

BUREAU OF CONSUMER FINANCIAL PROTECTION

[Docket No. CFPB-2017-0005]

Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process

AGENCY: Bureau of Consumer Financial Protection.

ACTION: Notice and request for information.

SUMMARY: The Consumer Financial Protection Bureau (CFPB or Bureau) seeks information about the use or potential use of alternative data and modeling techniques in the credit process. Alternative data and modeling techniques are changing the way that some financial service providers conduct business. These changes hold the promise of potentially significant benefits for some consumers but also present certain potentially significant risks. The Bureau seeks to learn more about current and future market developments, including existing and emerging consumer benefits and risks, and how these developments could alter the marketplace and the consumer experience. The Bureau also seeks to learn how market participants are or could be mitigating certain risks to consumers, and about consumer preferences, views, and concerns.

DATES: Comments must be received on or before May 19, 2017.

ADDRESSES: You may submit responsive information and other comments, identified by Docket No. CFPB-2017-0005, by any of the following methods:

- *Electronic:* Go to <http://www.regulations.gov>. Follow the instructions for submitting comments.

- *Mail:* Monica Jackson, Office of the Executive Secretary, Consumer Financial Protection Bureau, 1700 G Street, NW., Washington, DC 20552.
- *Hand Delivery/Courier:* Monica Jackson, Office of the Executive Secretary, Consumer Financial Protection Bureau, 1275 First Street, NE., Washington, DC 20002.

Instructions: Please note the number associated with any question to which you are responding at the top of each response (you are not required to answer all questions to receive consideration of your comments). The Bureau encourages the early submission of comments. All submissions must include the document title and docket number. Because paper mail in the Washington, DC area and at the Bureau is subject to delay, commenters are encouraged to submit comments electronically. In general, all comments received will be posted without change to <http://www.regulations.gov>. In addition, comments will be available for public inspection and copying at 1275 First Street, NE., Washington, DC 20002, on official business days between the hours of 10 a.m. and 5 p.m. Eastern Standard Time. You can make an appointment to inspect the documents by telephoning 202-435-7275.

All submissions, including attachments and other supporting materials, will become part of the public record and subject to public disclosure. Sensitive personal information, such as account numbers or Social Security numbers, or names of other individuals, should not be included. Submissions will not be edited to remove any identifying or contact information.

FOR FURTHER INFORMATION CONTACT: For general inquiries, submission process questions or any additional information, please contact Monica Jackson, Office of the Executive Secretary, at 202-435-7275.

AUTHORITY: 12 U.S.C. 5511(c).

SUPPLEMENTARY INFORMATION: The Bureau would like to encourage responsible innovations that could be implemented in a consumer-friendly way to help serve populations currently underserved by the mainstream credit system. To that end, in reviewing the comments to this request for information (RFI), the Bureau seeks not only to understand the benefits and risks stemming from use of alternative data and modeling techniques but also to begin to consider future activity to encourage their responsible use and lower unnecessary barriers, including any unnecessary regulatory burden or uncertainty that impedes such use.

The Bureau encourages comments from all interested members of the public. The Bureau anticipates that the responding public may encompass the following groups, some of which may overlap in part:

- Individual consumers;
- Consumer, civil rights, and privacy advocates;
- Community development and service organizations;
- Lenders, including depository and non-depository institutions;
- Consumer reporting agencies, including specialty consumer reporting agencies;
- Data brokers and aggregators;
- Model developers and licensors, as well as companies involved in the analysis of new or existing models;
- Consultants, attorneys, or other professionals who advise market participants on these issues;
- Regulators;
- Researchers or members of academia;

- Telecommunication, utility, and other non-financial companies that rely on consumer data for eligibility decisions;
- Participants in non-U.S. consumer markets with knowledge of or experience in the use of alternative data or modeling techniques for use in the credit process; and
- Any other interested parties.

All commenters are welcome to respond in any manner they see fit, including by sharing their knowledge of standard practices, their understanding of the market as a whole, or their own positions and views on the questions included in this RFI. Commenters may also choose to answer only a subset of questions. The information obtained in response to this RFI will help the Bureau monitor consumer credit markets and consider any appropriate steps. Comments may also help industry develop best practices. The Bureau seeks information predominantly pertaining to products and services offered to consumers. However, because some of the Bureau's authorities relate to small business lending,¹ the Bureau welcomes information about alternative data and modeling techniques in business lending markets as well. Information submitted by financial institutions should not include any personal information relating to any customer, such as name, Social Security number, address, telephone number, or account number.

For the purposes of this RFI, we define the following terms. None of these definitions should be construed as statutory or regulatory definitions or descriptions of statutory or regulatory coverage.

- “Traditional data” refers to data assembled and managed in the core credit files of the nationwide consumer reporting agencies, which includes tradeline information

¹ For example, the Equal Credit Opportunity Act covers both consumer and commercial credit transactions. 15 U.S.C. 1691 et seq. In addition, section 1071 of the Dodd-Frank Act requires data collection and reporting for lending to women-owned, minority-owned, and small businesses. The Bureau has yet to write regulations implementing that section but it has begun that process.

(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.

- “Alternative data” refers to any data that are not “traditional.” We use “alternative” in a descriptive rather than normative sense and recognize there may not be an easily definable line between traditional and alternative data.
- “Traditional modeling techniques” refers to statistical and mathematical techniques, including models, algorithms, and their outputs, that are traditionally used in automated credit processes, especially linear and logistic regression methods.
- “Alternative modeling techniques” refers to all other modeling techniques that are not “traditional,” including but not limited to decision trees, random forests, artificial neural networks, k-nearest neighbor, genetic programming, “boosting” algorithms, etc. We use “alternative” in a descriptive rather than normative sense and recognize that there may not be an easily definable line between traditional and alternative modeling techniques.
- “The credit process” refers to all the processes and decisions made by the creditor during the full lifecycle of the credit product, including marketing, pre-screening, fraud prevention, application procedures, underwriting, account management, credit authorization, the setting of pricing and terms, as well as the renewal, modification, or refinancing of existing credit, and the servicing and collection of debts.

