May 19, 2017

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

Re: Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process, Docket No. CFPB-2017-005

Thank you for the opportunity to provide comments in response to the Consumer Financial Protection Bureau’s (“CFPB” or “Bureau”) Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process.

Introduction

The Software & Information Industry Association (SIIA) is the principal trade association for the software and digital information industries. The more than 700 software companies, data and analytics firms, information service companies, and digital publishers that make up our membership serve nearly every segment of society, including business, education, government, healthcare and consumers. As leaders in the global market for software and information products and services, they are drivers of innovation and economic strength. SIIA members include companies that provide data and analytic services, including non-traditional data and modeling techniques, that are used by lenders in the credit process.

SIIA applauds the CFPB for seeking to explore the benefits associated with “alternative data,” and we strongly support the goal to encourage greater use of alternative data and modeling techniques while mitigating various risks. Given that alternative data represents an evolution and expansion of existing credit data, and the opportunities it presents to expand credit opportunities for so many Americans, SIIA often refers to this data as “next generation data.” These comments discuss several examples of non-traditional data and modeling techniques, and we highlight the substantial benefits of their use.

The Bureau is prudent to also consider the risks associated with the use of new data and techniques, but as we highlight below, the current legal regime is fully capable of controlling these risks, while allowing consumers to reap the benefits of this expanded use. SIIA urges the CFPB to encourage the expanded use of these data and techniques.

Please find below response to various questions raised in Part D, related to alternative data and modeling techniques used in the credit process.
Alternative Data

1) Types and Uses of Alternative Data

The national credit bureaus primarily report the current and historical status of consumer credit accounts, including the name of the lender, the type of debt, the current debt amount, the delinquency status, the available credit or original loan amount, and a summary of historical delinquency events. There are thousands of financial institutions that report credit tradeline data, and the process is highly automated using standard data reporting formats across all institutions.

Alternative data, when used in the context of these comments, can be defined as information not found on a traditional credit report that can reasonably and rationally be used for credit risk decisions because it is:

- Predictive information that bears on a consumer's credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living.
- Consumer information that is used, stored and disseminated in compliance with all relevant consumer protection laws and regulations, like the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA).

Alternative data includes the following types of information:

- **Public Records Data:** Data that is maintained by national, state and/or local governments, which could include property records, property tax values, evictions, and professional licenses.
- **Private Source Information:** Information compiled from sources not in the public domain. Examples include data regarding telecom and utility payments, prepaid cards, employment information, bounced check information, and deposit account information.

2) Application, Uses, Objectives and Availability of Alternative Data

Alternative data can be used for a variety of applications. It is currently used by different industries in the U.S. in different contexts, including credit card issuance, utilities, telecommunications, auto finance, demand deposit / account opening, small business financing, financial technology, and short-term consumer lending. Nearly all of the top bankcard issuers, auto finance companies, and wireless providers use alternative data in their account evaluation processes. Lenders utilizing alternative data typically have two main goals in mind:

1. Providing a way to evaluate potential customers who do not have a traditional credit file.
2. Supplementing traditional credit information to provide a more complete picture of consumers who have a thin credit file.
The specific data elements utilized is driven by the specific business objective. Scenarios include:

- A credit card company viewing a consumer’s asset ownership history to understand more about that consumer’s overall financial position.
- An auto finance company evaluating the way frequent address changes or evidence of a lien or judgment impacts the performance of a consumer’s loan portfolio.
- A credit union reviewing an applicant’s history with short-term loans to offer a refinancing product which shifts the consumer out of high-cost payday loans to more affordable loans.
- A mobile phone service provider analyzing utility bill payments to identify creditworthy consumers who may otherwise be viewed as credit risks based on traditional credit report data.
- A mortgage company searching eviction records to see if an applicant has had problems paying rent.

