HARVARD Kennedy School
MOSSAVAR-RAHMANI CENTER
for Business and Government

Revolution in Data:
How new technologies are upending borrowing

Marshall Lux
Guillaume Delepine

February 2019

M-RCBG Associate Working Paper Series | No. 107

Mossavar-Rahmani Center for Business & Government
Weil Hall | Harvard Kennedy School | www.hks.harvard.edu/mrcbg

Revolution in Data: How new technologies are upending borrowing

February 2019

Marshall Lux

Senior Fellow, Mossavar-Rahmani Center for Business and Government, John F. Kennedy School of Government, Harvard University

Senior Advisor, The Boston Consulting Group

marshall_lux@hks.harvard.edu

Guillaume Delepine

Research Assistant, Mossavar-Rahmani Center for Business and Government, John F. Kennedy School of Government, Harvard University

Master in Public Policy candidate and David M. Rubenstein Fellow, John F. Kennedy School of Government, Harvard University

Master of Business Administration candidate, Harvard Business School

Guillaume_delepine@student.hks.harvard.edu

Abstract

This working paper focuses on the implications of Big Data in consumer lending. It begins by examining the multi-dimensional definitions of Big Data, as a technology and as a credit assessment tool. We then present findings from a series of industry interviews to share the state of the art of Big Data after the latest round of investment and technological change. In consumer lending, Big Data appears to serve as a valuable tool in preventing fraud, engaging customers more deeply, and making more confident lending decisions. Our stakeholder analysis indicates that there is potential for Big Data to benefit lenders, startups, the overall economy, and individual consumers. However, regulators should be watchful that Big Data does not abuse consumer data, either through increased aggregation or expanded ability to generate insights. We conclude by offering a framework and set of recommendations for preparing government to regulate the consumer lending space in the era of Big Data.

Acknowledgements

We are deeply grateful for the support of the Mossavar-Rahmani Center for Business and Government at The Harvard Kennedy School. We also gratefully acknowledge the helpful insights of Bhaskar Dole, Ash Gupta, David Knott, Sushil Malhotra, Sonia Brodski, Melissa McSherry, Raul Vazquez, Suk Shah, Audra Schoenfeldt, Daniel Sumner, Todd Lowenberg, Andre Pimento, and Scott Bauguess. All errors are our own.

CONTENTS

Introduction

In this paper, we evaluate the role of Big Data in the provision of credit services. Big Data has held the attention of the financial sector for years, offering the promise of increased profits, new services, and finely-tuned risk management techniques. Big Data reached the peak of its hype cycle in 2014, and then outgrew its own position – in 2015, Forrester dropped Big Data off of its annual hype cycle for emerging technologies report, their lead analyst, Betsy Burton, explaining that "Big Data has become a part of many hype cycles."[1] Illustrating this fact, the Forrester Wave for Big Data included no fewer than fifteen companies in Q2 2018, highlighting the growth in investment and new entrants in the space.[2] In banking and consumer credit, The Boston Consulting Group has entered the fray, releasing thought leadership called

[1] Alex Woodie, "Why Gartner Dropped Big Data Off the Hype Curve," Blog, Datanami, August 26, 2015, https://www.datanami.com/2015/08/26/why-gartner-dropped-big-data-off-the-hype-curve/.
[2] Noel Yuhanna, "The Forrester Wave™: Big Data Fabric, Q2 2018," Denodo, Q2 2018, https://www.denodo.com/en/page/forrester-wave-big-data-fabric-q2-2018.

"Data Analytics for FIs: The Journey from Insight to Value."[3] In an attempt to do just that, the financial services industry invested nearly $9 billion in Big Data in 2018 alone.[4]

There is mounting evidence of the potential for Big Data to enable consumer lenders to design better products and better engage their customers. While some applications come in traditional enterprise mechanisms such as cyber security, [5] an increasing amount of thought leadership has focused on ways for Big Data to enhance consumer banks' product offerings, critical operations, and business models. These frameworks corroborate banks' investments to suggest that Big Data may lead to a high ROI for lending institutions. Chen et. al taxonomized the Big Data value chain into four phases: data generation, data acquisition, data storage, and data analysis, in 2014. They defined Big Data as datasets that cannot be "acquired, stored, and managed by classic database software," in volumes from several terabytes to several petabytes.[6] The reward for accomplishing the technical breakthroughs needed to harness Big Data were again discussed by Garg et al, who proposed three mechanisms for unlocking increased profits from Big Data: Boosting Traditional P&L Levers (Accelerating Growth, Enhancing Productivity, Improving Risk Control), Finding New Sources of Growth, and Delivering the Digital Bank. [7] In this paper, we address the second lever, through which banks will use Big Data to identify new growth opportunities – harnessing predictive tools to assess risk more precisely and enter markets typically considered too risky for traditional lending.

For policymakers concerned with consumer access to credit, Big Data may be a driver of positive social outcomes. The Consumer Financial Protection Bureau recorded 659 million open credit card accounts in the United States in mid-2017, [8] whose total credit lines totaled 20% of GDP.[9] As of September 30, 2017, 181 million consumers in the US had access to credit cards,[10] a large total, but only approximately 56% of the population overall.[11] In 2017, 22.7% of applicants for credit were denied, with particularly high denial rates among African-American and Hispanic groups earning less than $40,000 per year, who were denied at 50.5% and 44.1%, respectively. In addition to the applicants who were denied, another 16.1% were granted less credit than they applied for, and 16.7% decided not to apply because they thought they would be