PART A: TRADITIONAL AUTOMATED CREDIT PROCESS AND ITS ALTERNATIVES

Most of today's automated decisions in the credit process use traditional modeling techniques that rely upon traditional data elements as inputs. When lenders make decisions about consumers relating to applications for credit, increases or reductions in credit lines, extensions of new offers of credit, or other decisions in the credit process, lenders typically evaluate consumers using a standard set of information that includes consumer-supplied data (such as income, assets and, if secured, any collateral) and other traditional data supplied by one or more of the nationwide consumer reporting agencies. Many lenders base their decisions, in whole or in part, on scores using traditional data as inputs and generated from commercially-available, third-party models such as one of the many developed by FICO or VantageScore Solutions. Other lenders may base their decisions, in whole or in part, on proprietary scoring algorithms that use traditional data, and perhaps scores from these third-party models, as well as consumer-supplied information, as inputs. In addition to using common inputs, there is similar consistency in the modeling techniques used to generate these automated decision engines. They have predominantly been developed using multivariate regression analysis to correlate past credit history and current credit usage attributes to consumer credit outcomes to determine whether, based on the performance of other previous consumers who had similar attributes at the time credit was extended, it is likely that the consumer being evaluated will default on or become seriously delinquent on the loan within a certain period of time (often 1-2 years). These traditional data and modeling techniques have facilitated the standardization and automation of the credit process, leading to efficiencies in the provision of credit over the past few decades.

Yet the use of traditional data and modeling techniques has left some important gaps in access to mainstream credit for certain consumer groups and segments. The Bureau estimates that 26 million Americans are “credit invisible,” meaning that they have no file with the major

credit bureaus, while another 19 million are “unscorable” because their credit file is either too thin or too stale to generate a reliable score from one of the major credit scoring firms.² Most of these 45 million Americans are underserved by the mainstream credit system and they are disproportionately Black and Hispanic, low-income, or young adults. Some populations, like those recently widowed or divorced or recent immigrants, have difficulty accessing the mainstream credit system because they have not established a long enough credit history on their own or in this country. Some underserved consumers instead resort to high-cost products that may not help them build credit history.

Several commentators have suggested that alternative data and modeling techniques could address this problem and reach some of the millions of consumers currently shut out of the mainstream credit system and enable others to obtain more favorable pricing based on more refined assessments of their risks.³ Discussions point to the wide array of other data sources beyond traditional credit files that could be used to assess the creditworthiness of borrowers, including so-called “big data.”⁴ In addition, increased computing power and the expanded use of machine learning to mine massive datasets could potentially identify insights not otherwise discoverable through traditional methods. The application of alternative data and modeling techniques might also improve decisions in the credit process by improving the predictiveness of

² CFPB, *Data Point: Credit Invisibles* (May 2015), available at http://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf (figures are from 2010 Census).

³ See, e.g., PERC, *Give Credit Where Credit Is Due: Increasing Access To Affordable Mainstream Credit Using Alternative Data* (Dec. 2006), available at <http://www.perc.net/publications/give-credit-where-credit-is-due/>; CFSI, *The Predictive Value of Alternative Credit Scores* (Nov. 2007), available at <http://www.cfsinnovation.com/Document-Library/The-Predictive-Value-of-Alternative-Credit-Scores>;

⁴ “Big data” is a distinct concept from alternative data, though some alternative data may have the attributes generally ascribed to “big data.” In the FTC’s words, “A common framework for characterizing big data relies on the ‘three Vs,’ the volume, velocity, and variety of data, each of which is growing at a rapid rate as technological advances permit the analysis and use of this data in ways that were not possible previously.” FTC, *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues* (Jan. 2016), available at <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

credit-related models, by lowering the costs of sourcing and analyzing data, or through other process improvements such as faster decisions.

If these claimed benefits prove valid, the use of alternative data and modeling techniques could significantly reshape the consumer (and business) credit market. Potentially millions of consumers previously locked out of mainstream credit could become eligible for credit products that might help them buy a car or a home. An increasing ability for lenders to accurately assess risk could reduce the price of credit for those who are shown to be good risks (although it could *increase* the price of credit for those shown to be worse risks), and might even reduce the overall average price of credit for those who qualify for credit. The process of applying for credit could become more streamlined and convenient.

At the same time, other commentators have pointed out that alternative data and modeling techniques could present risks for consumers. These risks include but are not limited to potential issues with the accuracy of alternative data and modeling techniques; the lack of transparency, control, and ability to correct data that might result from their use; potential infringements on consumer privacy; and the risk that certain data could dampen social mobility, result in discriminatory outcomes, or otherwise disadvantage certain groups, characteristics, or behaviors.

The Bureau seeks to learn more about these potential benefits and risks. In further educating ourselves and the public, the Bureau seeks to encourage responsible uses of alternative data and modeling techniques while mitigating the various risks.

PART B: ALTERNATIVE DATA AND MODELING TECHNIQUES

Based on its research to date, the Bureau is aware of a broad range of alternative data and modeling techniques that firms are either using or contemplating. These innovations may be in different stages of development and market adoption. As set forth below, the Bureau seeks more information about the stages of development and extent of adoption of these innovations. In

some cases they are broadly used by a wide range of market participants, while others are in earlier stages of development. Some may be used often in fraud detection or marketing, for example, but rarely in underwriting. Some have been developed by established data aggregators or model developers who license their technologies or “platforms” to lenders; others have been developed for proprietary use by established lenders; and still others are being used by early stage lenders as a basis for lending at lower cost or profitably in certain channels or to consumer segments that established lenders have not traditionally served or can only serve at higher cost. Among the numerous online or marketplace lenders that have formed over the past few years, many have identified use of proprietary alternative data or machine learning techniques as central to their business strategies and comparative advantage.