The predominant uses of alternative data include initiation of firm offers of credit (e.g., prescreen, prequalifications), application decisioning, and account management:

- **Extending firm offers of credit**: Prescreened offers of credit are often thought to present a better customer experience and less friction than other forms of customer acquisition due to the targeted nature of these campaigns. The limited nature of credit bureau information can be a barrier to identifying new and traditionally underserved consumers to which prescreened offers could be made.
- **Improved application decisions and risk segmentation**: Over the past few years, the power of alternative data has become evident, particularly for companies that serve no-file and thin-file customers about whom credit bureaus have little or no data. There is, however, a growing recognition that alternative data can be used in ways beyond merely substituting for missing traditional data. It can supplement credit bureau data and give companies the ability to get a more complete view of a consumer’s economic life.
- **Account monitoring and maintenance**: Since alternative data captures consumer activity with entities that do not contribute information to the credit bureaus, users of alternative data can gain a better view of events in their customers’ lives and how that might impact a given customer’s risk profile. Alternative data can provide an early warning system, indicating when a customer’s capacity to pay might be at risk of changing. This could give companies more time to work with that customer to mitigate risk.

Similar to traditional credit data, alternative data used for credit decisioning can be provided as a score, report, or individual attributes. This flexibility allows lenders to consume the data in a way that is most convenient for their workflow. For example, a small lender who manually reviews all applicants may only require a consumer’s detailed alternative data report. On the other hand, a more sophisticated lender may identify a handful of attributes which they believe provide significant predictive lift over their existing models.

Alternative data is most often used within credit risk models, which function similarly to a model built from traditional credit bureau data (e.g., FICO Score, VantageScore). Often, these scores employ the
same three-digit format as mainstream credit scoring products. These empirically derived credit scores are designed to predict an applicant’s likelihood of default over an account’s lifecycle. Reason codes are also returned with the score, articulating the key factors driving the score. Additionally, models may vary by industry (e.g., bankcard, auto), population (e.g., subprime, no-hit), or use case (e.g., prescreen, account management).

Alternative data can be delivered online through a web interface, or it can also be delivered “machine-to-machine” (e.g., XML call) for real-time decisions or in a “batch” process for portfolio reviews or offline evaluation.

Although the aim of most lenders using alternative data is to increase access to credit, a sound underwriting strategy must consider a borrower’s stability and ability to repay. While traditional credit bureau data only uses historical payments to predict future behavior, alternative data can directly measure a consumer’s stability and ability to repay. Lending money to consumers who do not have the means to pay back their debts is unsustainable and detrimental to the financial health of individual consumers and the financial system as a whole. It is also unfair to consumers who expect financial institutions to offer products that are safe, non-predatory, and in the consumer’s interest.

Therefore, it is critical to balance information that reflects both positively and negatively on the creditworthiness of a consumer. Examples of alternative data elements and their relationship to financial behavior include:

- **Address Stability**: Consumers who do not move around a lot and are not associated with many different addresses are 5 times less likely to default than consumers with less stable address histories. Consumers who rent their place of residence can benefit from this data when rental payment history is not available; when a consumer’s rental address remains constant over a period of time, it suggests that the consumer makes regular rent payments.
- **Criminal History**: Consumers without previous convictions are 3 times less likely to default on a loan.
- **Derogatory Public Records**: Consumers without a lien or judgment are 50% less likely to default on a loan than those who have at least one lien or judgment. The presence or absence of evictions can be predictive of the risk of default as well.

Alternative data on the U.S. consumer population has long been available. However, access has been cumbersome, expensive and fragmented. The cost and difficulty of collecting and interpreting data from numerous sources for the purpose of determining creditworthiness has historically limited its use.

License and registration data can be obtained from individual state governments and from third-party aggregators and may include: voter registration, driver’s license, vehicle registrations, watercraft and aircraft registrations, and professional licenses. Landline phone directories, mobile phone listings, utility source records, and other consumer behavioral information are typically obtained from third-party furnishers.
One value alternative data providers offer is the ability to aggregate information across a wide variety of sources and handle linking and layout standardization so that an end user can easily consume this data.