[3] James Faucette et al., "Data Analytics for FIs: The Journey from Insight to Value" (The Boston Consulting Group & Morgan Stanley, October 25, 2016).
[4] Laura Wood, "Big Data Investments in the Financial Services Industry Will Account for Nearly $9 Billion in 2018 Alone," PR News Wire, August 1, 2018, https://www.prnewswire.com/news-releases/big-data-investments-in-the-financial-services-industry-will-account-for-nearly-9-billion-in-2018-alone-300690188.html.
[5] Mark Jeffs and Christy Williams, "Fiserv and Mastercard Help Financial Institutions Approve More Genuine Transactions Without Increasing Risk," accessed November 7, 2018, https://www.businesswire.com/news/home/20180306005515/en/Fiserv-Mastercard-Financial-Institutions-Approve-Genuine-Transactions.
[6] Min Chen, Shiwen Mao, and Yunhao Liu, "Big Data: A Survey," *Mobile Networks and Applications* 19, no. 2 (April 2014): 171–209, https://doi.org/10.1007/s11036-013-0489-0.
[7] Amit Garg et al., "Analytics in Banking: Time to Realize the Value," McKinsey, April 2017, https://www.mckinsey.com/industries/financial-services/our-insights/analytics-in-banking-time-to-realize-the-value.
[8] Consumer Financial Protection Bureau, "The Consumer Credit Card Market 2017" (Consumer Financial Protection Bureau, December 2017), https://www.consumerfinance.gov/about-us/newsroom/cfpb-releases-report-state-credit-card-market/. Page 35.
[9] Consumer Financial Protection Bureau. Page 38.
[10] Bob Sullivan, "State of Credit: 2017," Company Blog, Experian, January 11, 2018, https://www.experian.com/blogs/ask-experian/state-of-credit/.
[11] US Census Bureau, "Population Clock," November 7, 2018, https://www.census.gov/popclock/.

denied.[12] However, it would be incomplete to assume that the reason for unsatisfactory access to credit in America is a failure to confidently apply. Credit confidence is not the problem – 78.1% of respondents claim to be somewhat or very confident in their ability to secure a credit card. In marginalized populations, African-Americans and Hispanics with incomes below $40,000, 55.1% and 57.2% of respondents, respectively, are somewhat or very confident they would be approved for a credit card.[13]

There is no single answer to what prevents more people from accessing credit, but multiple drivers' impacts may be reflected in the general population's credit scores. Experian reports that 21.2% of Americans have subprime credit scores, or scores below 660. [14] Credit scores range from 300 to 850. And 660+ is considered prime. People without prime credit scores either hold scores below 660, or are non-scored. In 2016, there were 26 million users without credit scores and 19 million with a financial record but no credit score. [15] These credit applicants generally face four options: Unsecured general purpose cards from mass market issuers, accounts offered by subprime specialists, private label credit cards, and secured cards offered by mass market issuers. The last option most resembles a product available to someone with a prime credit score, but the credit line is capped by a deposit put down by the customer at card issuance. In some cases, users can "graduate" to the mass-market card option. [16] The inability to access unsecured, mass-market credit leads to limited credit lines and higher costs for people without prime credit scores.

There are multiple opportunities for Big Data to alleviate some of the difficulties of accessing affordable credit for populations without prime credit scores. They are dependent on the way Big Data is implemented by major consumer lenders, and carry external risks which we discuss below.

The first mechanism by which Big Data may increase access to credit is by identifying loan-worthy applicants with credit scores below 660, or no credit scores. In the absence of credit scores, Big Data-enabled organizations can examine other factors to determine creditworthiness. This can be done using regression analyses of various complexity.

Second, Big Data can enable lower credit prices for qualified individuals whose credit scores would otherwise trigger high annual percentage rates (APRs). As banks compete for creditworthy customers, identifying applicants with low probability of delinquency may offer a competitive advantage, enabling them to confidently out-bid each other for customers by offering cheaper credit. This would benefit the consumers identified by the algorithms.

Third, applying Big Data to the existing customer base can help lenders reduce their overall risk and, in turn, offer lower prices to their existing customers. Offering lower interest rates may convince more applicants to apply for credit, including from under-banked and under-lent populations.

Fourth, Big Data-enabled organizations that invest in improving the transparency of their services can play a role in improving financial literacy. Compared to the opacity of credit scores, where the contributing factors, and sometimes even the most up-to-date score, can be difficult to

---

[12] Board of Governors of the Federal Reserve System, "Report on the Economic Well-Being of U.S. Households in 2016 - May 2017," FederalReserve.gov, June 14, 2017, https://www.federalreserve.gov/publications/2017-economic-well-being-of-us-households-in-2016-banking-credit.htm.

[13] Board of Governors of the Federal Reserve System.

[14] Sullivan, "State of Credit: 2017."

[15] Consumer Financial Protection Bureau, "The Consumer Credit Card Market 2017." Page 206.

[16] Consumer Financial Protection Bureau. Page 231.

find, services from providers highly competent in Big Data may help demystify the process. With greater information comes an increased ability for consumers and small businesses to make adjustments to the financial behavior within their control so as to position themselves for successful credit applications.

The success of these four mechanisms is contingent upon the way that credit companies work Big Data into their strategies. In one scenario, lenders may identify only the safest borrowers using their algorithm. This would make for a lower-delinquency pool, lower interest rates, and higher thresholds for entering the pool. This would generate cheap credit, but restrict access to capital overall. In another scenario, the algorithm may allow credit companies to open their pools to more people while holding the risk profile of their overall borrower portfolio steady. This would enable companies to perform land grabs on parts of the market where credit applicants have lower credit scores, leading to an increase in overall credit access.

To assess the ways that Big Data has and will be used by the consumer lending industry, we conducted interviews with industry-leading players across the value chain. Our conversations with thought leaders included Argus, The Boston Consulting Group, and Ash Gupta, former chief risk officer at American Express. We also interviewed credit card issuers and service providers representing companies including HSBC, Visa, and Fair Square Financial. We also spent time with emerging startups using Big Data to provide new and innovative services, in particular the near prime and subprime lending platforms Avant[17] and Oportun. Finally, we interviewed public and private sector compliance professionals from the SEC and TD Ameritrade. We are very grateful to these companies' contributions to our research, as their insights and decisions are likely to shape the future of Big Data in the financial services industry.

Our findings suggest that a public policy approach must be careful to enable innovation and stability for lenders without abandoning the goals of expanding access to credit and protecting historically underbanked populations. It will be important to identify the potential winners and losers of unchecked Big Data investment and implementation to ensure that any policy solutions adequately level the playing field.

The first potential loser is the scoring vendor – in a world of Big Data, the off-the-shelf credit score may be rendered obsolete if it fails to show the predictive capabilities that Big Data can. Scoring vendors do have an opportunity to capture the market for creditworthiness assessments in Big Data. Today, lenders use the FICO score as a starting point. The continued dominance of the FICO score as the industry's first-choice method of credit decision-making depends upon scoring vendors' ability to evolve the metric and keep it more predictive than alternative methods that banks may develop in-house.

