

May 19, 2017

Via regulations.gov
Monica Jackson
Office of the Executive Secretary
Consumer Financial Protection Bureau
1700 G Street NW
Washington, DC 20552

Re: Comments in Response to Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process, Docket No. CFPB-2017-0005

The undersigned consumer groups are pleased to submit the following comments in response to the CFPB's Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process. Alternative data has the potential to benefit millions of consumers, whether they are "credit invisible" or they have impaired records with the traditional Big Three nationwide consumer reporting agencies (CRAs) - Equifax, Experian, and TransUnion. However, alternative data is not without its risks.

Much of the analysis of whether alternative data will benefit or hurt consumers depends on several key factors –

- What kind of alternative data is being used?
- How is the alternative data being used?
- What is the accuracy and predictiveness of the data?
- What level of disparate impact does the data have on protected groups, especially communities of color?

If alternative data is used for credit decisionmaking, its use is regulated by the Equal Credit Opportunity Act (ECOA). Unless it is firsthand experience information, most likely it is also regulated by the Fair Credit Reporting Act (FCRA). Compliance with both these laws will be critical for the purposes of accuracy, predictiveness, transparency, and minimizing disparate impact.

A. First Considerations

We appreciate the CFPB's focus on the issue of "credit invisibility." The Bureau has conducted very valuable and enlightening research estimating that 26 million Americans do not have a credit history, and that 18 million Americans do not have a credit score because their histories are too scant ("thin") or old to generate a score.¹ The research also found that credit invisibility has a disparate impact, in that African American and Hispanic (as well as low-income) consumers are more likely to have no credit history or to be unscorable.

¹ See Consumer Financial Protection Bureau, Data Point: Credit Invisibles, May 2015, *available at* http://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf.

The natural reaction to such data is the desire to address this invisibility, to make the invisible seen. To that end, the CFPB, industry, policymakers, and consumer and civil rights advocates have considered the use of alternative data to address credit invisibility. Before doing so, however, we should carefully weigh the risks of alternative data. We urge a cautious and thoughtful approach in developing solutions to credit invisibility, including the use of alternative data. As with so many aspects of credit and financial services, the old adage of “the devil is in the details” rings true.

First, we should challenge the premise that invisibility is always a negative. Certainly, credit invisibility poses real and significant problems for consumers who are actively seeking credit. But in some areas, the lack of a credit history could be better than a negative history. Employment and insurance are two uses of credit report information where no history may be better than solely negative information, and where invisibility may be a benefit. For insurance, a number of state laws are designed to ensure that a consumer with no credit score is not treated worse than someone with an average credit score.² Invisibility might also help consumers avoid lead generators and predatory lenders who target high-cost credit to consumers with low credit scores.

Another important consideration before tackling credit invisibility is to consider the ultimate goal. Access to credit by itself should not be the goal, if there is no consideration of the type of credit involved, whether that credit is fairly priced or predatory, and whether the consumer can afford to repay the credit. The goal should be increase access to affordably priced credit with fair terms that the consumer has the ability to repay.

Not all credit is good credit, as we should be reminded by how the freewheeling granting of credit in the early- and mid-2000s precipitated the foreclosure crisis and the Great Recession. And in the current era, some of the lenders using alternative data have offered credit at triple digit APRs, making them little better than high tech payday lenders.³ Using alternative data may result in giving subprime scores to formerly invisible consumers,⁴ boxing them into high-priced credit and making them targets for predatory lenders.

² The Supreme Court observed the existence of these laws in *Safeco Ins. Co. of Am. v. Burr*, 551 U.S. 47, 73, 127 S. Ct. 2201, 2217, 167 L. Ed. 2d 1045 (2007) (“[a] number of States permit the use of such “neutral” credit scores to ensure that consumers with thin or unidentifiable credit histories are not treated disadvantageously”).

³ See Persis Yu, NCLC, *Big Data: A Big Disappointment for Scoring Consumer Credit Risk*, March 2014, at 7, available at www.nclc.org/images/pdf/pr-reports/report-big-data.pdf (Big Data loan products with APRs of 134% to 448%). We are especially concerned about non-bank fintech lenders using alternative data to make triple-digit APR loans given the Office of Comptroller of Currency’s proposal to allow such lenders to obtain national bank charters, which they could use to evade state usury caps.

⁴ For example, a study by a proponent of utility credit reporting states: “For all those that become scoreable, about one-third [i.e., 33%] scored in the F category, 22% scored in the D category, and 45% scored in the C or higher category.” Michael Turner, et al., PERC, *A New Pathway to Financial Inclusion: Alternative Data, Credit Building, and Responsible Lending in the Wake of the Great Recession*, June 2012, at 13. Thus, over half (55%) of consumers without scores end

B. Conventional alternative data

One of the most critical points in discussing alternative data is that the type of data matters. Not all data is created equal. Some data shows promise, other data is a mixed bag, and some data is harmful enough that it should not be used. We discuss specific types of data below.

Before doing so, however, we have one taxonomic point. The CFPB has stated it considers any data outside of the nationwide CRA files to be alternative data.⁵ However, we would divide it into two types: conventional and Big Data. Conventional alternative data includes information about payment obligations that are periodic and recurring, similar to the loan payment obligations in a credit report, or reflect other mainstream financial information, such as bank account transactions. Big Data includes information such as web searches, social media, and other types of data that seem far removed from the information usually in a credit report.

1. Gas and Electric Utility Payment Information

One of the most often promoted types of alternative data is “full file” utility credit reporting, *i.e.* monthly reporting of customer payments, including late payments, by gas and electric utilities to the nationwide CRAs. Currently, the vast majority of electric and natural gas utility companies only report to the nationwide CRAs when a seriously delinquent account has been referred to a collection agency or written off as uncollectible.⁶

However, we are concerned that full file utility credit reporting could end up harming consumers who already have thick files and visible credit scores.⁷ Furthermore, credit invisible consumers will more likely end up with low scores instead of no scores. This is because many low-income customers would receive negative marks for 30- or 60-day late payments during months when utility costs are high, even though they eventually catch up when costs are lower (*e.g.*, in summer months for cold weather winter states). These negative marks can be damaging, as a single 30-day late payment damages a credit score by as much as 60 to 110 points.⁸

up with a probably subprime score (Ds and Fs) and about 35% end up with a C. Only a few percent of the formerly unscorable consumers ended up with an “A” or “B” score.

