disparate impact, or both.[83] To allege disparate treatment, the plaintiff must make a showing that the lender based its decision to extend credit based on "a discriminatory intent or motive."[84] To allege disparate impact, the plaintiff must show that the lender's practice resulted in a "disproportionately negative impact on a prohibited basis" regardless of whether the lender lacked the intent to discriminate.[85] Plaintiffs alleging disparate impact in the ECOA context may face heightened causation requirements if firms argue that their

---

77. *Id.* at 186.

78. *See id.* at 185–86.

79. Equal Credit Opportunity Act, Pub. L. No. 93-495, § 701(a), 88 Stat. 1500, 1521 (1974) (codified as amended at 15 U.S.C. §§ 1691–91f (2018)).

80. 15 U.S.C. § 1691(a)(1)–(3).

81. 12 C.F.R. § 202.2(m) (2020).

82. Hurley & Adebayo, *supra* note 1, at 191.

83. *Id.* at 192.

84. *Id.*

85. *Id.*

methodologies simply mirror "existing forms of systemic bias."[86] Even if a plaintiff is able to establish causation, the defendant can establish that the challenged practice or policy is in place for a valid business purpose.[87] The purpose need not be essential or completely necessary to business objectives; it must simply be relevant to the entity's business objectives.[88]

If a defendant can justify a challenged policy or practice with a legitimate business objective, the plaintiff must still provide an alternative method that mitigates the disparate impact but is still equally effective in fulfilling the defendant's business objectives.[89] Given that credit-scoring algorithms—regardless of whether they use traditional or alternative data—are trade secrets, it is an enormous challenge for a plaintiff to overcome information asymmetries and gain an understanding of the tools used by these firms.[90]

C. Gaps in the Existing Regulatory Framework

The existing regulatory framework, comprised primarily of the FCRA and ECOA, is inadequate to address the challenges posed by the use of alternative Big Data and complex proprietary algorithms in credit scoring. Hurley, Adebayo, and Lee's proposed model legislation, the Model Fairness and Transparency in Credit Scoring Act (FaTCSA) is designed to address the four major challenges the authors identified in the use of alternative credit-scoring tools and the insufficient regulatory framework surrounding them.[91] This model legislation addresses transparency and accuracy issues arising from the shift towards alternative Big Data, as well as the possibility of discriminatory and biased scoring practices and the opportunities for firms employing these scoring practices to identify and exploit consumers from disadvantaged backgrounds.[92] The following Subsections discuss the gaps in the regulatory framework and present potential enhancements to the recommendations incorporated in FaTCSA.

---

86. *Id.* at 194.

87. *Id.*

88. *Id.* at 194.

89. *Id.* at 194–95.

90. *Id.* at 195.

91. *Id.* (defining the four challenges as "1) insufficient transparency, 2) input data that are potentially inaccurate, 3) the potential for biased and discriminatory scoring, and 4) the risk that these tools will be used to target vulnerable consumers.").

92. *Id.* at 197–99.

1. Transparency and Accuracy Issues

The purpose of transparency in credit scoring is to ensure that scoring entities are held to standards reflecting their importance to society, as well as to ensure that consumers who are unable to access credit are able to understand the steps they must take to do so.[93] It is unlikely that consumers will be able to take steps to improve their behaviors or identify mistakes in their credit reports if they are unaware of the factors that impact their credit scores, especially given the use of nontraditional data derived from Big Data that may not have been adequately tested for accuracy.[94]

The biggest transparency-related issue with firms using alternative credit scoring methodologies is the secrecy of the methodologies used for developing the credit scores.[95] Because the methodologies are protected trade secrets, it is difficult to know whether they conform to industry best practices or have been evaluated and developed through consultation with experts.[96] The disclosure requirements in FaTCSA require firms to share their methodologies with a state attorney general or a body acting under the supervision of the state attorney general, but only upon request.[97]