Another is the underbanked groups themselves – existing subprime lending companies (e.g., Oportun) rely on extensive application processes that can force people with limited credit history to sacrifice a greater amount of privacy in exchange for their credit access than more privileged populations do. Especially if their data is housed without adequate precautionary security, personal-information-extractive Big Data programs may lead to serious equity concerns.

[17] Any references to Avant originations of unsecured loans in the United States relate to issuance of loans via the Avant platform by their partner bank, WebBank.

A third mechanism, and much discussed, is algorithmic justice, the theory of ensuring that the algorithms we implement do not re-entrench humans' discriminatory selection biases. Amazon recently shut down a piece of artificial intelligence that was found to penalize women in recruiting settings.[18] Silicon Valley tech company Palantir was recently found to be implementing a crime-predicting algorithm that disproportionately targeted Black and Latino men in Los Angeles and New Orleans.[19] Bartlett et al of the UC Berkeley Haas School of Business found that historically underbanked minorities pay 5.6 to 8.6 basis points more than their white and Asian counterparts on home purchase loans.[20] In credit, algorithms may similarly discriminate against marginalized groups based on previous trends, risking the establishment of self-fulfilling prophecies that perpetuate race-based differences in access to credit.

To balance corporate and public interests, we suggest that policymakers follow these general principles: Prepare to govern, Encourage and Enhance Innovation, Monitor the Algorithms, and Protect Privacy. Preparing to govern entails upgrading regulatory workforces to the same level of data literacy as the private sector's standard. Protecting innovation requires close collaboration with the private sector to craft policy compatible with ongoing research and development. Monitoring the algorithms will require recruiting private sector operators to track the results of algorithms and ensure their performance even in unusual economic times. Protecting privacy will require policymakers to continue to pursue existing standards and weigh algorithmic performance improvements against their associated privacy tradeoffs. These principles underpin the policy recommendations we propose, and should serve as a more stable guide for policymakers as the industry evolves.

In the short to medium term, we recommend that the Federal Trade Commission (FTC) and Consumer Financial Protection Bureau (CFPB) continue to increase the depth of their collaboration. While the FTC has taken a leadership role in holding hearings with the private sector and starting to educate consumers, the CFPB is poised to play a vital role in monitoring credit bureaus and lending institutions that apply new Big Data methods, and modifying and enforcing the law when appropriate situations arise.

Defining Big Data

As with many technology revolutions, Big Data is an evolution of previously emergent statistical management techniques. While the generally accepted understanding of the term Big Data evokes connotations of massive datasets, increased precision can show the unique structures and methods that allow Big Data to have force-multiplying effects on enterprise operations. For the purposes of this paper, we define Big Data in the sections below, both in terms of the technology in general and its applications in consumer lending.

---

[18] Adam Smith, "Amazon Shut Down Recruiting AI for Sexist Results | News & Opinion | PCMag.Com," October 10, 2018, https://www.pcmag.com/news/364315/amazon-shut-down-recruiting-ai-for-sexist-results.
[19] Dan Robitzski, "The LAPD's Terrifying Palantir-Powered Policing Algorithm Was Just Uncovered and Yes It's Basically 'Minority Report' – Futurism," News, Futurism, May 10, 2018, https://futurism.com/lapd-documents-show-their-policing-algorithms-continue-to-target-minorities-and-past-offenders.
[20] Robert Bartlett et al., "Consumer-Lending Discrimination in the Era of FinTech," October 2018, 44.

Big Data as a Technology

Various definitions of the term Big Data have been proposed by academics, technology groups, and business groups. The general concept of Big Data is simple – the field uses new tools to extract insight from unprecedented amounts of data. Practically, however, drawing the line between these new techniques and the ubiquitous practice of data analysis via familiar methods such as Microsoft Excel is useful to help laypeople understand the true disruptive capacity of the technology. An early definition came from Apache Software Foundation, developers of the open-source Hadoop platform of software for processing large datasets: "datasets which could not be captured, managed, and processed by general computers within an acceptable scope."[21] The Boston Consulting Group proposed a three-pronged definition in 2013:

> "Big data describes the large amounts and varieties of fast-moving information that can be processed and analyzed to create significant value. We'll address three of the key characteristics first:
>
> Volume. Data that have grown to an immense size, prohibiting analysis with traditional tools
>
> Variety. Multiple formats of structured and unstructured data—such as social-media posts, location data from mobile devices, call center recordings, and sensor updates—that require fresh approaches to collection, storage, and management
>
> Velocity. Data that need to be processed in real or near-real time in order to be of greatest value, such as instantly providing a coupon to customers standing in the cereal aisle based on their past cereal purchases."[22]

The additional dimensions of the Boston Consulting Group's definition help provide more specificity into the activities the technology requires of businesses looking to apply Big Data. It acknowledges that the "acceptable scope" delineated by the Apache Foundation includes the volume, variety, and velocity of the data. Further, it acknowledges that, consistent with Moore's Law, the amount of data that both traditional methods and more advanced systems can handle will likely continue to grow. The diagram below describes growth trends in all three categories:

[21] Chen, Mao, and Liu, "Big Data."
[22] Rob Souza et al., "How to Get Started with Big Data," The Boston Consulting Group, May 29, 2013, https://www.bcg.com/publications/2013/advanced-analytics-how-get-started-big-data.aspx.