⁵ 82 Fed. Reg. at 11,184.

⁶ CFPB, Does my history of paying utility bills, like telephone, cable, electricity, or water, go in my credit report?, May 27, 2016, *available at* <https://www.consumerfinance.gov/askcfpb/1817/does-my-history-paying-utility-bills-telephone-cable-electricity-or-water-go-my-credit-report.html>.

⁷ While we have significant concerns about full-file utility reporting in which consumers do not have a choice as to whether or not to participate, we do not oppose permitting consumers to voluntarily opt-in to utility credit reporting if consent is knowingly and meaningfully given.

⁸ Equifax, Can One Late Payment Affect My Credit Score? Feb. 7, 2014, at <https://blog.equifax.com/credit/can-one-late-payment-affect-my-credit-score/> (“[a]ccording to FICO data, a 30-day delinquency could cause as much as a 90- to 110-point drop on a FICO Score of 780 for a consumer who has never missed a payment on any credit account. In comparison, a consumer with a 680 FICO Score ... would experience a 60- to 80-point drop”).

One study claims that negative information will be minimal because its data shows that less than 3% of consumers earning \$50,000 or less have a single 60-day late utility payment during a one-year period.⁹ Yet data from utility companies and regulators shows that the percentage of utility consumers, especially low-income and energy assistance customers, who are delinquent in paying is much higher:

- Data from Pacific Gas and Electric shows that 6.6% of general residential customers and 17.1% of low-income/energy assistance customers were in arrears by 61 or more days in just one month--March 2017.¹⁰
- In Massachusetts, well over one-third (36.4%) of low-income/energy assistance customers of Eversource Electric were more than 60 days late in paying their bills in May 2016.¹¹
- Among investor-owned electric and natural gas companies in Iowa, nearly 18% of all customers and over 38% of energy assistance customers in March 2017 were late in paying a bill.¹² The number of days late for these delinquencies was not reported.
- For Indianapolis Power and Light Company, the monthly average 60-day delinquency rate among energy assistance customers was 33% between 2012 and 2014.¹³
- Among investor-owned electric and natural gas companies in Pennsylvania, the delinquency rate for low-income customers was 18% (electric) and 26% (gas), respectively, in 2015.¹⁴ The number of days late past 30 days for these delinquencies was not reported.
- In March 2017 in Washington D.C., the delinquency rate among all customers of Potomac Electric Power Company was 25% and for low-income customers was a whopping 75%.¹⁵ The number of days late for these delinquencies was not reported.

⁹ Michael Turner, et al., PERC, *The Credit Impacts on Low-Income Americans from Reporting Moderately Late Utility Payments*, August 2012, at 12.

¹⁰ Pacific Gas and Electric Company, *Quarterly Disconnect Data Report*, California Public Utilities Commission, R. 10-02-005, April 25, 2017.

¹¹ Eversource (f.k.a. NSTAR Electric), *Monthly Credit and Collections Report to Massachusetts Department of Public Utilities*, March 2017.

¹² Iowa Utilities Board, *Residential Customer Statistics: Breakdown by Company for March 2017*, April 17, 2017, *available at* https://iub.iowa.gov/sites/default/files/files/records_center/residential_data/MY2017/Moratorium_03-Mar17.pdf.

¹³ Indianapolis Power and Light Company Response to Citizens Action Coalition Data Request 1.6, Cause No. 44576, Indiana Utility Regulatory Commission, March 2015.

¹⁴ Pennsylvania Public Utility Commission- Bureau of Consumer Services, *Report on 2015 Universal Service Programs and Collections Performance*, *available at* http://www.puc.pa.gov/General/publications_reports/pdf/EDC_NGDC_UniServ_Rpt2015.pdf.

¹⁵ Potomac Electric Power Company, *Compliance Filing pursuant to Public Service Commission of the District of Columbia Orders No. 14293 and 15134*, April 19, 2017, *available at* http://edocket.dcpsc.org/edocket/docketsheets_pdf_FS.asp?caseno=ARDIR2017-01&docketno=8&flag=D&show_result=Y.

In addition to introducing a great deal of negative information into the nationwide CRA files, we are concerned that full file utility credit reporting conflicts with established state utility consumer protections. State laws often include consumer protections intended to protect vulnerable populations from the loss of electric and natural gas utility service during high cost months, and in times of illness or financial hardship.¹⁶ These consumers may sometimes defer full payment of utility bills until later in the year, knowing they are protected from shutoff. Adopting full file utility credit reporting would undermine these health and safety protections.

2. Telecommunications data

Over the past several years, there have been efforts to create scores using data from wireless phone and cable television providers. For example, FICO XD uses data from the National Consumers Telecom and Utility Exchange (NCTUE) as well as LexisNexis to generate an alternative score that can be used for otherwise credit invisible consumers.

Unlike regulated electric and gas service, telecommunications industries have fewer consumer protections that could be undermined by monthly reporting. Thus, they do not raise the same concerns as those sources of data. However, the level of accuracy of the data has not been assessed, nor the impact on consumers who lodge disputes over issues such as cramming, early termination fees, and questionable surcharges.

Transparency is also an issue; consumers may also not be aware that their cell phone and cable payment histories are being supplied to nationwide or specialty CRAs. This is critical for low-income consumers because when they juggle bills, they often do so in consideration of multiple factors including credit reporting.

3. Rental payments

Traditionally, rental data is only reported when a tenant is so delinquent that the account is sent to a debt collector. Efforts to add positive data appear to be promising, especially those efforts that do not report late payments prior to the debt being sent to collections. However, tenants who invoke their rights under state or local laws to withhold rent due to poor conditions should not be penalized.