Just how “alternative” or “traditional” certain data or modeling techniques are depends on one’s perspective. Labeling data or modeling techniques as “alternative” is not intended as a normative judgment, but to describe the fact that they have not customarily been used in decisions in the credit process. Any mention in this document of particular types of alternative data or modeling techniques should not be construed as endorsement or disapproval by the Bureau.

Data that some have labeled “alternative” include but are not limited to the following:⁵

- Data showing trends or patterns in traditional loan repayment data.
- Payment data relating to non-loan products requiring regular (typically monthly) payments, such as telecommunications, rent, insurance, or utilities.

⁵ This list is purely descriptive, and nothing should be implied from the inclusion or exclusion of any data.

- Checking account transaction and cashflow data and information about a consumer’s assets, which could include the regularity of a consumer’s cash inflows and outflows, or information about prior income or expense shocks.
- Data that some consider to be related to a consumer’s stability, which might include information about the frequency of changes in residences, employment, phone numbers or e-mail addresses.
- Data about a consumer’s educational or occupational attainment, including information about schools attended, degrees obtained, and job positions held.
- 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.
- Data about consumers’ friends and associates, including data about connections on social media.

Modeling techniques that some have labeled “alternative” include but are not limited to the following:

- Decision trees (or sets of decision trees, such as “random forests”).
- Artificial neural networks.
- Genetic programming.
- “Boosting” algorithms.
- K-nearest neighbors.

Given the rapidly evolving credit market landscape, the Bureau is eager to learn more about types of alternative data and modeling techniques, including but not limited to those listed above, and their uses and impacts.

PART C: POTENTIAL BENEFITS AND RISKS ASSOCIATED WITH USE OF ALTERNATIVE DATA AND MODELING TECHNIQUES IN THE CREDIT PROCESS

Prior Research and Interest in Alternative Data and Modeling Techniques

The Bureau is aware that several market participants,⁶ consumer advocates,⁷ regulators, and other commentators have identified the use of alternative data and modeling techniques as a source of potential opportunities and risks. Without seeking to summarize the full range of prior work, we note here a few relevant recent publications by other Federal entities.⁸ In September 2014, the Federal Trade Commission (FTC) held a public workshop on the topic of “Big Data” and subsequently published a report in January 2016 entitled “Big Data: A Tool for Inclusion or Exclusion?”⁹ This report outlined potential consumer benefits and risks broadly, rather than those specific to credit decisions. The FTC found that big data “is helping target educational, credit, healthcare, and employment opportunities to low-income and underserved populations” but could also contain “potential inaccuracies and biases [that] might lead to detrimental effects, including discrimination, for low-income and underserved populations.”¹⁰

⁶ See, e.g., FICO, “Can Alternative Data Expand Credit Access?” (Dec. 2015), available at <http://subscribe.fico.com/can-alternative-data-expand-credit-access>; TransUnion, “The State of Alternative Data,” available at <https://www.transunion.com/resources/transunion/doc/insights/research-reports/research-report-state-of-alternative-data.pdf>.

⁷ See, e.g., National Consumer Law Center, *Big Data: A Big Disappointment for Scoring Consumer Creditworthiness* (Mar. 2014), available at <http://www.nclc.org/issues/big-data.html>; Leadership Conference on Civil and Human Rights, “Civil Rights Principles for the Era of Big Data,” February 27, 2014, available at <http://www.civilrights.org/press/2014/civil-rights-principles-big-data.html>.

⁸ State policymakers and law enforcement officials have also looked into the potential risks and opportunities of alternative data, particularly on data privacy issues. For example, in March 2015 the National Association of Attorneys General held a meeting to discuss “Big Data: Challenges and Opportunities,” available at <http://www.naag.org/naag/media/naag-news/untitled-resource1.php>. In addition, the Massachusetts Attorney General hosted a March 2016 forum on data privacy in partnership with the MIT Computer Science and Artificial Intelligence Lab.

⁹ FTC, *Big Data: A Tool for Inclusion or Exclusion?* (Jan. 2016), available at <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf>.

¹⁰ *Id.* at 1.

Similarly, the Department of the Treasury’s May 2016 report on marketplace lending referenced the use of alternative data in underwriting by marketplace lenders as an area of both promise and risk: “While data-driven algorithms may expedite credit assessments and reduce costs, they also carry the risk of disparate impact in credit outcomes and the potential for fair lending violations.”¹¹

The Obama Administration completed two reports on big data, each referencing both the promises and risks posed by alternative data in the credit process.¹² The latter report notes, among other things, the importance of mitigating “algorithmic discrimination,” designing the best algorithmic systems, and algorithmic auditing and testing.

Finally, the Office of the Comptroller of the Currency (OCC), the Federal Reserve Board of Governors (FRB), and the Federal Deposit Insurance Corporation (FDIC) recently issued joint guidance¹³ referencing alternative data. The guidance identifies that banks’ use of “alternative credit histories” as a means “to evaluate low- or moderate-income individuals who lack sufficient conventional credit histories and who would be denied credit based on the institution’s traditional underwriting standards” could be considered an “innovative and flexible practice . . . to address the credit needs of low- or moderate-income individuals or geographies” that examiners would consider in evaluating banks’ lending practices under the Community

¹¹ U.S. Treasury, *Opportunities and Challenges in Online Marketplace Lending* (May 2016), available at https://www.treasury.gov/connect/blog/Documents/Opportunities_and_Challenges_in_Online_Marketplace_Lending_white_paper.pdf.

¹² Executive Office of the President, *Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights* (May 2016), available at https://www.whitehouse.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf; Executive Office of the President, *Big Data: Seizing Opportunities, Preserving Values* (May 2014), available at https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.