EXHIBIT 2 | The Volume, Variety, and Velocity of Data Are Expanding Exponentially

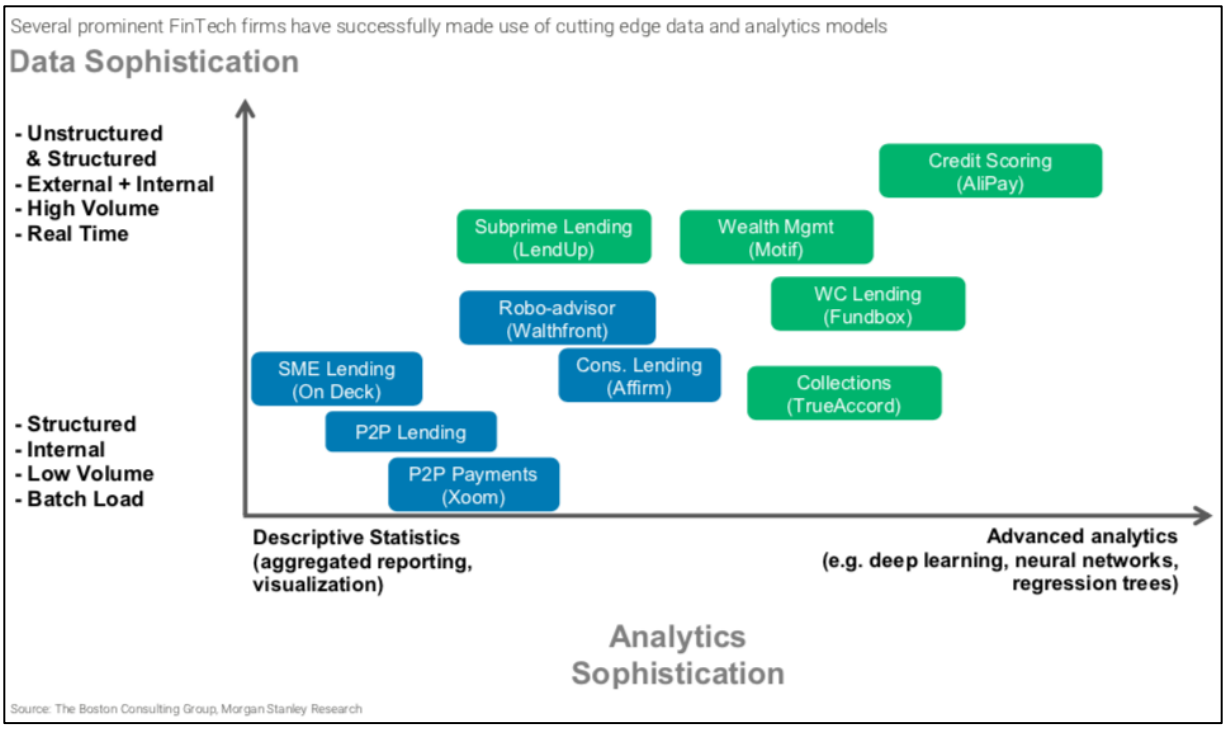
Data volume	**Data volume increases every day at a rate of 2.5 exabytes, or 2.5 billion gigabytes** • The volume of Wal-Mart's hourly transaction data is 67 times that of what is contained in the Library of Congress • Smart meters in China alone will generate 13.3 trillion readings in 2020 • Twelve terabytes of tweets are created each day
Data variety	**Content comes from more-diverse sources and formats** • Data can come from smartphones, sensors, RFIDs, GPS devices, transactions, and more • Data can be in the form of databases, video, speech, social-media posts, links, tweets, text messages, and more
Data velocity	**Faster analytics unlock value and reduce risk** • Organizations are analyzing millions of trading events per day to identify securities fraud • Organizations are analyzing hundreds of video feeds in real time to identify security threats

Source: BCG analysis.

[23]

We propose two criteria that must be met for an activity to be considered in the realm of Big Data for this paper's purposes: the dataset must be large enough that new tools and equipment are required, as well as new skills for the company. At the terabyte and petabyte scale, processing the data on personal computers with tools like Microsoft Excel is no longer an option. For these applications, companies may need to invest in cloud storage, on-premise data lakes, or third-party storage services. Accessing this data also requires new tools that abstract the data itself from the process of performing the analysis (anyone experienced in Excel modeling has experienced the frustration of waiting for the program to rearrange the entire dataset every time it makes a simple edit). In these data-rich environments, analysts will need to interact via scripting tools such as Python, SQL, and graphical user interfaces that execute pre-designed queries, either asynchronously or in real time.

[23] Souza et al.



[24]

The second criterion that sets Big Data apart from traditional data analysis is the need for a lender to change its organization. In conversation, Ash Gupta, former risk leader at American Express, suggested three key mechanisms: hiring the required talent, developing a partnership ecosystem, and adopting a test-and-learn culture.

> **Hire required skills.** While 85% of a data-literate workforce might be capable of adapting to new data analytics tools, 15% needs to be hired from environments that teach the latest statistical methods for machine learning and artificial intelligence – leading universities, coding boot camps, and pioneering companies in the technology sector.[25] Over time, improvements in the software ecosystems for managing Big Data may lower the technical threshold needed to add value in the space, and the need to hire new teams may diminish. However, in today's talent environment, the need to access talent trained in the latest statistical methods is a key factor in distinguishing Big Data activities from traditional analysis.
>
> **Develop partnerships.** The value chain for Big Data is too distributed for any one group to develop a customer facing solution end-to-end. To implement such a solution requires access to data, the hardware to store and process large amounts of information, the analytic software and skills to generate insight, and the business processes to moderate market-facing activities reflecting the system. The ecosystem contains independent

[24] Faucette et al., "Data Analytics for FIs: The Journey from Insight to Value."
[25] Ash Gupta (Former chief risk officer, American Express) interviewed by Marshall Lux and Guillaume Delepine, October 2018.

entities providing each of these services, and for success, companies seeking to provide Big Data services in a timely and efficient manner need to understand that they cannot do so alone.

Adopt a test-and-learn culture. As companies innovate in this space, they may be pulled into the faster cycles of software development, where the popular Agile methodology is comprised of a series of "sprints." In successful development cultures, failing fast is a key performance indicator, and lenders developing Big Data solutions to consumer lending need to be prepared to update their course in an iterative fashion. Ash Gupta points out that the guiding light in each adjustment must be a deep respect for the consumer and individual privacy, as judged by both law and good common sense, remembering that the purpose of these products is to serve people better.[26]

We note that defining Big Data around the needs to buy new equipment or upgrade talent is inherently a moving target. As the equipment and tools used to manage vast quantities of data improve, they will become more capable of scaling to meet organizations' expanding data needs. Further, as processing software improves, it will become more usable, and it will be easier for data-literate employees to upgrade to manage larger datasets. In today's environment, companies that claim to be working on integrating Big Data into their business operations have seen the need to upgrade both their equipment and workforce to be able to carry out their new activities.