Given how important these scoring methodologies are to disadvantaged populations, the existing regulatory scheme can be enhanced by making periodic alternative data-related methodology disclosures *mandatory*, at least to federal regulators. Some experts argue that to adequately test scoring systems, regulators would require the input data used by scoring algorithms, along with source code programmer notes, and other correlations integrated into these algorithms.[98] Additionally, commercial credit-rating agencies, such as Moody's, publish detailed summaries of their assessment methodologies.[99] Notably, commercial credit-rating agencies use data

93. *See* Citron & Pasquale, *supra* note 72, at 11, 18; Hurley & Adebayo, *supra* note 1, at 196–98 (noting that transparency issues in Big Data-driven alternative scoring systems stem from the complexity of the algorithms used, the volume, breadth, and quality of data employed, and the lack of insight into scoring methodologies developed and used by firms in the consumer credit scoring industry).

94. *See* Citron & Pasquale, *supra* note 72, at 11; *see also* Yu et al., *supra* note 48, at 14–16.

95. Citron & Pasquale, *supra* note 72, at 33.

96. *See id.* at 5, 25.

97. Hurley & Adebayo, *supra* note 1, at 207.

98. *See* Citron & Pasquale, *supra* note 72, at 25.

99. *See, e.g.*, *Rating Methodology: Banking*, MOODY'S INVS. SERV. (Jan. 7, 2016), https://www.moodys.com/researchdocumentcontentpage.aspx?docid=pbc_186998 [https://perma.cc/BYH4-WQKD].

that has been audited, and therefore regulated, for accuracy.[100] While there may be more competition in the consumer credit-scoring market due to the proliferation of fintech firms that employ consumer-scoring methodologies, it does not make sense for firms to equate sharing information with regulators with disclosing trade secrets to potential competitors.

The FCRA does not adequately address the transparency issues arising from the use of alternative data in consumer credit scoring because it does not place any limitations on the types or categories of data that can be used to evaluate consumer credit. Additionally, it is not feasible for consumers to assume that every data point collected about them may in some way impact their credit scores.[101] For example, the FCRA does not distinguish between the inclusion of positive consumer data on timely payments or negative consumer data reflecting late payments.[102]

While it is true that the inclusion of both positive and negative data may increase the overall accuracy of credit scoring, the collection of alternative data related to utility payments can further disadvantage financially vulnerable consumers, especially those who live in locations with harsher weather.[103] For example, an individual may need to choose between making or deferring a utility payment for heat so that she can obtain help under federal assistance programs that require her to defer payments before she becomes eligible for the program benefits. In such scenarios, consumers may be forced to choose between obtaining needed assistance for nondiscretionary products or damaging their credit scores.[104]

Even if Congress had passed the Credit Access and Inclusion Act of 2019, which would have amended the existing FCRA to permit the reporting of positive data on lease agreements, as well as utility and telecommunications services, there would still be gaps stemming from negative data that could harm the credit scores of financially disadvantaged consumers.[105] The amendment would have prohibited the reporting of negative data when the consumer has commenced a payment plan to remedy late payments but does not outline requirements for the timing of the reporting.[106] It also fails to place

100. *See Credit Rating*, CORP. FIN. INST., https://corporatefinanceinstitute.com/resources/knowledge/finance/credit-rating/ [https://perma.cc/78C2-U2QQ] (last visited Jan. 31, 2021).

101. *See* Hurley & Adebayo, *supra* note 1, at 189.

102. *See id.*; Carroll & Rehmani, *supra* note 21, at 12.

103. *See* Carroll & Rehmani, *supra* note 21, at 12; Yu et al., *supra* note 48, at 13–14.

104. *See* Yu et al., *supra* note 48, at 13–14.

105. *See* Credit Access and Inclusion Act of 2019, S. 1828, 116th Cong. (2019).

106. *See id.* § 2(a).

an obligation on the firms to discuss or offer payment plans prior to reporting negative data to CRAs.[107] Data can be reported and transferred much faster than payment plans can be established, and if the data reporting precedes the consumer's opportunity to protect her credit score, the limitations on negative data in the statute are not very effective.

