

September 10th, 2014

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

RE: Docket No. CFPB-2014-0012

Dear Sirs:

Reinvestment Partners is a 501 © 3 organization in Durham, North Carolina whose mission is to help traditionally-underserved borrowers and communities to access fair, transparent, and high-quality financial products.

Please accept our comments on the topic of mobile financial services.

We see potential pitfalls for under-served consumers in this new environment. Our main concern has to do with how data analytics can create new opportunities to harm protected classes. This new technology contemplates a system that could validate and reinforce the chance that under-served consumers will be disadvantaged. We think there are many reasons that justify further exploration on the question of how mobile financial services could have a disparate impact upon individuals in protected classes.

The Executive Office of the President produced a compelling report that included some caution about implications for big data with respect to protected classes, including the instance of financial services:

“A significant finding of th[e] report is that big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace.(President 2014)”

The unfortunate problem with big data is that it creates a system where disparate impacts are plentiful even if disparate treatments are not. Big data creates thousands of new possible methods for disparate and adverse impacts upon protected classes.

Mobile financial services are operationalized through new analytical technologies. Often referred to as “big data” or “data mining,” these methods create the sense that any disparate impact is excusable because of empirical evidence.

A second issue concerns general privacy problems. Worth noting in our opinion is that privacy concerns are often caused by the same big data techniques. In short, big data is the problem. Yet without a doubt, big data is here to stay. No regulatory action could conceivably reverse this change to our society. But this means that a regulatory response should address how big data is operationalized.

The proper unit for regulatory analysis is the creator of the analytics. An analytics firm might sell its insights to a number of lenders. The CFPB should attempt to make sure that impactful analysis is stopped before it is sold to lenders or other companies that use credit data to make determinations about their consumers.

Underwriting with geo-spatial data could harm consumers in under-served areas

Many alternative underwriters make use of geo-spatial data as part of their analysis. The systems usually gather a person’s geo-spatial history. Geo-spatial history is a more than merely where a person resides; it consists of a footprint of a person’s movements.

We believe that with the exception of products related to transportation (but not to auto lenders), geo-spatial history should not be a part of alternative underwriting models. Our reasoning is based on previous concerns that show how many lenders provide less capital to poor and minority neighborhoods. What if an algorithm determines that individuals who travel through certain areas tend to be less reliable credit risks? If those areas were minority areas, then it is hard to imagine how such a variable would not end up having a disparate impact. One employment consultant has decided that distance between home and work is a valuable estimator of employee performance (Peck 2013). In communities where employment opportunities are further away from lower-income neighborhoods (not an infrequent situation), such an algorithm poses the possibility of creating disparate impacts.

Imagine that an analytic determines that loans to residents in a minority neighborhood tend to default on loans more often. An underwriting model that used this as an independent variable would be more likely to produce disparate impacts upon minority loan applicants.

Geo-spatial data can also be useful in determining with whom a person tends to associate. If a data aggregator can determine the proximity of two smart phones to each other over time, then it is simple to conclude the nature of social networks. It is even easier to do this through new social media.

We do recognize that there are cases where geo-spatial data is proper to use. Currently, a number of car insurance companies are using geo-spatial data to better understand the driving habits of their customers. This seems to us like a virtuous application of geo-spatial data. Unlike inferential data, driving behaviors are legitimate factors for safety. The recording of geo-spatial data is not inferential. Rather, it tells car insurance companies exactly how their customers will drive. To be clear, this is different than is the case in many other uses. For example, a person’s geo-spatial data could infer something about a person’s credit worthiness, but it is very possible that an algorithm could incorrectly interpret the meaning of such a record.

Social media creates a number of opportunities for disparate impacts

We are concerned that relationships will be vetted in the process of qualifying individuals for various financial services. One lender told us that his company uses Facebook data to infer a variety of things about his loan applicants. His company taps records of interactions to create a hierarchy of relationship strength. The most intense connections are used as proxies for the applicant's trustworthiness. If an applicant has long-standing social media relationships, then his company interprets it positively. Our concern is that people will suddenly be downgraded because they belong in a group with poor financial attributes. This would thwart social mobility.

We suspect that many underwriters are using social networks as inputs for their estimates. Sociogramics, for example, has developed models that incorporate LinkedIn and Facebook profiles to a credit algorithm. Sociogramics sells those conclusions to lenders and other companies that use this kind of data. Their access to a mobile phone further informs how they use social media. For example, if a person attests that they work in a location that is not a part of their regular geo-spatial pattern, it throws up a red flag. Consumers must opt-in to provide Sociogramics with their data. But that leads to other challenges:

Terms and Conditions for Apps or In-Store Wi-Fi should not be allowed to garner the right to sell data to third-parties for underwriting.

While it is often true that consumers have agreed to terms and conditions that authorize the sharing of their information, in reality this is only rarely an intentional decision. Very few people read the entirety of a disclosure before they download an app. Those apps are often authorized to sell the wide variety of data that can be gleaned from a phone.