Big Data as a Credit Assessment Tool

Credit providers have used data-driven techniques for years to inform their assessments of risk. While credit models in the past emphasized specificity and focus, new techniques are distinguishing in the way that they require new tools across the Big Data lifecycle as well as new methods of analysis. Over the course of development, it is clear that the most successful enterprises collaborate extensively with ecosystem partners, and have adopted a test-and-learn culture that guides their Big Data activities.

The Big Data value chain consists of collection, management, analysis, and action. Traditional methods used on-premise databases from such providers as Microsoft, Oracle, and SAP. These databases collect data from customers and store it in servers, where they are managed by internal teams and compared to other incoming datasets from credit bureaus. Analysis can be performed using traditional querying tools such as MySQL, and turned into analysis and insight in Microsoft Excel and PowerPoint. To move to Big Data, companies need to acquire data lakes – either by managing their own datacenters or outsourcing the capabilities to companies such as Amazon Web Services. These lakes are capable of handling vast quantities of data, but need to be accessed via querying software that requires specific training. Upgrades are required to support Big Data activities across the value chain.

[26] Ash Gupta (Former chief risk officer, American Express) interviewed by Marshall Lux and Guillaume Delepine, October 2018.

Collection: Datasets are growing as credit cards become more ubiquitous – credit card infrastructure operator Visa's network serves 3.5 billion issued cards, that make over 180 billion transactions a year at 50 million merchant endpoints.[27] Lending platforms such as FairSquare Financial and Avant also purport to buy 3rd party data from companies such as Argus and Experian to understand their applicants in ways that traditional credit bureau analyses do not consider.[28,29] Greater volume and diversity of datasets are distinguishing features of Big Data solutions and have forced companies to hire software engineers to design management systems.

Management: Managing so many new sources of data at high volume presents unique challenges to lenders. HSBC's chief architect, David Knott, points out that 55% of the company's $5.2b technology budget is dedicated to adding new systems, rather than simply maintaining existing ones.[30] This budget applies to building management systems that are capable not only of holding large quantities of data, but organizing it in ways that are transparent. HSBC is subject to European data warehousing laws and has to ensure it can trace the path of any data that is used in the company. Managing this metadata at scale is a challenge unique to Big Data, as automated collection and management now takes place at a global scale within the organization.

Analysis: Analysis of Big Data can take two forms: static and dynamic. Static analyses are run on a case-by-case basis, like a report in today's companies. These types of analyses may require advanced data analytics capability, and may use machine learning techniques such as logistic or multiple regression, which help to assess trends across multiple dimensions in large datasets efficiently. Dynamic analysis consists of standalone software that performs continual analysis over the course of its lifetime. These systems integrate directly into the company's data architecture and can automate decision-making or perform automated reporting in ways that become more intelligent over time. These dynamic analyses often earn the moniker of artificial intelligence, as their analysis is performed absent a human operator. Companies we interviewed often start by performing static analyses as a validation of newer models – Oportun's development team claims to perform this static to dynamic design process in as few as 25 days.[31]

Action: The applications of Big Data are the subject of later sections of this paper. For the purposes of defining the actions from Big Data, firms can either act on the insights from Big Data or put Big Data in place to take action itself. Machine learning is differentiated from traditional methods of analysis in its capacity to make predictions via methods such as regression and random forests with more confidence than traditional methods of extrapolation from historical data.

[27] Melissa McSherry (Senior vice president of product, Visa), interviewed by Guillaume Delepine, November 2018.
[28] Bhaskar Dole (Head of Analytics, Fair Square Financial), interviewed by Marshall Lux and Guillaume Delepine, October 2018.
[29] Suk Shah (CEO, Avant), interviewed by Marshall Lux and Guillaume Delepine, October 2018.
[30] David Knott (Chief Data Architect, HSBC), interviewed by Marshall Lux and Guillaume Delepine, October 2018.
[31] Raul Vazquez (CEO, Oportun), interviewed by Marshall Lux and Guillaume Delepine, October 2018.

Across the various facets of the Big Data value chain, there are two dimensions by which we categorized players on the spectrum from traditional data analysis methods to Big Data methods: techniques and datasets. The techniques category refers to the connections companies make and the algorithms they use. The datasets category is a measure of the companies' ability to integrate non-traditional sources of information (defined as information unavailable to credit bureaus) on the problem into tools that generate insight. The data and analytics sophistication diagram presented in the "Big Data as a Technology" section proposes a model for assessing Big Data sophistication among the companies we interviewed. The leading companies engage in a combination of importing new sources of information into their existing datasets and analyzing their combined datasets with more advanced statistical methods.

Investments in Big Data

Big Data is a prime area for investment from financial institutions, credit card providers, and new entrants to consumer lending. Market research firm Research and Markets projects that the financial services industry will invest $9 billion in Big Data in 2018 alone, with that figure projected to grow at 17% compound annual growth rate (CAGR) for the next three years.[32] The investments require both capital investments and human investments – the new equipment and tools to process vast datasets, and the trained staff to operate them and convert data to insight and action.

The motivation for further investment can be found in a cursory analysis of a dataset provided by Argus Information. We used the dataset provided by their team to make comparisons between credit card accounts for consumers with FICO scores lower than 680 vs. the general population. From Q1 2007 to Q3 2016, consumers in the low FICO score range generated 2.02x as many dollars of charge-offs, and 1.4x as many account attritions as the general population.[33] That finding is consistent with two investment hypotheses for Big Data. First, that companies are struggling to identify creditworthy applicants with low FICO scores, and that applying Big Data to refine the credit pool in this FICO band could reduce losses. Second, companies are struggling to connect with their cardholders from these traditionally underserved FICO bands – using Big Data to personalize the consumer experience may help keep cardholders in the financial system longer. The promise of monetizing these mechanisms and others has prompted investment across the Big Data value chain.

Capital Investments

Capital investments can take the form of in-house investment by financial institutions, or equity investments in teams building out the next generation of data techniques. In our assessment, we consider capital investments to include operating costs associated with subscription services such as Amazon Web Services. Often, as in many emerging technologies, these investments are based on collaborative arrangements.