2. Discrimination, Biased Scoring Practices, and Potential Exploitation Issues

The existing ECOA does not adequately protect consumers, especially those from disadvantaged groups, against accidental biases built into alternative Big Data credit-scoring systems.[108] Additionally, it does not prevent firms from using this biased data to target specific consumer groups with financial products with exceptionally unfavorable terms that they would not offer other consumers.[109] Many ECOA-related issues stem from algorithms using nontraditional data that may evaluate consumers based on societal associations or protected characteristics rather than individual creditworthiness.[110]

The use of data highly correlated with a prohibited characteristic can introduce similar biases into alternative consumer credit-scoring methodologies that the ECOA was designed to prevent.[111] Because of the enormous volume of Big Data analyzed by consumer credit-scoring algorithms, they may "indirectly consider sensitive characteristics, such as race, even when those characteristics are not directly designated as input values."[112] These algorithms are designed to identify patterns and correlations between hundreds of variables. It is easy to imagine that these algorithms identify correlations between different behavioral traits that vary by culture, social status, and characteristics that are protected under the ECOA.[113] Given that these systems are so complex and analyze such a high volume of data that they are able to find relationships between unrelated and *random* variables, it is even more apparent certain variables, especially nontraditional behavioral variables that appear

107. *See id.*

108. *See* Hurley & Adebayo, *supra* note 1, at 197, 199.

109. *See id.* at 199.

110. *See* Citron & Pasquale, *supra* note 72, at 14; *see also* Yu et al., *supra* note 48, at 27–28.

111. 15 U.S.C. § 1691(a)(1)–(3) (2018); Hurley & Adebayo, *supra* note 1, at 196.

112. *See* Hurley & Adebayo, *supra* note 1, at 182.

113. *See* Yu et al., *supra* note 48, at 14.

nondiscriminatory and neutral, are highly correlated with protected characteristics.[114]

In addition to indirectly using prohibited characteristics in assessing consumer credit risk, these algorithms may also be developed and trained using data that does not have adequate coverage across all groups of people, one of the most important characteristics of a good data source.[115] Consumers from different racial and cultural backgrounds may access the internet in ways that leave different types of digital footprints (e.g., using a mobile phone versus a computer), and people from some cultures are more likely to visit certain social media platforms than others.[116] For these reasons, using nontraditional behavioral data that is highly correlated with certain protected characteristics in a consumer credit-scoring algorithm can introduce bias against a protected group.[117] The implication of introducing this type of bias is that a consumer who has a certain unchangeable characteristic (e.g., race or national origin) may receive a rejection or less favorable lending terms than she would have gotten had she not had those characteristics.[118] In other words, while alternative data can be very beneficial for the expansion of credit, other forms of alternative data may introduce prohibited biases into consumer credit scores.[119]

Additionally, the ECOA does not expressly protect consumers from discrimination based on sexual orientation.[120] The model FaTCSA includes sexual orientation as one of its protected characteristics.[121] It is probably fairer to consumers from all sexual orientations, especially those that are considered "nontraditional," to clearly prohibit discrimination on this basis rather than infer protections from a different characteristic, such as sex.[122]

III. CLOSING THE GAPS IN THE EXISTING REGULATORY FRAMEWORK

A disclosure regime such as the one outlined in FaTCSA that requires routine disclosures and attestation that the methodologies used are not discriminatory is a good first step in reducing the risk of

114. *See id.*
115. *See* Hurley & Adebayo, *supra* note 1, at 199; Carroll & Rehmani, *supra* note 21, at 9.
116. Yu et al., *supra* note 48, at 27.
117. *See id.* at 27–28.
118. *Id.* at 27.
119. *Id.* at 28.
120. Hurley & Adebayo, *supra* note 1, at 192.
121. *Id.* at 205.
122. *Id.* at 192.

ECOA violations.[123] However, additional steps can be taken to better ensure that credit scorers using alternative scoring systems are doing so in a nondiscriminatory way.