¹³ OCC, FRB, and FDIC, *Community Reinvestment Act; Interagency Questions and Answers Regarding Community Reinvestment; Guidance*, 81 FR 48506 (July 25, 2016), available at <https://www.gpo.gov/fdsys/pkg/FR-2016-07-25/pdf/2016-16693.pdf>

Reinvestment Act (CRA). The guidance lists a prospective borrower's rental and utility payments as examples of alternative credit history.

These agencies' attention to the use of alternative data and modeling techniques in the credit process reflects the growing importance of these methods and approaches in the marketplace. As a Federal agency designated by Congress to oversee compliance with the various consumer financial protection statutes and regulations as they apply to *both* banks and non-banks, and with its additional desire to foster consumer-friendly innovation in the marketplace, the Bureau is especially interested in increasing its understanding of the consumer benefits and risks that are likely to accompany these developments and how they relate to established consumer protections. Through this RFI, the Bureau seeks to build on the foundation of existing research by other Federal agencies and develop a deeper understanding of these potential benefits and risks. The Bureau seeks to encourage responsible and consumer-friendly uses of alternative data and modeling techniques that leverage such benefits while providing a clearer path whereby market participants can mitigate risks to consumers.

Potential Consumer Benefits

Alternative data and modeling techniques have the potential to benefit consumers in several ways listed below. These benefits, as well as others not identified here, could accrue differently in different product markets – what helps consumers in the credit card marketplace may not help consumers in the mortgage marketplace – or could provide different levels of benefits to different consumer segments – what helps consumers with no credit records may not help consumers with long traditional credit histories.

- **Greater credit access:** The Bureau estimates that approximately 45 million Americans lack access to mainstream credit because they have no credit history or because their

credit history is insufficient or stale. The use of alternative data or modeling techniques could increase access to credit for that population by providing more information about them and enabling them to be reliably scored. For example, some consumers might not have traditional loan repayment history but might pay their mobile phone bills on a regular basis, a pattern that might be sufficient to reassure some lenders that they are viable credit risks. Of course, only some portion of that 45 million might be reliably scorable using alternative data and modeling techniques, and some of those scores might not qualify consumers for mainstream credit.

- **Enhanced creditworthiness predictions:** Alternative data and modeling techniques could allow lenders to better assess the creditworthiness of consumers who are already scored. For example, a lender might not currently lend below a credit score of 620, but might be willing to do so if, by adding some new data source, it could distinguish those sub-620 consumers who present greater or lesser risks of default. It is important to note that, to the extent alternative data or modeling techniques could help a creditor identify consumers who are *more and less* likely to default than their current credit score suggests, alternative data could in fact *decrease or increase* a given consumer's likelihood of receiving credit, or could *raise or lower* the price that any individual is offered for that credit. Though this could be seen as a detriment to consumers who are less likely to receive credit (or whose prices increase), it could also be seen as an improvement in risk assessment, which may provide greater certainty and allow a lender to increase credit availability for those who qualify. Indeed, in the longer term consumers whose credit scores understate their true risk may be better served if they do not obtain additional credit that they cannot repay.

- **More timely information:** The credit process could be improved by relying on more timely information about the consumer being assessed. While all risk assessments use data from the present or past to predict outcomes in the future (e.g., likelihood of default), traditional data often lags actual events. For example, the opening of a new credit account might take months to show up on a consumer’s credit report and in some cases it may not show up at all. Alternative data could provide more timely indicators, such as real-time access to a consumer’s outstanding credit card balance. It could also help lenders recognize whether a particular consumer’s finances are trending in a particular direction, such as through a job status change appearing on social media. Such information could help to distinguish those consumers whose low scores are a function of prior financial problems that they have surmounted from those consumers whose financial challenges have just begun and who may pose a greater risk than the score indicates. Alternative modeling techniques might also generate more timely feedback to the extent they dynamically change as new data are ingested, though such dynamism could also carry certain risks.
- **Lower costs:** The use of alternative data and modeling techniques may have the potential to lower lenders’ costs – these cost savings might, in turn, be passed along to consumers in the form of lower prices or in lenders’ ability to make smaller loans economically. For example, a lender might currently verify employment and income by calling the consumer’s employer or manually reviewing tax returns. If, instead, the lender could automate such tasks by processing data associated with the individual’s employer, tax returns, or other methods, its processing costs might significantly decline.

- **Better service and convenience:** Alternative data and modeling techniques might also be able to drive operational improvements that enable better customer service outcomes for consumers or greater convenience. For example, to the extent more tasks can be automated, it might speed up application processes or reduce any discretionary judgments that may sometimes lead to discrimination.

Through this RFI, the Bureau seeks to understand how consumers might benefit from the use of alternative data and modeling techniques (including in the ways identified above), the degree to which those benefits impact different consumer segments or products, and any specific empirical evidence relevant to the likelihood and extent of those benefits.

Potential Consumer Risks

Use of alternative data and modeling techniques also carries several potential risks. The Bureau lists some such risks below not to dissuade the use of alternative data and modeling techniques but rather to highlight some of the challenges with such use, to encourage responsible use that takes consideration of and manages these risks, and to invite commenters to discuss their views about how these and other risks could be mitigated. As with the consumer benefits, this list of consumer risks may not encompass all of the perceived or potential consumer risks, and some risks may apply differently to different consumer or product segments.

- **Privacy:** Some types of alternative data could raise privacy concerns because the data are of a sensitive nature and consumers may not know the data were collected and shared nor expect or be aware it will be used in decisions in the credit process.
- **Data quality issues:** Some types of alternative data could raise accuracy concerns because the data are inconsistent, incomplete, or otherwise inaccurate. Though

traditional data raises accuracy concerns,¹⁴ it could be that certain types of alternative data have greater rates of error due to their nature or the fact that the quality standards for their original purpose are lesser than those associated with decisions in the credit process. Such concerns may arise in part because such data have not historically been used in credit or other eligibility decisions and, as a result, the sources of such data may not have been subject to the type of accuracy and quality obligations that would commonly be expected for data to be used in decisions in the credit process.