Consumer credit providers have made investments in the infrastructure required to carry out Big Data activities. American Express, for example, uses a Hadoop platform licensed from

[32] Wood, "Big Data Investments in the Financial Services Industry."
[33] Argus Information. (2019). Bespoke Harvard Study Credit Dataset (Version 1). Excel.

MapR Technologies starting in 2014.[34] Our interviewees suggested that both the total technology budget and the portion dedicated to new system development – 55% in HSBC's case – will continue to grow in the coming years.[35] Continuing investment in enabling technology is likely to depend upon the rate at which new technologies reach the market, the scalability and adaptability of existing solutions, and the extent to which Big Data becomes a business-critical function for lending organizations.

The market for equity positions in Big Data for consumer lending has also been accelerating, with new entrants based on Big Data drawing substantial venture investment, and multiple high-profile acquisitions in the space. According to startup data aggregator Crunchbase, Fair Square Financial has raised $300 million in two rounds, Oportun has raised $250.3 million over 9 rounds, and Avant has raised over $659m in 6 funding rounds.[36,37,38] As startups continue to proliferate, the fact that the three movers interviewed for this study have raised over $2 billion over the course of 20 fundraising rounds suggests that capital is easy to access for companies looking to disrupt consumer lending with Big Data.

Further, acquisition has been demonstrated as a pathway for companies to enter the Big Data arena. Mastercard has completed multiple high-profile acquisitions that are designed to enhance its Big Data capabilities. Below we summarize Mastercard's Big Data-related acquisitions including passages from the parent company's official press releases.

> **5One Consulting. (2014, undisclosed amount).** "ShopperCentre, 5one's proprietary software, helps optimize insights drawn from their clients' data. This solution, combined with trends from MasterCard Advisors' Information Solutions business, will help deliver retailers a unique set of insights and decision-making resources."[39]
>
> **Applied Predictive Technologies (2015, $600 million).** "This acquisition advances MasterCard's agenda to deliver differentiated services to merchants, complementing its recent acquisition of 5One, a London-based retail consulting and analytics firm. As part of the company's services portfolio, Applied Predictive Technologies will now have access to MasterCard's analytics suite, consulting capabilities, marketing services and global footprint that will expand its reach and value."[40]
>
> **Brighterion (2017, undisclosed amount).** "Artificial intelligence plays a critical role in enabling consumer convenience, while delivering enhanced security. This advanced

[34] Alex Woodie, "How Credit Card Companies Are Evolving with Big Data," Datanami, May 3, 2016, https://www.datanami.com/2016/05/03/credit-card-companies-evolving-big-data/.
[35] David Knott.
[36] Crunchbase. "Fair Square Financial." *Crunchbase.com.* https://www.crunchbase.com/organization/fair-square-financial (Accessed: 3 January 2019).
[37] Crunchbase. "Oportun." *Crunchbase.com.* https://www.crunchbase.com/organization/progreso-financiero/funding_rounds/funding_rounds_list (Accessed: 3 January 2019).
[38] Crunchbase. "Avant." *Crunchbase.com.* https://www.crunchbase.com/organization/avant-credit/funding_rounds/funding_rounds_list (Accessed: 3 January 2019).
[39] "MasterCard Acquires 5one Marketing Limited," MasterCard Social Newsroom, accessed December 20, 2018, https://newsroom.mastercard.com/press-releases/mastercard-acquires-5one-marketing-limited/.
[40] "MasterCard Announces Acquisition of Applied Predictive Technologies," MasterCard Social Newsroom, accessed December 20, 2018, https://newsroom.mastercard.com/press-releases/mastercard-announces-acquisition-of-applied-predictive-technologies/.

technology delivers greater insights from every transaction to assist in making even more accurate fraud decisions… Brighterion's unique Smart Agent technology will be added to Mastercard's advanced suite of security products already using artificial intelligence."[41]

NuData Security (2017, undisclosed amount). "The addition of NuData will build on our layered security strategy to safeguard each and every transaction across the globe. The combination of session and biometric information will provide even richer context around potential cyber and device-specific threats, enabling us to deliver even greater trust and peace of mind."[42]

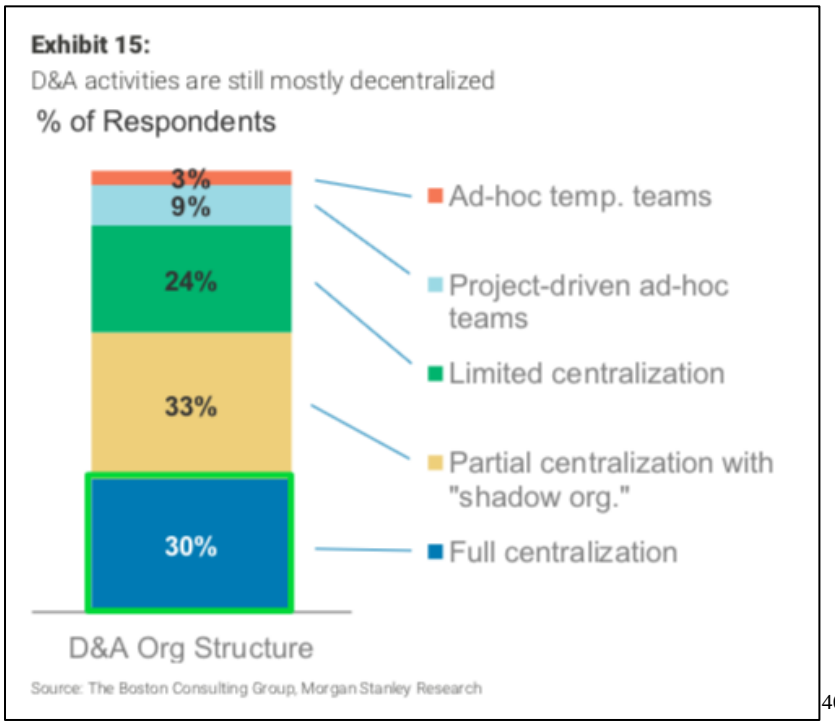
Mastercard's acquisitions have been integrated into the parent's product offerings, leveraging the enormous amounts of transactional information in Mastercard's proprietary database. As companies with access to large amounts of data look to expand their capabilities along the "methods" dimension of Big Data, acquisition may present an attractive entry strategy in the coming cycles.