Alternative data can have unique data structures depending on the type of data and the source from which it derives. Alternative data typically includes a subject’s name and address, and sometimes date of birth, social security number, phone number, or driver’s license number.

As a result of the number of unique reporting sources and the event-driven nature of alternative data creation, there is more variability in data update cycles for alternative data than with monthly credit reporting cycles. However, coverage of the U.S. population is higher than traditional credit bureau trade-line data.

Consumer disputes of data accuracy are handled according to applicable law, regardless of whether the data sources are traditional or “alternative.” We discuss this in greater detail below, in response to questions 18 and 19.

The ability to resolve input information to a single consumer file with high confidence is critical to delivering high performing data and models. Linking technology and matching logic considers multiple components of the consumer identity, including full name, SSN, date of birth, current address, previous address, phone number, and driver’s license number in its processes.

Access to, and application of, alternative data is not equal across industry segments. Consumer reporting agencies that leverage advanced linking and identity technology alongside expansive data sources have been able to achieve high levels of data reliability for public records. Companies who source records directly and maintain control over the retrieval, custody, and overall supply chain of data have a distinct advantage over providers who indirectly obtain this data.

4) Combination of Traditional and Alternative Data

Financial institutions employ alternative data to gain additional insight within a broader consumer risk assessment strategy. Many find that a more inclusive data strategy is critical to market expansion and risk segmentation across their books of business.

Many users of alternative data do so to gain a “secondary look” at consumers who were initially declined credit either because of poor credit history or a lack of credit history. In this scenario, the lender may find that while a traditional credit score is low or non-existent for a given consumer, the alternative data score may be high due to the inclusion of previously overlooked variables. The lender can now successfully extend credit to the consumer while maintaining appropriate risk levels in their portfolio.

According to research from McKinsey & Company, “new alternative data models have cut credit losses in experimental forays into lower-income segments by 20 to 50 percent and doubled their application approval rates.” Another study on an auto lending portfolio found a model that uses only credit bureau data would lead to approval of 74% of applicants at a default rate of 3%. By incorporating alternative
data, the lender could increase their approval rate to 85% while still maintaining a 3% default rate, resulting in up to 24% more loan approvals than a bureau only model. The incremental benefit comes from the inclusion of new consumers who would have previously been denied credit due to a lack of information or presence of negative information on their credit profile. Even among the lowest scoring consumers in a traditional credit bureau based model, alternative data can help identify the handful of borrowers who exhibit financially responsible behaviors.

In other cases, a “hybrid” model that contemplates both traditional and alternative data within a unified score may prove more meaningful. These holistic models can be applied to almost any consumer across the credit risk spectrum for improved decision making.

**Potential Benefits and Risks to Consumers and Market Participants**

9 and 10) Evidence Regarding the Potential Benefits for Consumers of Using Alternative Data and Modeling

Alternative data can address the significant challenges faced by financial institutions when trying to serve the unbanked or underbanked adults. Many of these individuals are currently shut out of mainstream credit offers due to the lack of a file at a national credit bureau. Estimates put the number of “unscorable” U.S. consumers at between 50 – 60 million depending on product or scoring methodology. Research shows that alternative data can score between 80-90% of these otherwise unscorable consumers, increasing the scorable population by more than 40 million. Of the newly scorable consumers, more than half exhibit credit patterns similar to “prime” or “near-prime” borrowers.

Credit bureau thin-files are traditionally defined as consumers having very few credit accounts (i.e. 1 or 2) or very limited time since their first credit account was opened (i.e. less than 2 years). In addition, thin-file populations tend to be young adults with little public record history to go along with their limited credit history. Typically 30% to 40% of young adults who have thin files have other public record or other alternative data history such as utility records, occupational licensure, court or law enforcement records or other behavioral data.