Mandating disclosures from firms involved in consumer credit scoring is a potentially effective way to increase transparency.[124] While the model FaTCSA's disclosure requirements are robust and shift the onus of verifying accuracy from the consumer to the credit-scoring firms, they should more stringently trace how input data is transformed into data consumed by scoring algorithms, also known as "data lineage."[125] The model FaTCSA requires routine and public disclosures regarding the types and classifications of data, the sources and transformations of this data, the methods used to collect it, and the particular data points or set of data points that the scoring models treat as significant. It also outlines credit-scoring standards, including requirements that "data must be regularly tested for accuracy, verifiability, and traceability."[126] However, the model FaTCSA does not have explicit disclosure requirements related to the traceability of the data.[127] Consumers may find it difficult to understand where a listed data category or source was truly derived from because many types of alternative data are transformed several times before they are ultimately used by algorithms.[128]

The FCRA or supplemental legislation should also include enhanced standards for adverse action notices for all CRAs and firms that use alternative credit-scoring systems. Because it is permissible for credit scorers to provide vague explanations in adverse action notices, consumers have very limited insight into why an adverse decision was made.[129] A potential enhanced reporting notice should not only include granular data points denoting specific types of accounts and behaviors that influenced the decision to deny credit or a favorable rate to a consumer, but it should also provide the consumer with steps she can take to improve her score, as these may not be immediately clear given the breadth of alternative data and techniques used to collect it. Most importantly, an enhanced reporting notice should inform consumers how the factors were weighed relative to one another so the consumer is aware of which steps to prioritize.

---

123. *Id.* at 199–200.
124. *Id.* at 197.
125. *Id.*
126. *Id.* at 206.
127. *See id.* at 204.
128. *Id.* at 175–76.
129. Citron & Pasquale, *supra* note 72, at 17.

Additionally, lenders using alternative data-scoring systems should disclose how they define "creditworthiness" so that consumers can gain a better understanding of the standards they are being held to. These disclosures will also hold lenders to an appropriate standard for credit scoring.[130] The model FaTCSA, like the regulations promulgated to enforce the ECOA, requires that credit-scoring systems are in place with the purpose of predicting a consumer's creditworthiness.[131] However, neither the FCRA nor the ECOA explicitly requires that consumer credit scorers disclose what their definitions are.[132] The lack of such a requirement is significant because "a poorly-crafted definition could also lead to inadvertent discrimination" if the definition is not tailored in a way that prevents bias against protected characteristics.[133] Implementing a regulatory requirement that consumer credit scorers must define creditworthiness could also be a good starting point to develop or enhance model risk-management practices for consumer credit-risk methodologies.

The use of a consumer's internet-browsing data, including data based on her social media network, should either be curtailed or prohibited in consumer credit scoring because it introduces the risk that a consumer will be evaluated based on negative attributes that she, specifically, does not possess.[134] In November 2019, the state of New York adopted Assembly Bill A5294, which prohibits a CRA or lender from using data derived from an individual consumer's social media network in its credit-scoring methodology.[135] The amendment was proposed to target a potential move toward integrating

130. Hurley & Adebayo, *supra* note 1, at 173.

131. *Id.* at 199–200.

132. *See* Yu et al., *supra* note 48, at 20.

133. Hurley & Adebayo, *supra* note 1, at 173.

134. *See id.* at 174–75. *But see* Arnoud Boot, Peter Hoffmann, Luc Laeven & Lev Ratnovski, *Financial Intermediation and Technology: What's Old, What's New?* 11–12 (Eur. Cent. Bank, Working Paper No. 2438, 2020), https://www.ecb.europa.eu/pub/pdf/scpwps/ecb.wp2438~d0d447b9b6.en.pdf [https://perma.cc/A82D-LMTX] (suggesting that "Bigtech finance is most effective when traditional financial intermediation is undersupplied" and noting that Alibaba's affiliate company Ant Financial "extends more credit lines in rural areas of China with a limited presence of banks").