A variation on using rental payments is to use address stability (*i.e.* number of times a consumer has moved) as a proxy for on-time payments. This is an overly crude measure of rental payment history, because consumers often move due to many reasons other than nonpayment of rent. As the CFPB notes, military servicemembers are one population that moves often. 82 Fed. Reg. at 11187. Another population that moves frequently are students. This lack of precision makes address stability a poor measure, and one goal of improving tools to assess creditworthiness should be greater precision, not less. Also, address stability to assess creditworthiness probably

¹⁶ National Consumer Law Center, Access to Utility Service, Appx. A, (5th ed. 2011), *updated at* www.nclc.org/library.

results in disparate impact, since renters move more often than owners and communities of color have lower rates of homeownership.¹⁷

4. Subprime Credit

Payday loans and other forms of subprime credit are often not reported on traditional credit reports. Adding these types of credit could damage the credit records of these borrowers. High-cost credit is often designed to lead to a cycle of debt, and even merely using a subprime form of credit can negatively affect a credit score.¹⁸ Creating a database of consumers who have used high-cost credit could make them even more vulnerable. Adopting the use of subprime credit data would likely harm consumers who use such credit.

Furthermore, the necessity of using subprime credit as a form of alternative data is questionable, as most payday borrowers are not credit invisible. The authors of one research study analyzing payday loan applications noted that the results of their matching payday borrowers to nationwide CRA files “imply that nearly all of the payday loan applicants had a credit record at the time they applied for their first payday loan.”¹⁹

¹⁷ The homeownership rate for non-Hispanic whites was about 72% at the end of 2016, compared to 42% for African Americans, 46% for Hispanics and 57% for Asians. U.S. Census Bureau, Quarterly Residential Vacancies and Homeownership, First Quarter 2017 - CB17-55, Table 7: Homeownership Rates by Race and Ethnicity of Householder: 2013 to 2017, April 27, 2017, available at <https://www.census.gov/housing/hvs/files/currenthvspress.pdf>

¹⁸ Experian states:

Do finance companies have a negative impact on a credit score?

The presence of a loan finance account can negatively affect your score because these *accounts often carry high interest rates which may hamper your ability to repay and which many lenders view negatively*. However, when paid on time, these accounts can also have a positive effect on your score (if the loan helps you to make your payments in a more timely fashion, for example). (emphasis added)

Experian, Credit Score Basics FAQs, available at www.experian.com/blogs/ask-experian/credit-education/faqs/credit-score-faqs/ (viewed May 15, 2017).

If a finance company loan is viewed negatively because of high interest rates, the use of even more expensive credit sources (such as payday loans) to develop consumer reports would logically create a database of “bad risks.” See, e.g., Ruth Lythe, A Payday Loan Can Stop You Getting a Mortgage: Worrying News for Three Million Who Have Needed Fast Cash, The Daily Mail, July 13, 2012, available at <http://www.dailymail.co.uk/money/mortgageshome/article-2173443/Payday-loans-Worrying-news-million-needed-fast-cash.html#ixzz20beuLWIm> (in United Kingdom, reporting that lenders refusing to grant mortgages to payday loan borrowers).

¹⁹ Neil Bhutta, Paige Marta Skiba, & Jeremy Tobacman, Payday Loan Choices and Consequences (2014) at 12, available at

<http://www.calcfa.com/docs/PaydayLoanChoicesandConsequences.pdf>. The authors matched 250,000 payday loan application files to credit report files in the Federal Reserve Bank of New York’s Consumer Credit Panel (CCP), composed of Equifax data. Since the primary CCP dataset is a five-percent random sample, the authors had expected to match roughly 12,400

5. Bank account transaction data

Bank account transaction data appears to be a promising form of alternative data. First, it incorporates an analysis of ability to repay, since it includes information about both income being deposited into the account as well as expenses being paid out of the account.

Second, bank account data allows for less emphasis on historical events that could have been due to no fault of the consumers. One of our critiques of traditional credit reporting and scoring is that it is overly blunt, failing to distinguish between “can’t pay” and “won’t pay.”

Data suggests that many of the consumers with impaired credit were the victims of unfortunate events such as illness or job loss.²⁰ Bank account data can avoid the need to rely on long historical timeframes such as seven years, given that lenders review 12 months of statements at most when they manually review bank account activity.²¹ This means that negative marks from a period of economic hardship from several years ago need not impair a consumer’s ability to obtain credit for an extended period of time. Also, bank account data can show when there has been a healthy sustained recovery from an extraordinary life event such as a job loss or illness.

However, bank account transaction information raises significant privacy issues, and should only be used when the consumer has knowingly and actively consented to its use. Transaction data includes sensitive information beyond just financial patterns, such as debit card purchases showing where the consumer shops or even which political causes she supports. There will need to be methods to protect the privacy of this information while being able to use account information for credit decisionmaking. One method might be for a third party to analyze the data and then provide its own type of score or recommendation, with protections in place that prevent the third party from revealing individual account level transactions.

Another issue is providing access to third parties, whether they be aggregators or lenders, in order to obtain and analyze the bank account data. As the CFPB knows from the comments submitted in response to its Request for Information regarding consumer access to their own

applicants to the CCP. Indeed, they were able to match 12,151 individuals, suggesting that most payday borrowers had a file with Equifax. *Id.* at 11.

²⁰ About 70 to 80% of consumers with impaired credit or a low score, such as a 600, will actually not default. These may be victims of extraordinary life circumstances who do not default again once they have recovered economically. *See* Chi Chi Wu, NCLC, Solving the Credit Conundrum: Helping Consumers’ Credit Records Impaired by the Foreclosure Crisis and Great Recession, Dec. 2013, at 9-11, *available at* www.nclc.org/images/pdf/credit_reports/report-credit-conundrum-2013.pdf (summarizing research).