A more progressive approach to data allows lenders to give a “second look” to applicants who would have previously been denied credit due to a poor score or lack of traditional credit information. Research by FICO found that the majority of the 28 million consumers with scant or stale bureau data, traditional scoring would not provide for them to establish credit. About 65% of these consumers have

---

a negative item and no active account. With no positive data flowing into their files to offset the negative, they would likely score too low to obtain credit. Additionally, the same research revealed two key elements about consumers that become scorable as a result of alternative data. First, more than a third of these individuals receive a score of 620 or above. Second, and perhaps more importantly for the long-term credit opportunities of thin-file individuals, once these consumers have been brought into the mainstream economy, they can build a more traditional credit profile. FICO recently conducted a study to predict the future creditworthiness of consumers using FICO Score XD, their credit score using alternative data. The test determined that nearly 80% of customers with a >620 FICO Score XD achieved a traditional FICO Score 9 of >740 two years later. At least 25% achieved a FICO Score 9 of 740 or greater, indicating a large number of consumers who are unscorable will become prime or near-prime consumers once given the opportunity to access credit.

Individuals with no file at the credit bureaus include populations of young adults who have a similar public record footprint to credit bureau thin-files. However, such individuals also include those substantial populations who have chosen to not use credit or are unable to obtain credit. Such individuals create public record and alternative data behavioral histories over time as a result of residential address histories (utilities, phone service, etc.), government interactions (licenses, vehicle registrations, etc.), court and law enforcement interactions (judgments, liens, convictions, etc.), occupational records, and even in some cases property ownership (unrecorded mortgages, family deed transfers, marriage, etc.) or bankruptcy caused by debts not typically reported to credit bureaus. There is also a significant population of former credit users who have had no credit activity in the past seven years and, therefore, are no longer reported by the credit bureaus despite their often-rich behavioral history.

The challenges with scoring no-file or thin-file individuals differentially affects historically disadvantaged minorities. A recent Lexis-Nexis study found that 41% of historically underserved minority populations of Hispanics and African-Americans could not be scored using traditional methods, while the unscorable rate for the general population was only 24%. Minorities face an unscorable rate that is 1.7 times the rate — almost twice — the rate for the general population.

Alternative data sources can substantially help to provide accurate scores for many of these individuals. For instance, an alternative credit score, called RiskView, built by Lexis-Nexis relies on public and institutional data such as educational history and professional licensing, property asset and ownership data such as home ownership, and court-sourced items such as foreclosures, evictions, bankruptcies, and tax liens. The Lexis-Nexis report demonstrated the extent to which credit risk scores built from alternative data can help to extend credit to unscorable consumers, finding that fully 81% of the unscorable minorities received a RiskView score.

---


5 Ibid.

Moreover, incorporating alternative data sources into traditional measures of calculating credit scores can differentially increase access to credit for low-income and minority consumers without over-extending credit. For example, with the addition of utility payment data in VantageScore’s credit score, acceptance rates to access credit for those making less than $20,000 per year, African-Americans and Hispanics increased by more than 20%. Notably, incorporating utility payments also did not “overextend credit.” In other words, the Political Economy Research Center (PECR) found that these higher acceptance rates did not offer credit at unaffordable prices or to those who could not afford to repay the loan.\(^\text{7}\) Clearly, a major benefit of alternative credit scores is the improvement in the availability of credit for historically underserved minority groups.

The longer an adult lives and works, the more opportunities he/she will have to create a history in public records or behavioral data that supports credit scoring. Many no-file populations have had a longer time to develop such histories.

Another demonstrable benefit of the application of alternative data is its usefulness in creating competition amongst credit providers, thereby decreasing the overall cost of credit to consumers. Before widespread adoption of alternative data, many lenders lacked the necessary proprietary tools to better segment customers. Today, companies use alternative data to more appropriately price products in relation to a consumer’s ability to pay—in many cases the outcome is better terms for borrowers. Moreover, smaller and more innovative lenders have been able to capitalize on their familiarity with new data and advanced modeling techniques to offer more affordable products to consumers.