- **Lost transparency, control, and ability to correct:** Some sources of alternative data may not permit consumers to access or view data that is being used in decisions in the credit process, or to correct any inaccuracies in that data. In some cases, consumers might not be able to determine the sources of the data. These issues are compounded if creditors are not transparent about the type of data they are using and how those data figure into decisions in the credit process. Certain alternative modeling techniques could compound the transparency problem if they do not permit easy interpretation of how various data inputs impact a model's result.
- **Harder to change credit standing through behavior:** Traditional credit factors are heavily influenced by the consumer's own financial conduct, such as whether the person paid their loans on time or how much credit the person has obtained and utilized. Alternative data that cannot be changed by consumers or that are not specific to the

¹⁴ See FTC, *Report to Congress Under Section 319 of the Fair and Accurate Credit Transactions Act of 2003* (Jan. 2015), available at <https://www.ftc.gov/system/files/documents/reports/section-319-fair-accurate-credit-transactions-act-2003-sixth-interim-final-report-federal-trade/150121factareport.pdf> (26% of consumers found material errors on their credit reports, 13% experienced a change in their credit score as a result of modifying their reports, and 5% experienced a significant change that changed their risk tier).

individual, but relate instead to peers or broader consumer segments, do not enable consumers to improve their credit rating.

- **Harder to educate and explain:** The more factors that are integrated into a consumer's credit score or into decisions in the credit process, or the more complex the modeling process in which the data are used, the harder it may be to explain to a consumer what factors led to a particular decision. This may be true for lenders, who are required to provide adverse action notices to consumers in certain circumstances, as well as for financial educators, who wish to improve consumers' understanding of the factors that impact their credit standing. These complexities make it more difficult for consumers to exercise control in their financial lives, such as by learning how to improve their credit rating.
- **Unintended or undesirable side effects:** The use of alternative data and modeling techniques could penalize or reward certain groups or behaviors in ways that are difficult to predict. For example, members of the military may frequently move and the perceived lack of housing stability or continuity may give a false impression of overall instability. Or negative inferences could potentially be drawn about consumers who are not found in the alternative data source being used by the lender. Foreseeable or otherwise, using alternative data and modeling techniques could also cause potentially undesirable results. For example, using some alternative data, especially data about a trait or attribute that is beyond a consumer's control to change, even if not illegal to use, could harden barriers to economic and social mobility, particularly for those currently out of the financial mainstream.

- **Discrimination:** Alternative data and modeling techniques could also result in illegal discrimination. For example, using alternative data that involves categories protected under Federal, State, or local fair lending laws may be overt discrimination. In addition, certain alternative data variables might serve as proxies for certain groups protected by anti-discrimination laws, such as a variable indicating subscription to a magazine exclusively devoted to coverage of women’s health issues. And the use of other alternative data might cause a disproportionately negative impact on a prohibited basis that does not meet a legitimate business need or that could be reasonably achieved by means that are less disparate in their impact. Machine learning algorithms that sift through vast amounts of data could unearth variables, or clusters of variables, that predict the consumer’s likelihood of default (or other relevant outcome) but are also highly correlated with race, ethnicity, sex, or some other basis protected by law. Such correlations are not per se discriminatory but may raise fair lending risks. The use of alternative data and modeling techniques could potentially lead to disparate impact on the part of a well-intentioned lender as well as allow ill-meaning lenders to intentionally discriminate and hide it behind a curtain of programming code.
- **Other violations of law:** The use of alternative data and modeling techniques could potentially raise the risk of violating consumer financial laws, such as the Equal Credit Opportunity Act (ECOA) and Regulation B, the Fair Credit Reporting Act (FCRA) and Regulation V, and the prohibitions on unfair, deceptive, or abusive acts or practices (UDAAPs, collectively). The Bureau also recognizes that there may be uncertainty about how certain aspects of these laws apply to alternative data and modeling techniques, and the Bureau seeks to understand specifically where greater certainty would be helpful.

Through this RFI, the Bureau seeks to understand risks to consumers from the use of alternative data and modeling techniques (including in the ways identified above), the degree to which those risks impact different product or consumer segments, and any specific empirical evidence relevant to the likelihood and extent of those risks. The Bureau also seeks to understand what steps market participants are taking to manage risks and realize benefits. The Bureau intends to use information gleaned from the questions below to help maximize the benefits and minimize the risks from these developments.

PART D: QUESTIONS RELATED TO ALTERNATIVE DATA AND MODELING TECHNIQUES USED IN THE CREDIT PROCESS

This RFI is intended to cover past, current, and potential uses of alternative data and modeling techniques. The Bureau is interested in learning more about the specific types of alternative data and modeling techniques utilized for various decisions in the credit process, as well as the policies and procedures used to ensure the responsible use of these alternative data and methods. In addition, the Bureau seeks to learn how the use of alternative data and modeling techniques compares and contrasts with the use of traditional data and modeling techniques for those same decisions. Finally, of particular interest is a specific and empirical understanding of the current and potential consumer benefits and risks associated with the use of alternative data and modeling techniques, including risks related to specific statutes and regulations.

While the Bureau recognizes that some commenters may feel that answering the questions below raises concerns about revealing proprietary information, we encourage commenters to share as much detail as possible in this public forum.¹⁵ We also welcome comments from

¹⁵ We do not seek, nor should commenters provide, actual alternative data about consumers. Rather we seek information about different types of alternative data.

representatives, such as attorneys, consultants, or trade associations, which need not identify their clients or members by name.

The questions below are divided into four sections: (1) Alternative Data; (2) Alternative Modeling Techniques; (3) Potential Benefits and Risks to Consumers and Market Participants; and (4) Specific Statutes and Regulations. Each question speaks generally about all decisions in the credit process, but answers can differentiate, as appropriate, between uses in marketing, fraud detection and prevention, underwriting, setting or changes in terms (including pricing), servicing, collections, or other relevant aspects of the credit process. The questions are phrased in the present tense, but the Bureau is equally interested in information about any past but discontinued uses or in any potential future uses that commenters are considering or are aware of. The Bureau welcomes any relevant empirical research or studies on these topics.