²¹ For example, Fannie Mae requires lenders to review 12 months of bank account statements to establish payment activity. Fannie Mae Selling Guide, B3-5.4-03: Documentation and Assessment of a Nontraditional Credit History, August 30, 2016, *available at* <https://www.fanniemae.com/content/guide/selling/b3/5.4/03.html>. Anecdotally, we have heard that some lenders only require 3 to 6 months of bank account statements.

financial information,²² some banks have put up roadblocks to a consumer's ability to share his or her bank account data with third party entities. If a bank can block the consumer from sharing data, then the only lender who can utilize bank account transaction data for credit decisionmaking would be the consumer's own bank. This would limit the consumer's choice and ability to shop around. That is why this issue goes hand in hand with the Bureau's examination of account aggregator issues and whether to require data sharing under Section 1033 of the Dodd-Frank Act.²³

Last, but certainly not least, any meaningful effort to expand access to credit using bank account transaction data must be accompanied by significant reform of overdraft practices. As the CFPB well knows, abusive and unfair practices by financial institutions in the triggering of overdrafts and imposition of related fees has cost consumers billions of dollars.²⁴ These practices then result in shutting out millions of consumers from mainstream bank accounts,²⁵ and will similarly shut them out from credit access based on account transaction data. Bank account transaction data will be contaminated from the impact of unfair overdraft practices unless there is significant reform on this issue.

6. Trended data

The CFPB mentions data showing trends in traditional loan payments, *i.e.*, "trended data," as a type of alternative data. 82 Fed. Reg. at 11185. While it does not quite meet the definition of alternative data, since this data does originate from the files of the nationwide CRAs, the use of trended data is certainly a novel development. One of the drivers of its use is Fannie Mae, which now uses trended data in its Desktop Underwriting program.²⁶

A significant issue around trended data will be the accuracy of payment information. In order for trended data to work accurately, information furnishers, most particular credit card issuers, must provide complete and correct information to the nationwide CRAs about the amount of each monthly payment – not just whether a payment was made that met or exceeded the

²² Request for Information: Consumer Access to Financial Records, Docket No. CFPB-2016-0048, November 17, 2016.

²³ We commend those banks that have made the decision to allow data sharing, such as JPMorgan Chase. Press Release, Intuit, Chase, Intuit to Give Customers Greater Control of Their Information (Jan. 25, 2017), *available at* <https://www.intuit.com/company/press-room/press-releases/2017/Chase-Intuit-to-Give-Customers-Greater-Control-of-Their-Information/>.

²⁴ Consumer Financial Prot. Bureau, CFPB Study of Overdraft Programs: A white paper of initial data findings 15 (June 2013), *available at* www.consumerfinance.gov (estimating that banks earned roughly \$12.6 billion in consumer NSF and overdraft fees in 2011).

²⁵ See Chi Chi Wu, Katie Plat, NCLC, Cities for Financial Empowerment Fund, Account Screening Consumer Reporting Agencies: A Banking Access Perspective (October 2015), *available at* <http://www.nclc.org/images/pdf/pr-reports/Account-Screening-CRA-Agencies-BankingAccess101915.pdf>.

²⁶ Fannie Mae, Trended Credit Data and Desktop Underwriter (DU), 2016, *available at* https://www.fanniemae.com/content/fact_sheet/desktop-underwriter-trended-data.pdf.

minimum required. We have seen that several credit card issuers do not provide such information.

7. Automating income data

The CFPB mentions automating verification of employment and income data as a sort of alternative data or modeling. 82 Fed. Reg. at 11186-87. An example of a CRA that provides automated employment or income verification is The Work Number, a subsidiary of Equifax. The Work Number collects information from employers regarding employment and salary information that can be used to verify income.

We are not opposed to the use of automated income data per se. It does have the potential to make income verification faster and less burdensome. But there are issues around accuracy, in that we have seen several instances where consumers were erroneously reported to be employed and have income when they did not. Furthermore, a limited number of employers furnish information to The Work Number. Employers of other organizations should not be disadvantaged because their employer does not participate in The Work Number. We do not want to create another sort of “invisibility” problem.

C. Big Data

“Big Data” encompasses many types of information, and not all of them are created equal. There are many unanswered questions about the predictiveness and accuracy of these data sources. For example, in one study in which NCLC employees ordered their own reports from four data brokers, these reports were found to be riddled with inaccuracies, ranging from the mundane (a wrong e-mail address or incorrect phone number) to the seriously flawed (wrong occupation or salary estimate).²⁷ Of the 15 records sought, some brokers had as many as 13 reports with errors in them.²⁸

Furthermore, compliance with statutory requirements seems questionable. Any data used for credit decisionmaking, whether conventional or Big Data, implicates the ECOA. In addition, we suspect that many “Big Data” purveyors do not realize the broad scope of the FCRA, as discussed in Section F.1 below, and do not consider themselves to be CRAs when they actually are covered as such.

1. Educational or occupational attainment

The CFPB asks about the potential of using data about a consumer’s educational or occupational attainment. We are highly concerned about the use of this data. First, there are obvious racial disparities in educational and occupational attainment. For example, 93% of non-Hispanic

²⁷ See Persis Yu, NCLC, Big Data: A Big Disappointment for Scoring Consumer Credit Risk, at 18, March 2014, available at www.nclc.org/images/pdf/pr-reports/report-big-data.pdf. (reviewing the accuracy of reports by eBureau, ID Analytics, Spokeo, and Intelius).

²⁸ *Id.*

whites graduate high school, but only 67% of Hispanics and 87% of African Americans do.²⁹ Over 36% of non-Hispanic whites and nearly 54% of Asians have a bachelor's degree, but less than 16% of Hispanics and 23% of African Americans do.³⁰ As for occupation, 40% of non-Hispanic whites and 51% of Asians are employed in management, professional, and related positions, but only 30% of African Americans and 22% of Hispanics are similarly employed.³¹

The use of education and occupational attainment also reinforces other types of inequality. As one researcher summarized the literature “In most studies, parental education has been identified as the single strongest correlate of children’s success in school, the number of years they attend school, and their success later in life.”³² In other words, a consumer’s educational attainment is most strongly linked with the educational level of his or her parents. Thus, judging the consumer’s creditworthiness based on educational attainment is essentially basing credit decisions on the socio-economic class that the consumer was born into, perpetuating economic inequality.