It is the same with giving consent to use in-store Wi-Fi. Most consumers are shocked to know that "big box" retailers have the ability to track in-store movements. Most are not aware that spending time in a paint store can conceivably lead to an advertisement on their email folder. Most are not aware that the "offer" on their phone could be factored by a "Customer Lifetime Value" score. Unfortunately, all of these outcomes are made possible by a decision to accept free Wi-Fi or to download an app. We doubt that even a tiny fraction of consumers are aware of how their acceptance of terms and conditions can subsequently affect their financial affairs.

Media consumption should not be utilized in alternative underwriting for financial services:

One CEO of a non-bank payday lender told Reinvestment Partners that while his company had not done so, some of his professional peers are using media preferences as an indicator for their underwriting. In such a system, an analyst would look for different customer profiles based upon reading history, favorite bands, or television shows. It is hard to imagine how media preferences would not create proxies for race, gender, religion, sexual orientation (federally-protected), marital status, and age. It would not be right if a car lender turned down an applicant because their underwriting vendor discounted the person for what they read.

Consumers are free to release any of this data on their own volition, but the arrival of big data should not mean that this information is used to gauge their qualification to buy a car or receive a mobile "offer."

Given how publications serve specialized audiences, media preferences are very nearly explicit in how they identify consumers. Would it be difficult to ascertain the sexual orientation of a subscriber to Q-Notes and The Advocate? Would it be hard to determine the age of a subscriber of Modern Maturity (senior over 55)? What

can be inferred about a person who posts reviews on Good Reads/Amazon of books published by Zondervan Press and Saddleback Church (evangelical Christian)?

The issue is not limited to consumption, either. Being a “fan” on Facebook of a television show is almost as explicit. When “fandom” of one media product is combined with records showing “fandom” of other media, a marketer can quickly derive a highly predictive inference of all of the characteristics that make up membership in a protected class. For example, what if data determined that fans of Tyler Perry were less credit-worthy than those of the Dukes of Hazzard? It is likely that this kind of input would end up creating a disparate impact. There are protections in place for library check outs. A similar firewall would be helpful in stopping problems that are likely to develop from using media preferences. It is probably a foregone conclusion that marketers will record our media habits (as well as other choices) but regulators should address the possibility that this information will become an element of credit scoring.

The end result could be a future where people consider the opinion of alternative data analysts when they make personal choices about media choices. If credit scoring begins to incorporate search engine requests, then this could lead to a system that curbs a consumer’s choices during scores of moments every day. This is an over-reach. In truth, we all need protection from this kind of result.

Establish a sunset period for social media data

The online marketers association has published a voluntary set of best practices. One of those is that there should be a limit to how far back a company can “scrape” data from social media. This is similar to current rules that prohibit traditional credit bureaus from using defaults and delinquency data that is more than seven years old. We think that there should be such a limitation in place for analytics and that it should be far shorter than seven years.

Separately, he confirmed that his company has an internal practice of blocking its ability to view pictures. This is an important behavior and one that we think should become an element of a rule-making. He cautioned that there would be natural concerns if his underwriting could infer the racial makeup of applications.

Disclosures for denial of credit

For some time, lenders have been expected to explain why they turn down a borrower for a loan. In practice, this means that they refer to the credit report that they received from a bureau. At the moment, this same expectation is not in place for decisions made based upon alternative “big data” underwriting. This loophole should be closed.

Options for Enforcement

Under existing law, a disparate impact claim can assert that a defendant should have used an alternative method without the same discriminatory results. This seems like a fairly easy remedy, given that there are thousands of data points that could conceivably be chosen from as alternatives. Given that alternatives are readily available, the claim that such a choice is a “business necessity” seems to hold little merit.

Big Data is readily available to defend itself on the grounds that it studies the behaviors of individuals and not of classes. Moreover, outputs from big data are not made by individuals but instead by machines. Presumably, machines have none of the ability to demonstrate prima facie discrimination. But that argument is false. Machines are the products of the individuals that create them.

The hard part is that exclusions can lower the predictive power of algorithms. Even the slightest deterioration in the quality of a model could potentially serve as the grounds for insisting that the cost of conforming to a standard is too high. Other exclusions will do little to change the ability to identify protected class characteristics. Others have argued that removing one data point from an algorithm that taps many inter-related data points may do nothing to change the ability of a model to identify race, gender, religious preference or other variables.

One benefit, though, is that “rogue individual” excuse will no longer be viable. In our experience, banks and their prudential regulators begin any response to a fair lending complaint by asserting that it is a product of a “rogue individual.” The rogue individual is fair lending’s answer to the “rogue trader.” There is no “rogue data algorithm.”

One potential alternative enforcement approach would be to examine the outcomes that result from analytics. If purchasers of an alternative data set repeatedly use that data as a basis for adverse treatment, then this could be a basis for a disparate impact suit.