Alternative Data

This section asks questions about the types, sources, and purposes of alternative data. Comments referencing specific practices, firms, or data are especially helpful.

1. What types of alternative data are used in decisions in the credit process? Please describe not only the broad categories (e.g., cashflow data) but also the specific data element or variables used (e.g., rent or telephone expense). The questions below refer back to each type of alternative data listed in response to this question.
2. For each type of alternative data identified above:
 - a. Please describe the specific decisions in which this type of alternative data is used, the specific purpose for using it, and the product(s) and consumer segment(s) for which it is used. For example, are certain data used to create a proprietary score for underwriting mortgage loans for non-prime applicants while

- other data are used to determine whether credit line increases or decreases are appropriate for existing credit card users?
- b. Please describe any goals, objectives, or challenges that the use of this type of alternative data is designed to accomplish or address. For example, a certain type of data might be used in order to provide a more timely assessment of the consumer's current income while another type of data might be used to more accurately predict the stability of future income streams. Please describe the extent to which use of alternative data has in fact advanced or addressed these goals, objectives, or challenges.
- c. Please describe the source of the data, being as specific as possible, including if the data are provided by the consumer or obtained from or through a third party. If obtained from a third party, please indicate if that third party considers itself to be a consumer reporting agency subject to the FCRA.
- d. Please describe the format in which the data are received or generated, being as specific as possible.
- e. Please describe the breadth or coverage of the data. Are there certain consumer segments for whom the data are unavailable?
- f. Please describe whether the data include both positive and negative observations. For example, do records of rental payments include instances where consumers paid on time as well as when they were late?
- g. Please describe if the data are specific to the individual consumer (e.g., the consumer's actual income) or attributed to the consumer based upon a perceived

peer group (e.g., average income of consumers obtaining the same educational degree).

- h. Please describe the quality of the data, in terms of apparent errors, missing information, and consistency over time.
 - i. Please describe the methods or procedures used to assess the coverage, quality, completeness, consistency, accuracy, and reliability of the data, as well as who is responsible for overseeing those methods or procedures.
 - j. Please describe the original purpose for which the data were initially generated, assembled, or collected, and the standard for coverage, quality, completeness, consistency, accuracy, and reliability that the original data provider applied. Was the consumer able to see, dispute, or correct the data at the time they were originally collected or with the original collector of the data or with the subsequent user?
 - k. Could this particular type of alternative data feasibly be furnished to one or more of the nationwide consumer reporting agencies? What would be the investment(s) required to do so? What prevents such furnishing today?
 - l. Please describe whether and how the data are used in identifying and constructing target lists for marketing credit online, by mail, or in person (i.e., firm offers of credit or invitations to apply).
 - m. Please describe whether and how the data are used to screen for potential fraud prior to assessing creditworthiness.
3. For each type of alternative data identified above, please describe the process for deciding whether to use that type of data, including the criteria used for evaluating the

data and its potential use. If applicable, please describe the basis for determining the relationship between the data and the outcome they are designed to predict. If the relationship is empirically derived, describe the type(s) of data used to derive the relationship (e.g., internal loan performance data, third-party reject inference data, etc.).

4. For each type of alternative data identified above, please describe whether the data are used alongside other traditional or alternative data. How much impact does the alternative data have on the relevant decision? Is this data used only after a preliminary decision based on the exclusive use of traditional data, for example, to re-evaluate consumers who failed a model that used only traditional data? Or is it used at the same time? Are there particular decisions or particular products or consumer segments where firms rely exclusively or predominantly on the use of alternative data?
5. Are there types of alternative data that have been evaluated but are not being used in decisions in the credit process? If so, please describe and explain the evaluation process and outcomes and the reason(s) why the alternative data are not being used for the particular credit-related decision.
6. For questions 1 through 5 above, please describe any differences in your answers as they pertain to lending to businesses (especially small businesses) rather than consumers.

Alternative Modeling Techniques

This section asks questions about alternative modeling techniques. Comments referencing specific practices, firms, or data are especially helpful.

What types of alternative modeling techniques are used in decisions in the credit process? Please describe these modeling techniques in as much detail as possible, including but not limited to:

- a. A detailed explanation of the modeling technique, and how it transforms inputs into outputs.
 - b. The product or consumer segment(s) it is used for.
 - c. The outcome(s) the modeling technique aims to predict.
 - d. The final output that the modeling technique generates, such as a score within a defined range or a pass/fail decision, including any identification of the main factors impacting the final output.
 - e. A detailed explanation of the specific data types used as inputs, including both traditional and alternative data.
 - f. Whether the modeling technique is used concurrently with, subsequent to, or in conjunction with other traditional or alternative modeling techniques. How much impact does the alternative modeling technique have on the decision it informs?
7. For each type of alternative modeling technique identified above, please describe the model development and governance process (e.g., initial development, training, testing, validation, beta, broader use, redevelopment, etc.) in as much detail as possible, including but not limited to:
- a. Whether the process differs based upon the type of outcome being predicted.
 - b. Whether the process differs for alternative versus traditional modeling techniques.
 - c. Whether the process differs when alternative versus traditional data are used.
 - d. Whether specific tests or validations are performed to assess compliance with fair lending or other regulatory requirements. Are these similar to or different from those used for traditional modeling techniques?

- e. A description of any judgmental, subjective, or discretionary decisions made in the development phase. For example, for machine learning techniques, what are decisions the developer must make in supervising the training phase, or providing parameters or limits on its operation?
 - f. A description of how, if at all, the process handles:
 - i. Sample selection for model testing/validation.
 - ii. Potential measurement error.
 - iii. Overfitting.
 - iv. Correlations with characteristics prohibited under fair lending laws.
 - v. Direction of the relationship between features and outcomes (e.g., monotonicity).
 - vi. Any other noteworthy considerations.
8. For questions 7 and 8 above, please describe any differences in your answers as they pertain to lending to businesses (especially small businesses) rather than consumers.