Some products have also looked at the specific school that the consumer attended.³³ This is concerning given research on widespread undermatching – where students do not apply to or enroll in the best colleges likely to admit them – and its converse, overmatching enrollment.³⁴ Undermatching is often a result of a lack of financial resources or general information about the admissions process, which disproportionately affects low-income students. In contrast, researchers found that “[s]tudents from the wealthiest families, from neighborhoods where many adults have college degrees, and from high schools where many students go on to college are less likely to be undermatched but also more likely to be overmatched.”³⁵

The CFPB has raised concerns in the RFI that certain alternative data could harden barriers to economic and social mobility. 82 Fed. Reg. at 11,187. Educational or occupational attainment would probably be top on the list of mobility-impeding data, and would ossify the gaping racial and economic inequality in our country.

²⁹ Camille L. Ryan and Kurt Bauman, U.S. Census: Current Population Reports, Educational Attainment in the United States: 2015, March 2016, at <https://www.census.gov/content/dam/Census/library/publications/2016/demo/p20-578.pdf>.

³⁰ *Id.*

³¹ Bureau of Labor Statistics, BLS Reports, Report 1062: Labor Force Characteristics by Race and Ethnicity, 2015, September 2016, *available at* <https://www.bls.gov/opub/reports/race-and-ethnicity/2015/home.htm>.

³² Anna J. Egalite, How Family Background Influences Student Achievement, *EducationNext*, Vol. 16, No.2, Spring 2016, *available at* <http://educationnext.org/how-family-background-influences-student-achievement/>.

³³ <https://www.credible.com/refinance-student-loans>.

³⁴ Scott Jaschik, Admissions Mismatch, *InsideHigherEd.com* (August 13, 2013) *available at* <https://www.insidehighered.com/news/2013/08/13/study-finds-widespread-mismatching-students-and-colleges-primarily-due-student>.

³⁵ *Id.*

2. Friends and family data, including social media

The CFPB has asked about the use of data related to consumers' friends and associates, including data about connections on social media. Creditworthiness by association raises serious concerns about racial disparities.

Although it is not yet widely used in the United States, consideration of social media and other social networks is more common outside of the United States. One such company, Lenddo, which operates in twenty different countries, may deem a consumer to be less creditworthy if he or she is friends on Facebook with someone who was late paying back a loan to Lenddo.³⁶ Given the segregation that still exists in American society, most people's friends and family are likely to be of the same race, class, and cultural background. Combined with the fact that for the past two decades, study after study has found that African American and Latino communities have lower credit scores as a group than whites (and Asians, when the data is available), using this type of data would raise great concerns.³⁷

Creditors have used other methods to base a consumer's creditworthiness on the characteristics of others. For example, at one point it was reported that American Express lowered a customer's credit limit from \$10,800 to \$3,800, not based on his payment history with the company, but because "[o]ther customers who have used their card at establishments where you recently shopped have a poor repayment history with American Express."³⁸ With this type of analysis, low-income consumers with pristine credit histories could find their big data credit scores lowered simply because they save costs by shopping at low-end outlets whose customers include people who have trouble paying their bills.

Location or geographic neighborhood is another way that creditors have based creditworthiness by association. A recent report by TransUnion highlights this ominous trend:

...aggregated credit data is...helpful to [debt] collectors because it can identify local credit conditions clustered around common demographics. This is especially true for consumers with little or no credit history. For example, if the consumer is living in a ZIP code where the mortgage delinquency rates are climbing or always high, the chance for collection may be significantly less than for those in ZIP codes where the delinquency rate is relatively low and stable.³⁹

³⁶ Katie Lobosco, Facebook friends could change your credit score, CNN.com (August 27, 2013) *available at* <http://money.cnn.com/2013/08/26/technology/social/facebook-credit-score/index.html>.

³⁷ National Consumer Law Center, Past Imperfect: How Credit Scores and Other Analytics "Bake In" Past Discrimination and Perpetuate It, May 2016, *available at* http://www.nclc.org/images/pdf/credit_discrimination/Past_Imperfect050616.pdf.

³⁸ Lori Andrews, Facebook Is Using You, *New York Times*, Feb. 4, 2012, www.nytimes.com/2012/02/05/opinion/sunday/facebook-is-using-you.html?pagewanted=all.

³⁹ Jeffrey S. Morrison & Andy Feltovich, Leveraging Aggregated Credit Data and in Portfolio Forecasting and Collection Scoring, *The RMA Journal*, Oct. 2010, at 47, *available at* www.forecastingsolutions.com/publications/RMA_OCT2010.pdf.

There is already evidence that some companies target different zip codes differently. In 2016, Bloomberg compared Amazon same-day delivery areas with U.S. Census Bureau data and found that in six major metropolitan areas, predominantly black ZIP codes are largely excluded.⁴⁰ While there is no evidence that Amazon makes decisions on where to deliver based on race, and Amazon’s spokesperson said that the ethnic composition of neighborhoods is not part of the data Amazon examines when drawing up its maps, the impact is clear.

Census data, zip code, and location can function as a proxy for race and income and its use by creditors would reflect, and likely exacerbate, racial and socio-economic disparities.

3. Behavioral data, including web browsing

The CFPB has asked about the use of behavioral data about consumers, such as how consumers interact with a web interface or answer specific questions, or data about how they shop, browse, use devices, or move about their daily lives. The use of behavioral data has consistently shown racial bias, despite relying on seemingly racially neutral algorithms. In 2013, Latanya Sweeney, a professor of government at Harvard University, led a research project that concluded that Google searches of names more likely associated with black people often yielded advertisements for a criminal records search in that person’s name.⁴¹ On one website, searches of black-sounding names were 25 percent more likely to yield ads with offers to view the person’s arrest or criminal record.⁴² One industry analyst explained that this result could have occurred because black Google users clicking on those ads more than white users.⁴³ He then asked, “[i]f most of the searches for black names are done by black people...is that racially biased?” The answer is yes. If people of a certain race disproportionately engage display a certain behavior and that behavior is then imputed to *all* members of that race, then that would be racist.