135. *See, e.g.*, Act of Nov. 25, 2019, ch. 540, § 2, 2019 N.Y. Laws 540 (codified as amended at N.Y. GEN. BUS. LAW § 380-j(h) (McKinney 2021)). Prior to forwarding the legislation to Governor Cuomo for his signature, the New York legislature substituted Assembly Bill A5294 for Senate Bill S2302. *See Senate Bill S2302*, THE N.Y. ST. SENATE, https://www.nysenate.gov/legislation/bills/2019/s2302 [https://perma.cc/JH73-75WJ] (last visited Jan. 31, 2021). The text of A5294 is identical to S2302. *See id.*; Assembly Bill A5294, THE N.Y. ST. SENATE, https://www.nysenate.gov/legislation/bills/2019/A5294 [https://perma.cc/CR89-WB8X] (last visited Jan. 31, 2021).

nontraditional data in FICO scores and to protect individuals from bias based indirectly on geography, as many people develop social networks based on their geographic regions.[136] The amendment to New York's general business and banking laws defines "members of a consumer's social network" as "a group of individuals authorized by a consumer to be part of his or her social media communications and network."[137] The amendment specifically states: "No consumer reporting agency shall . . . evaluate . . . the credit worthiness . . . of members of the consumer's social network for purposes of determining the credit worthiness of the consumer."[138]

Similar legislation should be enacted at the federal level rather than leaving it up to individual states to decide whether they want to protect more vulnerable consumers. From the perspective of distributive justice, it is unfair for vulnerable consumers in some states to receive better protections against discrimination in credit scoring while consumers in other states do not. Consumers from financially disadvantaged backgrounds may not have the resources to move from a state with poor consumer protection laws to consumer-friendly states, such as California or New York. Further, many historically underprivileged communities experience regionalized inequality.[139] These communities may live in areas where many policies are unfavorable to them (e.g., social services, criminal justice, and education), not just consumer protection laws.[140] Legislation that enables all consumers in these communities to access credit as a vehicle to increase their financial wealth may be helpful in alleviating these regional inequalities as well.[141]

If passing legislation limiting the use of nontraditional, social media-derived data is not feasible at a federal level, then the agencies regulating firms using alternative credit-scoring systems should operate under the presumption that alternative data is discriminatory. The US Department of Financial Services recommends that insurers located in New York conduct their due diligence and ensure that alternative data does not introduce bias

136. *See Senate Bill S2302*, *supra* note 135.

137. N.Y. GEN. BUS. LAW § 380-a(u).

138. *Id.* § 380-j(h).

139. Jay Shambaugh, Ryan Nunn & Stacy A. Anderson, *How Racial and Regional Inequality Affect Economic Opportunity*, BROOKINGS (Feb. 15, 2019), https://www.brookings.edu/blog/up-front/2019/02/15/how-racial-and-regional-inequality-affect-economic-opportunity/ [https://perma.cc/B8C6-GFLH].

140. *Id.*

141. *See* Citron & Pasquale, *supra* note 72, at 11, 18.

based on any protected characteristics, even if the data is purchased from a third party.[142] Most importantly, the guidance indicates that insurers should be extremely cautious in employing alternative data and states that alternative data should not be used in an "algorithm or predictive model in underwriting or rating unless the insurer can establish that the underwriting or rating guidelines are not unfairly discriminatory."[143] Therefore, this guidance may encourage lenders to employ more rigorous practices in selecting the data used for consumer credit scoring, as well as ensure that the data does not introduce biased correlations into the scoring system. If regulators conduct audits under the rebuttable presumption that alternative data is discriminatory, the burden of proving fairness would shift to the lenders, an idea similar to the FaTCSA model legislation.[144]

If legislative and regulatory changes are slow, then regulatory agencies should encourage firms to seek no-action letters with terms that benefit both the lender and the regulatory agencies such that the regulatory agency can gain insight into the methodologies and credit-risk management practices of the lenders.[145] In 2017, the CFPB issued a no-action letter to Upstart Network, a firm that uses alternative data in addition to traditional data for the purposes of consumer credit underwriting and pricing.[146] The CFPB's issuance of the no-action letter was conditioned on Upstart Network maintaining "a model risk management and compliance plan that requires it to analyze and appropriately address risks to consumers, as well as assess the real-world impact of alternative data and machine learning."[147] These impacts are shared with the CFPB, along with data comparing outputs of alternative and traditional models, information that could be very valuable to the CFPB's efforts to appropriately regulate CSAs and users of consumer reports.[148]

---

142. *Insurance Circular Letter No. 1*, N.Y. ST. DEP'T OF FIN. SERVS. (Jan. 18, 2019), https://dfs.ny.gov/industry_guidance/circular_letters/cl2019_01 [https://perma.cc/RB4T-EHY4].