Potential Benefits and Risks to Consumers and Market Participants

This section asks questions about the potential benefits and risks related to the use of alternative data and modeling techniques. The Bureau encourages commenters to be as specific as possible when describing the potential benefits and risks, including but not limited to which consumer segments or groups (e.g., no traditional credit file, different demographic groups), which products (e.g., auto loans, credit cards), and which channels (e.g., online, storefront) are most affected.

9. What does available evidence suggest about the potential benefits for consumers of using alternative data present to:

- a. Improved risk assessment so that consumers are more accurately paired with appropriate credit products.
 - b. Increases in access to affordable credit.
 - c. Lower prices.
 - d. Quicker or more convenient decisioning process.
10. What does available evidence suggest about the potential benefits for consumers of using alternative modeling techniques? Such benefits could include, but are not limited to:
- a. Improved risk assessment so that consumers are more accurately paired with appropriate credit products.
 - b. Increases in access to credit.
 - c. Lower prices.
 - d. Quicker or more convenient decisioning process.
11. What does available evidence suggest about the potential benefits for market participants of using alternative data? Such benefits could include, but are not limited to:
- a. An increased ability to accurately predict the likelihood of a certain outcome (e.g., a 90 day delinquency within 24 months).
 - b. Risk assessment that is more reactive to real-time information.
 - c. Ability to assess and grant credit to more consumers.
 - d. Lower operational costs.
 - e. Quicker or more convenient decisioning process.
 - f. Competitive advantage, including the ability to compete with traditional methods.

12. What does available evidence suggest about the potential benefits for market participants of using alternative modeling techniques? Such benefits could include, but are not limited to:

- a. An increased ability to accurately predict the likelihood of a certain outcome (e.g., a 90 day delinquency within 24 months).
- b. Risk assessment that is more reactive to real-time information.
- c. Ability to assess and grant credit to more consumers.
- d. Lower operational costs.
- e. Quicker or more convenient decisioning process.
- f. Competitive advantage, including the ability to compete with traditional methods.

13. What does available evidence suggest about the potential risks for consumers of using alternative data? In addition, what steps are being taken to mitigate these risks? Such risks could include, but are not limited to:

- a. Impacts on consumer privacy.
- b. Decreased transparency about the use of one's data and about how decisions in the credit process are made.
- c. Decreased ability to dispute inaccurate information or correct errors.
- d. Decreased ability of consumers to improve their credit standing.
- e. Decreased completeness, consistency, accuracy, or reliability of data that affects decisions in the credit process.
- f. Illegal discrimination.
- g. The hardening of barriers to social and economic mobility.
- h. Decreased access to affordable credit.

- i. Decreased ability to inform and educate consumers about the factors affecting their credit standing.
14. What does available evidence suggest about the potential risks for consumers of using alternative modeling techniques? In addition, what steps are being taken to mitigate these risks? Such risks could include, but are not limited to:
- a. Decreased transparency about the use of one's data and about how decisions in the credit process are made.
 - b. Decreased ability to dispute inaccurate information or correct errors.
 - c. Decreased ability of consumers to improve their credit standing.
 - d. Illegal discrimination.
 - e. Decreased ability to inform and educate consumers about the factors affecting their credit standing.
15. What does available evidence suggest about the potential risks for market participants of using alternative data? In addition, what specific steps are being taken to mitigate these risks? Such risks could include, but are not limited to:
- a. Decreased transparency about how decisions in the credit process are made.
 - b. Lack of historical performance data related to certain alternative data.
 - c. Decreased completeness, consistency, accuracy, or reliability of data.
 - d. Decreased ability to inform and educate consumers about the factors affecting their credit standing.
 - e. Decreased consumer trust or acceptance of lender decisions.

16. What does available evidence suggest about the potential risks for market participants of using alternative modeling techniques? In addition, what specific steps are being taken to mitigate these risks? Such risks could include, but are not limited to:

- a. Decreased transparency about how decisions in the credit process are made.
- b. Lack of historical performance data related to certain modeling techniques.
- c. Decreased ability to inform and educate consumers about the factors affecting their credit standing.
- d. Decreased consumer trust or acceptance of lender decisions.

17. For questions 10 through 17 above, please describe any differences in your answers as they pertain to lending to businesses (especially small businesses) rather than consumers.

Specific Statutes and Regulations

This section asks questions about specific statutes and regulations as they pertain to alternative data and modeling techniques. Nothing below should be interpreted as a legal conclusion or interpretation by the Bureau. While the questions below are focused on the activities of market participants, the Bureau is equally interested in information from researchers, consultants, and other third parties about the issues raised below. The Bureau also recognizes that market participants may be reluctant to comment publicly on potential legal uncertainties and invite such parties to submit comments through anonymized channels such as law firms, trade associations, and the like.

18. The ECOA and Regulation B prohibit discrimination on the basis of race, color, religion, national origin, sex, marital status, age, the fact that all or part of the applicant's income derives from any public assistance program, or the good faith exercise of any right under

the Consumer Credit Protection Act. Evidence of disparate treatment and evidence of disparate impact can be used to show discrimination under ECOA and Regulation B.