Although, industry analysts dismissed Sweeney’s research as “isolated,” when working for the FTC, Sweeney found similar effects when reviewing the advertisers on *omegapsiphi2011.com*, a website for Omega Psi Phi, a high profile historically black fraternity.⁴⁴ Looking at the total number of ads shown on that website, she found large number of ads prompting viewers to click for “your” arrest records. She also looked at the type of credit cards companies that advertised on

⁴⁰ David Ingold and Spencer Soper, Amazon Doesn’t Consider the Race of Its Customers. Should It?, Bloomberg, April 21, 2016 *available at* <https://www.bloomberg.com/graphics/2016-amazon-same-day/>.

⁴¹ Hiawatha Bray, Racial bias alleged in Google’s ad results, Boston Globe (February 6, 2013) <https://www.bostonglobe.com/business/2013/02/06/harvard-professor-spots-web-search-bias/PtOgSh1ivTZMfyEGj00X4I/story.html>.

⁴² *Id.*

⁴³ *Id.*

⁴⁴ Latanya Sweeney, Federal Trade Commission, Online ads roll the dice (Sep 25, 2014) *available at* https://www.ftc.gov/news-events/blogs/techftc/2014/09/online-ads-roll-dice?utm_campaign=Submission&utm_medium=Community&utm_source=GrowthHackers.com.

this site and found that the credit card ads that appeared on *omegapsiphi2011.com* were offers for cards that were highly criticized on the internet or were generic. None of the highly desirable credit card companies advertised on this site. On an individual ad level, this may not indicate a problem. Taken across the millions of websites on the internet, this differential treatment may lead some segments of our population towards inferior or more expensive products. While these are not examples of companies using this information for credit determinations, they show how difficult it is to generate racially neutral algorithms.

Some other companies, such as German-based Kreditech, also gather information from the manner in which a customer fills out an online application. For example, Kreditech uses information such as the time spent reading information about the loan on Kreditech's website or whether the application is typed in all-caps (or with no caps), in its credit decisionmaking. Though the discriminatory impact of these criteria is unclear, it is worth exploring whether these types of behaviors produce disparities based on race or other protected classes. While this information might be relevant to fraud prevention for some reason, its predictiveness as to creditworthiness is open to question. Consumers should be judged based upon information that is truly predictive of their creditworthiness.

Behavioral data also raises accuracy concerns. As discussed in greater detail in Section F.1, consumers have the right to access their reports to check for accuracy and to dispute items. Given the type and volume of data that would be collected, the process of proactively checking their credit file would be nearly impossible for the typical consumer to do.

Finally, there is information that is clearly discriminatory in its targeting of a specific group. In fact, some companies intentionally target potential customers whose interests and demographic information they have identified through social networking data, web browsing history, and purchase information. Target, for example, can reliably predict which shoppers are pregnant based on the history of products purchased at the store, combined with other demographic information purchased from third-party data brokers.⁴⁵ The CFPB should be very skeptical about the use of this type of data for credit decisionmaking or other financial product or service related purpose.

D. How Alternative Data is Used is Critical

The manner in which alternative data is used is a crucial factor. Using alternative data to create special scores for otherwise unscorable consumers is preferable to the wholesale addition of the same data to traditional nationwide CRA credit reports, where it might harm consumers who already have a credit score. Thus, an effort such as FICO XD, which currently uses alternate data only when a consumer has a thin file or no file, is preferable to a situation where the information in NCTUE is added to the main credit reporting files of the nationwide CRAs.

⁴⁵ Kashmir Hill, *How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did.*, Forbes, Feb. 16, 2012, available at www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did.

In the same vein, the use of alternative data as a “second chance” screen, as opposed to an initial screen, is preferable. With a second chance screen, there is a potential that alternative data could help a credit invisible consumer but would not present risks to a thick file consumer. The riskiest part of a “second chance” score is the possibility that a lender might use it as either an initial screen despite its intended use, or as a sequential second score that could result in a denial for a thick file consumer.

E. Alternative Models

As for alternative models, one of the key issues is transparency – whether consumers understand not only *what* information is being used to assess them, but *how* that information is being used. Having “black boxes” that evaluate creditworthiness should be a thing of the past, as a matter of both fairness and ensuring that consumers are fully educated about financial issues. Consumers should always be able know, access, and understand both what is the information being used in credit decisions and how it is used. They should be able to review the information for inaccuracies so they can dispute errors.

A key element of transparency will be whether consumers receive the information they are entitled to under the ECOA and FCRA, as discussed in Section F below. Users of alternative modeling techniques will need to not only provide ECOA and FCRA adverse action notices, but make sure the notices contain sufficient information to satisfy the ECOA standard that the notice disclose the specific reasons for the action taken,⁴⁶ as defined by Regulation B and its corresponding Official Staff Interpretations.⁴⁷

Another issue with alternative models, as well as existing ones, are assumptions that they are automatically fair, unbiased, or judgment-free. Algorithms and models, despite their perception as impartial, reflect the goals and perspectives of those who develop them. They are not truly free of value judgments or preconceptions. And machine learning models can end up “learning” the prejudices of the society they were created in, as shown by the examples in Section C.3 of how Internet search algorithms “learned” racial biases from search behavior.

F. Regulatory Considerations

When data is used in the decision whether or not to grant credit, its use is regulated by the Equal Credit Opportunity Act – period. If that data is not firsthand experience data, it is also covered by the Fair Credit Reporting Act. These statutes provide the regulatory framework that governs the use of all data for credit decisionmaking, including alternative data.

The key requirements for any data used for credit decisions are that it be accurate, predictive, and transparent. The regulatory schemes for the FCRA and ECOA speak to each of these requirements.