**11 and 12) Evidence Regarding the Potential Benefits for Market Participants of Using Alternative Data and Modeling**

Alternative data can be used to inform auto lenders in a way that results in benefits to consumers. In auto lending, dealers solicit bids from multiple lenders. The financial institutions compete by offering aggressive terms while keeping an eye on their target loss rate. Lenders and service providers that use alternative data have more information about the applicant and are often able to provide better interest rates, discounts and more. The auto lending industry has embraced alternative data as a source of competitive advantage, offering the right terms to the right customers.

Similarly, credit card companies have upgraded their offers to “borderline” credit customers without taking on more risk. Cable companies have revised their deposit requirements in much the same way as the telecom industry. Sophisticated online “small-dollar” installment lenders have created more consumer-friendly products to displace expensive and punitive payday loans. Alternative data allows these industries to expand their addressable market in a responsible manner, improve their risk management and increase their sales efficiency.

---

15 -16) Risks to Market Participants of Using Alternative Data and Modeling, and Mitigation of These Risks

Despite the benefits of using alternative data to inform credit decisions and customer relationship management, many companies are hesitant to fully incorporate alternative data into their decisioning processes because they have difficulty realizing a positive economic return from the data. That is, the benefits do not yet always justify the costs they would incur. While progress has been made in recent years with respect to the collection and management of alternative data, there are improvements yet to be made:

- **Data attributes:** Industry executives are most interested in information regarding the payment of telecom and utility bills, since that is information that credit bureaus currently do not capture. Other areas of interest include housing and asset ownership information, educational data, employment information, income, and transactional banking account information.

- **Ease of use:** Given the high level of automation in the credit card and other lending industries, companies want to be able to incorporate new data sets into their current systems with as little disruption and additional cost as possible. Some data providers currently provide data that is not compatible with all systems. Furthermore, some companies are hesitant to use stand-alone systems due to the additional step that such a system would require. To mitigate these concerns, alternative data providers should look into ways to make alternative data as easily ingested as possible. In addition to making their data sources more easily accessible through current systems, providers could explore working with the big processors (such as TSYS and First Data) so that lenders do not need to each have a stand-alone system or data connection. Another possible solution would be for the providers to work with the credit bureaus to tie their databases into the credit bureaus so that one query could hit both credit bureau and alternative data attributes. Alternative data providers should look into ways to make accessing their data as easy as possible.

Specific Statutes and Regulations

18-19) Applying the FCRA and ECOA to Alternative Data

Existing statutes have proven flexible enough to address alternative data and new modeling techniques. For instance, the Fair Credit Reporting Act (FCRA) requires credit bureaus (and other businesses that report information on consumers for employment, credit, or insurance purposes) to maintain standards designed to ensure “maximum possible accuracy” and completeness with respect to consumer information that they report. Covered businesses must provide consumers with access to the data and have a process in place for investigating and resolving consumer disputes in a timely manner—this infrastructure needs to provide for compliance at a significant rate when used in high-volume scoring. The Federal Trade Commission (FTC) promotes the efficiency of the nation’s banking and consumer credit systems through enforcement of eligibility decision-making, a role they have taken very seriously since the law was first enacted in 1970.
Companies that regularly engage in the practice of assembling non-traditional, or alternative data, about individual consumers should be considered as providing a “consumer report” under the FCRA when the information involved is “used or expected to be used or collected in whole or in part for the purpose of serving as a factor in establishing the consumer’s eligibility for credit...” 8  Their obligations with respect to non-traditional data are the same as those which apply to any other kind of data currently provided in a consumer report, including those relating to disclosure of information to consumers, accuracy or completeness of information, and consumer rights of access and correction. This is consistent with the Federal Trade Commission’s settlement in its complaint against Spokeo in 2012 that alleged that Spokeo had acted as a credit reporting agency in the employment context. In that case, the FTC alleged that Spokeo marketed its service to HR departments through the slogan “Explore Beyond the Resume,” and created special portals and services for HR departments. 9

Additionally, the Equal Credit Opportunity Act (ECOA) prevents discrimination in the granting of credit, and it creates liability for both intentional (e.g., race-based) and unintentional (e.g., disparate impact) theories. Credit systems, like FICO scores for instance, must be “empirically derived,” as well as “demonstrably and statistically sound” in order to be used without fear of liability. A careless approach in the use of such data can easily lead to lawsuits if, for example, the use of a facially neutral criteria had the effect of discriminating against married people, the elderly, or another protected class. For any organization that provides scores for millions of consumers, that liability could be devastating.