Conclusion:

There are many virtuous opportunities that could be the product of the advent of mobile financial services. It addresses a long-standing known concern with traditional credit scoring, where many consumers were shifted to the periphery because of a lack of information or because the relevant information only told one part of a fuller picture of their financial lives. With “big data” and mobile financial services, the problem is one of plenty. There are unlimited sources of data, and even though some individuals are still “under-datified(Lerman 2013),” their ranks are fewer.

In many cases, data drawn from smart phones will be the difference that brings people into the mainstream. In 2011, the FDIC reported that sixty-nine percent of the unbanked had a mobile phone. In those cases where there was a mobile phone in use, half had smart phones. Among the underbanked, the rates are even higher: 88 percent have a mobile phone and almost two-in-three of those underbanked consumers have a smart phone(FDIC 2013). By 2013, 90 percent of the unbanked and underbanked had a mobile phone, of which 71 percent had a smart phone(FDIC 2014).

One of the key things that we want to underscore is how the assumptions drawn from big data can create falsehoods. A traditional critique of credit scoring is that it sorts applicants by a standard that is has derived to serve as a predictor of credit worthiness. For some individuals, the reason their credit score is poor is only because of how the model gathers its independent variables. It turns out that this problem is not infrequent. Millions of Americans find it difficult to access prime credit solely because they have not yet developed a full set of inputs for a FICO algorithm.

But big data’s image of precision can be misleading. There are a number of ways that analytics can fail. For one, even the most scientific effort can be undermined with bad inputs that are themselves given to discriminatory conclusions. For example, research by professors from the Wharton Business School and Case Western Reserve University found that lenders participating on peer-to-peer lending service Prosper were systematically discriminating in offering and pricing loan applications made by borrowers whose pictures revealed that they were either African-Americans or female (Pope 2012). While these transactions were not made on a mobile platform, they do show how there is a “discrimination in-discrimination out” possibility in

any alternative underwriting model. In many instances, the selection of characteristics for potential inclusion in an algorithm are the product of human bias(Barocas 2014).

There may be significant and unforeseen consequences as well

Imagine what happens when factors drawn from as far and as wide as should be contemplated with mobile become in play as factors for credit-worthiness. The scenario becomes possible where people allow credit scoring to influence many of their personal decisions. Going back to an earlier example, consider what happens if credit scoring models determine that a subscription to the Advocate suggests the chance that a person is more likely to default? In that world, a person would have to weigh the cost in determining what and how they read.

Lenders should be able to employ underwriting. Access to credit is not an entitlement. We agree that it is more than reasonable to restrict someone's access to credit if they have defaulted on loans in the past or if they have presented too many bad checks. Those actions demonstrate the kinds of behaviors that underwriters deserve to use in their decision-making. But it is entirely different to use many of the "odd but true" kinds of variables that can be gleaned.

There are many companies with good intentions who use alternative financial data to increase access. I spoke with a subprime auto lender recently on the value of alternative data. In his opinion, a traditional FICO score lost most of its predictive power at the bottom end of the credit scoring spectrum. They had developed their own internal set of metrics. The company was making loans to many borrowers that would otherwise been considered risky. Using the alternative data was a net benefit for some of who might have been turned down in a FICO-only system.

But even a well-intentioned company may include inputs of concern. If those inputs increase the predictive power of their logistic regression model, it will present an attractive proposition to include such factors. But our concern is that nothing will be off-limits. Without a doubt, some factors will have disparate impacts on protected classes.

"Disparate treatment is viable because data mining systems treat everyone differently; that is their purpose. Disparate impact should be viable because, as the Article shows, data mining can have various discriminatory effects. But as the Part concludes, data mining combines some well-known problems with discrimination doctrine with new challenges particular to data mining systems, such that liability for discriminatory data mining will be hard to find.(Barocas 2014)"

It is unlikely that any rule is going to return privacy to where it was before the advent of the smart phone. The world has changed, so the question is how to adapt rules to this new era.

This comment has focused on big data and its potential to create disparate impacts. We want to underscore that while it is commonly in use outside of mobile financial services already, the increasing adoption of smart phones will make its impact far more widespread. Mobile financial services will create the data that becomes the input for these potentially discriminatory algorithms. As such, regulators should be focusing on this data source.

We support a strong rule from the CFPB which will protect consumers from discriminatory algorithms in mobile financial services.

Thanks for your concern.

Sincerely,

Adam Rust
Director of Research
Reinvestment Partners

Bibliography

Barocas, S. a. A. S. (2014). Big Data's Disparate Impact. Social Science Research Network.

FDIC (2013). 2011 Household Survey, FDIC.

FDIC (2014). "Mobile Financial Services and Inclusion."

Lerman, J. (2013). "Big Data and its Exclusions." Stanford Law Review **66**.

Peck, D. (2013). They're Watching You at Work. The Atlantic.

Pope, D. a. J. R. S. (2012). "What's in a Picture? Evidence of Discriminator from Prosper.Com." Journal of Human Resources **46**(1).

President, E. O. o. t. (2014). Big Data: Seizing Opportunities, Preserving Values.