⁴⁶ 15 U.S.C. § 1691(d)(3).

⁴⁷ Reg. B, 12 C.F.R. § 1002.9(b)(2); Official Interpretations of Reg. B, 12 C.F.R. pt. 1002, supp. I, § 1002.9(b)(2). *See generally* National Consumer Law Center, Credit Discrimination § 10.5.4.2 (6th ed. 2013), *updated at* www.nclc.org/library.

1. Fair Credit Reporting Act (FCRA)

As the CFPB knows, the FCRA has a very broad scope of coverage. Information is a consumer report if it is:

- Used or expected to be used or collected in whole or in part to serve as a factor in establishing eligibility for consumer credit;
- Pertains to seven characteristics, which cover an extremely far-reaching range of information – credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, and mode of living;
- Issued by third party that regularly assembles or evaluates such data for money or on a nonprofit cooperative basis.

Thus, almost all third party data collected for credit decisionmaking purposes is likely a “consumer report.”

One of the key issues with alternative data is the level of accuracy of the data. Accuracy is not just a good idea; it’s required by the FCRA. Section 607(b) of the FCRA, 15 U.S.C. § 16181e(b), requires consumer reporting agencies (CRAs) to follow “reasonable procedures to ensure maximum possible accuracy.” As discussed in Section C, our 2014 Big Data report revealed serious levels of inaccuracy in certain Big Data sources, and these purveyors may be in violation of Section 607(b).

The FCRA also has specific notice requirements, which are intended to ensure transparency when information from a CRA is used. Mostly importantly, Section 615(a) and (h) of the Act, 15 U.S.C. § 1681m(a) and (h), require users of consumer reports to provide adverse action and risk-based pricing notices when information from a CRA has been used to deny them credit or charge them a higher price. This ensures that consumers are aware of the sources and types of information that are used against them in credit (and other) decisions, so that they are not left in the dark as to the reasons for credit decisions, which may have critical consequences for their lives.

Furthermore, even if third party information is somehow not considered a consumer report, the FCRA includes a little-known provision that requires transparency in its usage. Section 615(b), 15 U.S.C. § 1681m(b), requires that lenders provide a specific notice if information that fits the seven characteristics listed in the definition of “consumer report” is obtained from a person other than a CRA and used to deny credit or charge more for it. This notice must inform consumers of their right to make a written request for the reasons for the adverse action, and then must disclose the nature of such information. Section 615(b) should apply to alternative data used for credit decisionmaking even if it somehow escapes the definition of a consumer report.

2. Equal Credit Opportunity Act (ECOA)

If alternative data is used for credit, there will be implications under the ECOA. Like the FCRA, the ECOA is a statute with a broad scope. It prohibits discrimination “with respect to any aspect of a credit transaction” on the basis of, *inter alia*, race, color, religion, national origin, sex or marital status, or age. 15 U.S.C. § 1691(a). “Credit” is broadly defined, as is the concept of “creditor,” which is not limited to banks or traditional lenders. 15 U.S.C. § 1691a(d) and (e). Finally, the ECOA is not limited to consumer credit but applies to certain types of business credit as well.

Most importantly for our purposes, Regulation B, which implements the ECOA, expressly notes that “legislative history of the Act indicates that the Congress intended an ‘effects test’ concept ... be applicable to a creditor's determination of creditworthiness.” 12 C.F.R. § 1002.6(a). The effects test is another name for the disparate impact test, and the Official Staff Interpretations explain that the test:

may prohibit a creditor practice that is discriminatory in effect because it has a disproportionately negative impact on a prohibited basis, even though the creditor has no intent to discriminate and the practice appears neutral on its face, 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.

Official Interpretations of Reg. B, 12 C.F.R. pt. 1002, supp. I, § 1002.6(a)-2.⁴⁸

As discussed in both Section C above and in Section G below, much of the data used for credit decisionmaking will have a disparate impact on protected classes. Thus, lenders must be extremely cautious in using any alternative data given the effects test. This test has a three-step analysis, which teasing out the Official Staff Interpretation quoted above, consists of:

1. Does the practice have a disproportionately negative impact on a protected class even if appears neutral on its face?
2. If so, does the practice meet a legitimate business need?
3. Can the same need be reasonably achieved using a less discriminatory alternative?

Like the FCRA, the ECOA also has specific notice requirements. It requires creditors to notify consumers of the action on an application. 15 U.S.C. § 1691(d)(1). If the creditor takes an adverse action, it must provide either a statement of reasons for the action or written notification of the right to such a statement. 15 U.S.C. § 1691(d)(2). This notice must be specific, and must meet the requirements of Regulation B and its corresponding Official Staff Interpretations.⁴⁹

⁴⁸ See generally National Consumer Law Center, Credit Discrimination § 4.3 (6th ed. 2013), updated at www.nclc.org/library.

⁴⁹ Reg. B, 12 C.F.R. § 1002.9(b)(2); Official Interpretations of Reg. B, 12 C.F.R. pt. 1002, supp. I, § 1002.9(b)(2). See generally National Consumer Law Center, Credit Discrimination § 10.5.4.2 (6th ed. 2013), updated at www.nclc.org/library.

G. Alternative Data, Structural Racism, and Reaching for Equity

“If we can agree that racialized wealth disparities are, in fact, unjust, then simply leaving these disparities alone to perpetuate themselves indefinitely is not an option. This is true regardless of whether there are currently racist policies creating wealth disparities, or whether the disparities, because of how wealth functions in a capitalist economy, are perpetuating themselves.”

-- john a. powell, “Understanding Structural Racialization”⁵⁰

It is very likely that most alternative data sources will have a disparate impact. This is because our society itself is deeply imbued with racial disparities. As Congressman John Lewis has noted “The scars and stains of racism are still deeply embedded in the American society.”⁵¹ This extends to credit scoring itself, which has stunning racially disparate impacts.⁵²

As discussed more in one of our Policy Briefs,⁵³ the explanation for these disparities lies in the aftereffects of historical inequalities and the racial wealth gap. African American families own less than seven cents for every dollar in wealth owned by white families, while Latino households own less than eight cents for every dollar of white wealth. With fewer assets to draw on, people of color – and the friends and family to whom they might turn – are far less able to cushion the blows of financial catastrophes, such as job losses, income reductions, sickness, or unplanned expenses.