Credit scoring agencies regularly update their algorithms not only to avoid litigation, but also in response to new information about how their algorithms could be improved. Just three months after the CFPB released a data study 10 in May 2014 indicating that FICO credit scores overstated medical debt collections, FICO introduced new measures to collect consumer information. 11 In an effort to differentiate medical and non-medical collections, FICO bypassed collections accounts and offered sophisticated software to ensure that medical collections do not proportionally have a larger impact than non-medical collections. These measures are expected to improve credit scores by 25 points for consumers whose only negative references are from medical collections. Following the data point study in May 2014, FICO simply recognized the problem and tweaked their software to ensure that credit scores are more accurately reported and to protect the integrity of their product. This FICO response to the CFPB data point study clearly indicates that private companies in the credit scoring business respond to government studies that suggest the need for reform in order to protect the integrity of their product.

Given the broad application of these statutes, companies have legal and regulatory incentive to evaluate non-traditional data and modeling techniques for fair lending risk, just as they do with traditional data

---

8 15 U.S. Code § 1681a(d)(1), available at: https://www.law.cornell.edu/uscode/text/15/1681a
and modeling techniques, including disproportionate adverse impact on a protected class such as those defined by age, sex, race, and ethnicity. Such evaluations include:

- Steps based on recognized thresholds, standards or baselines to determine whether alternative data may be serving as a proxy for a protected class or whether the use of such data has a disproportionately negative impact on a protected class.
- Steps 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.
- Steps 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.
- Companies should undertake these evaluations for all models related to lending process including, underwriting, pricing, fraud, and marketing, with the extent of the evaluation proportional to the likelihood and size of the harm in each use.

Companies using non-traditional data and modeling techniques should establish and maintain risk management programs pursuant to the guidance on model risk management issued by prudential regulators. They should use the principles and processes discussed in that guidance in connection with their management of fair lending risk, including requirements for validating the accuracy of predictions of creditworthiness at the time of model development and on a regular basis as the models are in use.

It is important for marketplace participants and regulators to recognize a distinction between non-traditional data that has been shown to be reliable and predictive, such as the examples highlighted above, and non-traditional data whose usefulness has not yet been demonstrated and is not likely to stand up to regulatory scrutiny and comply with the existing legal framework. Additionally, some social media companies do not allow other companies to know information about their users for eligibility decisions, therefore precluding its appropriate use in these contexts at this time.12

Summary and Conclusion

Again, thank you for the opportunity to provide comments regarding the use of alternative data and modeling techniques in the credit process. SIIA applauds the CFPB for seeking to explore the benefits and risks associated with “next generation data” and new modeling techniques. In summary, the key objectives of using alternative data are to two-fold: To identify new pools of credit borrowers who would not be selected under traditional standards, and to increase the ability to accurately assess credit risk. Both of these objectives are in the best interest of consumers and market participants, as they lead to a greater expansion of credit, and better credit decision-making. Therefore, SIIA strongly supports the goal to encourage greater use of alternative data and modeling techniques, while ensuring that this is

12 Facebook, for instance, says that their users’ data should not be used to “make decisions about eligibility, including whether to approve or reject an application or how much interest to charge on a loan.” See section 3.15 of their platform policy, available at https://developers.facebook.com/policy/.
done responsibly to mitigate various risks. Finally, we are confident in the ability of existing statutes to address broader adoption of alternative data and new modeling techniques.

I hope these comments are helpful in your review of this very important topic. If you have follow-up questions or would like to discuss further, please contact David LeDuc, SIIA Senior Director for Public Policy, at dleduc@siia.net or (202) 789-4443.

Respectfully Submitted,

Ken Wasch
President & CEO
Software & Information Industry Association