Human Investments

The second area in which consumer loan providers need to invest to become Big Data leaders is human capital. Industry veteran Ash Gupta suggests that 85% of staff can be retrained in the latest data techniques, but that it is critical to seed around 15% of any Big Data team with people with advanced training. This smaller category of staff can be populated by graduates of high-profile institutions or veterans of industry-leading data science companies (e.g., Google).[43] Accessing technical talent is a well-documented challenge, especially in non-technology markets, but nonetheless the industry leaders all boast efficient teams with high levels of expertise.

The Boston Consulting Group and Morgan Stanley Research suggest that large firms have an advantage in building teams quickly due to their existing recruiting and funding resources.[44] In fact, the advantage may come from the ability to identify internal talent and assemble it into decentralized Big Data networks consisting of ad-hoc temp teams, project-driven ad-hoc teams, or formally centralized Big Data teams. The report's findings from a broad survey of financial institutions are summarized in the table below:[45]

[41] "Mastercard Enhances Artificial Intelligence Capability with the Acquisition of Brighterion, Inc.," MasterCard Social Newsroom, accessed December 20, 2018, https://newsroom.mastercard.com/press-releases/mastercard-enhances-artificial-intelligence-capability-with-the-acquisition-of-brighterion-inc-2/.
[42] "Mastercard Acquires NuData Security Inc: Fraud Prevention," NuData Security, March 29, 2017, https://nudatasecurity.com/press/mastercard-enhances-security-internet-things-acquisition-nudata-security-inc/.
[43] Ash Gupta (Former chief risk officer, American Express) interviewed by Marshall Lux and Guillaume Delepine, October 2018.
[44] Faucette et al., "Data Analytics for FIs: The Journey from Insight to Value."
[45] Faucette et al.



[46]

The form that human investments take can vary firm to firm, and may as often reflect an organization's internal power structures as an unbiased framework for team design. The nature of Big Data teams will also depend on the level of business criticality that the Big Data services carry out for their organizations. Below we present a small set of anecdotes from around the industry on the structure of Big Data teams:

American Express. Although headquartered in New York City, American Express maintains big data facilities in both Bangalore, India and Palo Alto, CA. In Bangalore, the company has assembled a team of 300 data scientists from around the world.[47] From the American Express press release announcing the establishment of the Palo Alto site: "Over time, American Express expects to employ some 200 individuals in the Palo Alto hub. The company chose the location, in part, to cultivate the technology talent building their skills locally at technology firms, in the startup community and at area universities."[48]

Avant. The company employs over 400 employees across product, analytics, data, and technology. 50-100 work in product analytics, assessing product risk, credit cards, and loan products. Another 30-40 work on the business intelligence data infrastructure team. Approximately 100 work in specific industry verticals to perform analysis. Finally, around

[46] Faucette et al.
[47] Ash Gupta (Former chief risk officer, American Express) interviewed by Marshall Lux and Guillaume Delepine, October 2018.
[48] "American Express Creates New Technology Hub in Silicon Valley | American Express," accessed December 20, 2018, https://about.americanexpress.com/press-release/american-express-creates-new-technology-hub-silicon-valley.

100 software developers spend their time developing new algorithms, generally in Ruby on Rails.[49]

HSBC. The chief data officer runs a separate organization focused on data policy (how data is classified and curated) and shared data services (data engineering, warehousing infrastructure, and cloud adoption). HSBC has a core team of 120 dedicated senior technology strategists and architects, out of a distributed practice of approximately 1,200 staff doing project and product level architecture.[50]

Oportun. Oportun's data analytics teams are divided into core engineering (reports to chief technology officer), data and analytics (reports to chief risk officer), and mobile (a new group projected to triple in FY19). Across all areas, data teams number between 200 and 250 people, comprising around half of Oportun's back-office organization.

Applications of Big Data

The potential of Big Data is still unfolding across the personal lending landscape. While investment varies across like organizations, further differences in Big Data capabilities arise from various stages of the lending value chain's access to data. For example, while community banks have direct customer interactions, major players do not see the customer in person. Further, credit card network operators such as Visa are the stewards of cross-bank transactional data, but have limited visibility into application data.[51] Scale is insufficient to determine winners and losers, rather the winners will be the firms who are most adaptive to new technologies. Still, it is possible to taxonomize the applications of Big Data in consumer lending into three broad categories: anti-fraud, customer engagement, and creditworthiness decisions.

Risk Management

Fraud is a significant profitability factor in the consumer lending industry. Experian's 2018 Global Fraud Report says that from consumers to providers, fraud is a great concern – four of five consumers trust that businesses are making fraud prevention a top priority, and on the supply side, 72% of business executives expressed a growing concern about fraud over the last 12 months.[52] Further, 75% of businesses said that they were "very interested in more advanced security measures and authentication,"[53] indicating that there is market demand for advanced anti-fraud capabilities. These capabilities are likely to come from the IT organization, which 57% of companies claimed to hold directly responsible for fighting fraud.[54] IT organizations also play a critical role in establishing Big Data programs, and may play a critical role in averting fraud going forward.

---

[49] Suk Shah.
[50] David Knott.
[51] Melissa McSherry.
[52] Experian Information Solutions, "Experian Global Fraud Report 2018," Experian.com, 1/18, https://www.experian.com/assets/decision-analytics/reports/global-fraud-report-2018.pdf.
[53] Experian Information Solutions.
[54] Experian Information Solutions.

Big Data implementations bring two new primary advantages to the fight against fraud: immediacy and scale. Because Big Data requires the rapid ingestion, collection, and analysis of data, many systems are fully automated. HSBC, for example, automates the transfer of data between internal datacenters to comply with data residency regulations in Europe – abdicating to computers the job of transferring and tagging data.[55] The immediate nature of Big Data implementations helps identify fraud at the moment of attack, enabling real-time interventions. The scale of Big Data also enables automated fraud detection at levels of precision greater than humans are capable of. While humans can intuit a baseline of data trends and detect anomalies to the extent they are capable of surveying data in real time, automated Big Data systems can use machine learning to identify anomalies along dimensions that are not key focus areas for fraud monitoring staff. The Boston Consulting Group summarizes the way Big Data brings new tools to the fight against fraud:

> "Banks and fintech companies use data learnings for effectively underwriting credit risk, while payments processing companies have long used transaction data to assess fraud risk through use of predictive analytics on a wide range of factors such as past purchasing behavior, purchase location, IP address detection, etc. The use of sophisticated predictive analytics has helped the payment networks consistently minimize fraud while limiting user authentication requirements, thereby reducing friction in the consumer experience."[56]

The new advantages Big Data brings to the fight against fraud enhance lenders' and regulators' ability to fight fraud on two fronts: application fraud and transaction fraud.