Despite the fact that credit scoring has a disparate impact, it is also predictive. That is why it is legal under the ECOA – there is a legitimate business need for credit scoring given its predictiveness.

Thus, any alternative data will need to pass this same test of legitimate business need, because any alternative data will almost certainly exhibit racial disparities. Credit scores are not unique in showing economic disparities by race. Inequality is ingrained in many, if not, most of the social and political institutions and practices in our society. This is due to the phenomenon of “structural racism,” the idea that:

⁵⁰ Clearinghouse Review - Journal of Poverty Law and Policy, at 149, September–October 2013.

⁵¹ Athena Jones, Selma 50 years later: John Lewis's memories of the march, CNN, March 6, 2015, *available at* <http://www.cnn.com/2015/03/06/politics/selma-50-years-john-lewis-bridge-anniversary/>

⁵² For the past two decades, numerous studies has found that African American and Latino communities have lower credit scores as a group than whites (and Asians, when the data is available). National Consumer Law Center, Past Imperfect: How Credit Scores and Other Analytics “Bake In” Past Discrimination and Perpetuate It, May 2016, *available at* http://www.nclc.org/images/pdf/credit_discrimination/Past_Imperfect050616.pdf.

⁵³ National Consumer Law Center, Past Imperfect: How Credit Scores and Other Analytics “Bake In” Past Discrimination and Perpetuate It, May 2016, *available at* http://www.nclc.org/images/pdf/credit_discrimination/Past_Imperfect050616.pdf.

public policies, institutional practices, cultural representations, and other norms work in various, often reinforcing ways to perpetuate racial group inequity. . . . Structural racism is not something that a few people or institutions choose to practice. Instead it has been a feature of the social, economic and political systems in which we all exist.⁵⁴

In other words, systems replicate themselves, and racism is a system that keeps perpetuating itself long after individual animus-driven bias has eroded over the decades. Structural racism impacts the very foundations and institutions of our society – education, income, employment, housing, and criminal justice. This is why all the indicators discussed above – education, employment, income, and assets – exhibit such racial disparities.

Thus, because any economically based data is likely to show racial disparities, the Regulation B test for disparate impact will require a “legitimate business need” to use it for credit decisionmaking purposes. At a minimum, this means the data must be accurate and predictive. But we believe it should be more than just predictive – it should be more predictive than traditional data or should provide more access to credit (not less) to credit invisible or credit impaired consumers, and such credit should be affordably priced.

Also, even if predictive, there is certain data that is troubling because it entrenches and perpetuates inequality in such an obvious and stark way. One obvious example is educational attainment level, which as discussed in Section C.1 is highly correlated with parents’ income and education. Another type of data would be the consumer’s geographic neighborhood, given the obvious racial and economic segregation in housing in this country. Using these factors would only worsen the already herculean barriers that disadvantaged children face in making economic progress. And beyond race, we know that income and wealth inequality are already increasingly stark; using factors like educational level or neighborhood would just make it worse.

What data would be free of disparate racial impact and would not increase economic inequality? That is the brass ring. What kind of data would be future-looking data and not reproduce the inequality of the past? Bank account data could be promising. It is possible that alternative data and Big Data – used very carefully and judiciously – could have a role in this.

We urge significant research by data scientists, whether from the CFPB or industry or academia, to develop new tools or to adjust the current tools to account for discrimination. We also urge research as to whether the data models, whether traditional or alternative, could be programmed to reduce racial and economic disparities while maintaining – and hopefully, improving -- predictiveness. Models could be refined and modified so as to reduce racial disparities, to create the third prong of the disparate impact test - “a less discriminatory alternative.”

⁵⁴ The Aspen Institute, Glossary for Understanding the Dismantling Structural Racism/Promoting Racial Equity Analysis, June 2004, *available at* <https://assets.aspeninstitute.org/content/uploads/files/content/docs/rcc/RCC-Structural-Racism-Glossary.pdf>.

Such modifications to create a less discriminatory alternative might need to take race actively into account.⁵⁵ But that is not a radical concept. As the late Supreme Court Justice Harry Blackmun once noted, “In order to get beyond racism, we must first take account of race. There is no other way.”⁵⁶ To address invisibility and disparity, we must look *at* race, not away from it.

**never
trust anyone
who says
they do not see color.
this means
to them,
you are invisible.**

— from salt. by Nayyirah Waheed

* * *

Thank you for the opportunity to submit these comments and for your work to explore both the benefits and the risks of alternative data and modeling. If you have questions about these comments, please contact Chi Chi Wu at [ccwu@nclc.org](mailto:cwu@nclc.org) or 617-542-8010.

Respectfully submitted,

National Consumers Law Center
(on behalf of its low income clients)
California Reinvestment Coalition
Consumer Action
Consumers Union
National Association of Consumer Advocates
U.S. PIRG
Woodstock Institute
World Privacy Forum

⁵⁵ For example, one modification proposed by researchers would require including minority status as a “control variable” during the development of a credit scoring model. Michael LaCour-Little and Elaine Fortowsky, *Credit Scoring and the Fair Lending Issue of Disparate Impact* in *Credit Scoring for Risk Managers: The Handbook for Lenders* (Elizabeth Mays ed. South-Western Educational Pub. 2003); Elaine Fortowsky and Michael LaCour-Little, *Credit Scoring and Disparate Impact* (Dec. 2001), available at <https://www.phil.frb.org/-/media/consumer-credit-and-payments/payment-cards-center/events/conferences/2002/lacourpaper.pdf>.

⁵⁶ *University of California Regents v. Bakke*, 438 U.S. 265, 98 S. Ct. 2733 (1978).