- a. Are there specific challenges or uncertainties that market participants face in complying with ECOA and Regulation B with respect to the use of alternative data or modeling techniques?
- b. In the absence of data on applicants' ethnicity, race, sex, or other prohibited basis group membership, how prevalent is the practice of proxying for those characteristics in order to test for potential fair lending risks in the use of alternative data or modeling techniques?
- c. How, if at all, are market participants using demographically conscious model development techniques to ensure that models or modeling techniques do not result in illegal discrimination?
- d. For respondents (such as market participants or consultants, attorneys, or other professionals who advise market participants) that evaluate models for potential fair lending risk, please answer the following questions. For each activity described in your answers, please specify the point(s) in time (e.g., model development, validation, implementation, or use) at which the activity is conducted; the function(s) within the company responsible for conducting the activity; the type(s) of models reviewed (e.g., underwriting, pricing, fraud, marketing); how those models are prioritized for review; the level (e.g., attribute, model, or decisioning process) at which the activity is conducted; and which prohibited bases (e.g., age, sex, race, ethnicity) are evaluated.

- i. In general, what methods do market participants use to evaluate alternative data and modeling techniques for fair lending risk?
- ii. What steps, if any, do market participants take to determine whether alternative data may be serving as a proxy for a prohibited basis? What thresholds, standards, or baselines are used to make this determination?
- iii. What steps, if any, do market participants take to determine whether use of alternative data has a disproportionately negative impact on a prohibited basis? What thresholds, standards, or baselines are used to make this determination? To what extent, if any, do market participants use traditional data (or scores generated therefrom) as a baseline for making this determination?
- iv. What steps, if any, do market participants take to determine if the use of alternative data meets a legitimate business need notwithstanding any disproportionately negative impact that use may have on a prohibited basis?
- v. What steps, if any, do market participants take to ensure that a legitimate business need met by the use of alternative data cannot reasonably be achieved as well by means that are less disparate in their impact?
- vi. What other steps, besides those already discussed in response to questions 19(d)(i)-(v) above, do market participants take to evaluate or manage potential fair lending risk arising from the use of alternative data or modeling techniques?

- vii. When a lender identifies disparities affecting a prohibited basis group or other fair lending risks that arise from the use of a particular variable or model, what steps does the lender take as a result? To what extent do these steps mitigate that risk?
- viii. How do the activities described in response to questions 19(d)(i)-(v) compare with the activities conducted when using traditional data or modeling techniques?
- e. Many entities subject to the Bureau’s supervisory or enforcement jurisdiction have risk management programs in place pursuant to guidance on model risk management issued by prudential regulators.¹⁶ To what extent do market participants use principles or processes discussed in that guidance in connection with their management of fair lending risk?
- f. Are market participants using alternative data or modeling techniques as a “second look” for those who do not meet initial eligibility requirements based on traditional data or modeling techniques? If so, what issues and challenges, if any, arise in that context? Have data that were first used in “second looks” eventually become included in initial screening processes?
- g. When using alternative data or modeling techniques, or using multiple models, are there challenges in determining and disclosing to applicants the principal reasons for taking adverse action or describing the reasons for taking adverse

¹⁶ See Federal Reserve Board SR Letter 11-7 (“Guidance on Model Risk Management”) (April 4, 2011); Office of the Comptroller of the Currency (OCC) Bulletin 1997-24 (“Credit Scoring Models”) (May 20, 1997); OCC Bulletin 2000-16 (“Risk Modeling”) (May 30, 2000); OCC Bulletin 2011-12 (“Sound Practices for Model Risk Management”) (April 4, 2011); Federal Deposit Insurance Corporation (FDIC) Supervisory Insights (“Model Governance”) (last updated December 5, 2005); FDIC Supervisory Insights (“Fair Lending Implications of Credit Scoring Systems”) (last updated April 11, 2013).

action in a manner that relates to and accurately describes the factors actually considered or scored?

19. The FCRA and Regulation V regulate the collection, dissemination, and use of consumer information, including consumer credit information.

- a. Are there specific challenges or uncertainties that market participants face in complying with the FCRA with respect to the use of alternative data or modeling techniques?
- b. What challenges do companies generating, selling, and brokering alternative data face in determining whether they are a consumer reporting agency subject to the FCRA?
- c. What challenges do consumer reporting agencies assembling or evaluating alternative data face in implementing accuracy and dispute procedures and disclosing file information to consumers?
- d. What challenges do lenders face when they obtain alternative data? Is it typically clear whether the data provider is a consumer reporting agency subject to the FCRA?
- e. How, if at all, do market participants treat alternative data differently when they receive it from data providers or other sources that do not appear to be subject to the FCRA?
- f. When using alternative data or modeling techniques, or using multiple credit scores, are there challenges in providing adverse action notices or risk-based pricing notices? For example, when using alternative modeling techniques, are there challenges in determining the key factors that adversely affected the

consumer's score? Are there challenges in providing the source of the information? Do you have information showing whether consumers understand the information on these notices or take appropriate follow-up actions?

- g. When using alternative data or modeling techniques, are there challenges in disclosing, pursuant to Section 615(b) of the FCRA, the nature of the information used in credit-related decisions when such information comes from a third party that is not a consumer reporting agency?
 - h. The FCRA permits consumer reports to be obtained for some non-credit decisions, such as employment and tenant screening. What potential impacts could alternative data and modeling techniques have on these non-credit decisions?
20. The Dodd-Frank Act prohibits unfair, deceptive, or abusive acts or practices in connection with consumer financial products or services. Section 5 of the FTC Act similarly prohibits unfair or deceptive acts or practices in connection with a broader set of transactions.
- a. Are there specific challenges or uncertainties that market participants face in complying with the prohibitions on UDAAPs with respect to alternative data or modeling techniques?
 - b. What steps, if any, do users of alternative data or modeling techniques take to avoid engaging in UDAAPs?
 - c. What steps, if any, can the Bureau take to help minimize the risk of UDAAPs from the use of alternative data and modeling techniques?

[THIS SIGNATURE PAGE PERTAINS TO THE NOTICE AND REQUEST FOR
INFORMATION TITLED “REQUEST FOR INFORMATION REGARDING USE OF
ALTERNATIVE DATA AND MODELING TECHNIQUES IN THE CREDIT
PROCESS”]

Dated: February 14, 2017.

Richard Cordray
Richard Cordray,
Director, Bureau of Consumer Financial Protection.