143. *See id.*

144. *See* Hurley & Adebayo, *supra* note 1, at 198–99 (explaining how the Model FaTCSA shifts the burden of ensuring accuracy from consumers to credit scorers).

145. *See* Patrice Alexander Ficklin & Paul Watkins, *An Update on Credit Access and the Bureau's First No-Action Letter*, CONSUMER FIN. PROT. BUREAU (Aug. 6, 2019), https://www.consumerfinance.gov/about-us/blog/update-credit-access-and-no-action-letter/ [https://perma.cc/Z5UR-7VPJ].

146. *See id.*

147. *Id.*

148. *Id.*

The results provided as part of the no-action letter plan were promising for both consumers and lenders.[149] They indicate that improving transparency, by requiring disclosures and employing adequate risk-management practices, to prevent discriminatory lending practices can improve financial inclusion in consumer credit scoring using alternative data.[150] The Upstart data was promising, with extension of credit increasing by upwards of 20 percent and average APRs decreasing by 15–17 percent across all "tested race, ethnicity, and sex segments."[151] It is notable that the use of alternative data did not completely supplant traditional data in the consumer credit-scoring methodology, and the primary forms of alternative data used were educational attainment or employment history.[152] There was no use of behavioral data from consumer browsing history or social media networks incorporated into these models, as well as no utilities-related data.[153]

IV. FUTURE OUTLOOK

As the use of alternative data continues to gain momentum across various parts of the financial sector, regulatory bodies will need to define the roles they will play. Because regulatory agencies are still seeking to understand the impacts and methodologies surrounding alternative credit systems, fintech firms and their product offerings may be regulated on either a consolidated or fragmented basis.[154] These regulatory decisions will have enormous implications not just for financial inclusion on the consumer side but also for profitability and risk management of lenders who seek to extend consumer credit.

To ensure consumer credit is more accessible, regulators should adopt a disclosure regime and consumer-friendly regulatory framework that provide consumers with a more financially inclusive

---

149. *Id.*

150. *Id.*

151. *Id.*

152. *Upstart Request for No-Action Letter*, CONSUMER FIN. PROT. BUREAU 4 (2017), https://files.consumerfinance.gov/f/documents/201709_cfpb_upstart-no-action-letter-request.pdf [https://perma.cc/DX9A-2664].

153. *See id.* at 5.

154. Luz Maria Salamina, Pratibha Chhabra, Shalini Sankaranarayan & Collen Masunda, *Disruptive Technologies in the Credit Information Sharing Industry: Developments and Implications*, WORLD BANK GRP. 29–31 (2019), http://documents1.worldbank.org/curated/en/587611557814694439/pdf/Disruptive-Technologies-in-the-Credit-Information-Sharing-Industry-Developments-and-Implications.pdf [https://perma.cc/XHJ5-PSG9].

credit system.[155] Such a system would not only ensure more access to consumers in the United States but also to those in developing countries, where there are even larger information asymmetries.[156]

Sahiba Chopra[*]

155. *See id.* at 31.

156. *See id.* at 26; Thomas Wong, *Information Asymmetry and Competition in the Informal Financial Sectors of Developing Countries*, MEDIUM (Feb. 6, 2014), https://medium.com/@thomasawong/information-asymmetry-and-competition-in-the-informal-financial-sectors-of-developing-countries-eb948e9564a2 [https://perma.cc/E2GM-KF4U].

* J.D. Candidate, 2021, Vanderbilt University Law School; B.A., B.S., Boston University 2015. The Author would like to thank her parents Harpreet and Amarjit, her sister Arshi, and her partner Michal for their continued patience and support. The Author would also like to thank the staff of the *Vanderbilt Journal of Entertainment and Technology Law* for their thoughtful feedback and careful edits. Finally, the Author would like to thank Professor Yesha Yadav for teaching an enlightening seminar that spurred her interest in the topic.