Application fraud. Big Data is helping companies systematically identify false applications submitted by consumers with no intention of paying back their principal. At Oportun, one driver of fraud is the creation of false accounts, that are used to take out loan. With Big Data monitoring, Oportun has developed techniques to identify these actors to the extent that CEO Raul Vazquez points out that fraud rates are no longer the primary concern for further algorithm development. With the maturity of Big Data tools to fight application fraud, Oportun has been able to shift resources toward other profit-driving activities.

Avant is another player using Big Data to prevent application fraud. In addition to algorithmic monitoring, Avant incorporates non-traditional data sources into its risk algorithms. CFO Suk Shah mentioned a few key sources of information that inform Avant's fraud detection tools:

- ***Ownership authentication:*** Sourced from LexisNexis, these datasets can be triangulated against an applicant's stated asset ownership (e.g., did you own a 1992 purple sedan?) to verify the applicant's identity.
- ***Work number:*** These datasets can validate whether the applicant truly worked for an employer that was listed on the application to further support the identity verification process.

[55] David Knott.
[56] Faucette et al., "Data Analytics for FIs: The Journey from Insight to Value."

- ***Email age:*** The age of an email account has been shown to be predictive of fraud. For example, an email created six days prior to an application has an increased likelihood of having been instantiated for fraudulent purposes.
- ***Telesign:*** The history of a telephone number can help determine fraud in similar ways to email age verification.

Avant's algorithms achieve measurable results in mitigating non-right-party-contact (non-RPC) fraud. While the industry loses around 60 basis points to fraud per year, Avant holds fraud to around 30 basis points. This success rate has attracted substantial attention to Avant from financial institutions looking to white-label the fraud and servicing capabilities. The company now has signed partnerships with Regions, Banco Popular, Honeycomb PLC, and HSBC through its technology platform: Amount.[57] As in all industries, the methods of fraud are constantly evolving, but for the moment, Big Data has swung the advantage in the battle toward the lending platforms.

Transaction fraud. The algorithms combating transaction fraud are still developing across the financial services industry, and are being developed by both credit card networks and banks. Visa makes fraud mitigation a priority, seeing the $600m annual opportunity that exists in mitigating fraud worldwide.[58] HSBC looks beyond its own profits to detect the broader set of financial crimes carried out over credit networks, including money laundering, and work with the authorities to intervene.[59] Scott Bauguess of the SEC Division of Economic and Risk Analysis points out that the present state Big Data systems can detect insider trading algorithmically, but still require human intervention to definitively identify credit fraud.[60]

Visa's challenge with transaction fraud is identifying attacks in real time, which are often carried out in parallel by distributed crime networks, with losses accumulating quickly if the schemes proceed unchecked. Visa senior vice president of product Melissa McSherry gave a scenario of international cash-outs wherein an international organization would mobilize agents in multiple countries to make cash withdrawals from automated teller machines (ATMs) simultaneously, using a bulk set of counterfeit card numbers. In these attacks, losses can reach $250k per minute that the fraud continues unchecked.

Previous methods to mitigate these attacks included implementing chip cards, including photography and facial recognition at the teller machines, and increased encryption sophistication to protect card numbers from compromise by fraudulent actors. The new generation of techniques is enabled by Big Data, and Visa builds models to predict whether transactions are likely to be real or fraudulent based on the owner of the card. These models need to react quickly – when a charge reaches Visa's network, the company needs to scan the transaction for fraud in addition to the complex process of soliciting credit approval from the bank. Big Data paired with automation at scale allows Visa to avert fraud in real time.

Visa's programs have encountered success in mitigating transaction fraud, but additional challenges loom. Visa's sophisticated methods of thwarting transaction fraud have discouraged

[57] Suk Shah.
[58] Melissa McSherry.
[59] David Knott.
[60] Scott Bauguess (SEC Division of Economic and Risk Analysis) interviewed by Marshall Lux and Guillaume Delepine, January 2019.

financial criminals from even attempting counterfeit schemes. As a result, application fraud is doubling year-over-year as scammers seek the path of least resistance for theft. Application fraud, however, lies in the hands of issuers, so Visa may be increasing efforts to enable banks with their advanced mitigation tools in the near future.

Transaction fraud may be complicated further by the number of Big Data-enabled services that are now available to consumers. In Europe, new CSV2 technology enables increasing connections to consumers' bank accounts, but the additional exposure increases the attack surface for fraudsters to exploit vulnerable connections for financial gain. Big Data will need to work in tandem with rigorous authentication and authorization programs to monitor these additional connections and avert financial crime.

HSBC algorithmically monitors its large datasets to identify and report money laundering and other suspicious activities. Monitoring these datasets often requires inspecting sensitive information, so the datasets are typically stored on-premise, rather than in cloud databases. As automation takes over the process, monitoring for fraud is making a transition from a rules-based activity to a machine learning activity, where anomalies can be detected even if a human did not explicitly instruct the machine to look for a particular trend. The challenge with moving to the new approach is that it comes with a high false positive rate. At present, the algorithms flag approximately 100 incidents for every detected financial crime, leading human-staffed investigative teams to allocate a large share of their resources toward incidents that do not lead to any arrests.

As fraud techniques continue to evolve, Big Data will have to evolve as well in order to remain a relevant tool in thwarting financial crimes. Adding new algorithmic techniques, enhancing the speed of fraud identification and automated response, and incorporating new datasets are all levers that might help Big Data keep financial institutions one step ahead of fraudsters. Achieving improvements in these areas will require, and likely continue to draw, substantial investment from major industry players.

Servicing

Customer engagement is another primary motivating factor for consumer lenders to leverage Big Data. Instead of improving profitability at a unit or overall level, improving customer acquisition costs and retention rates are ways for lenders to capture and defend market share. The Boston Consulting Group summarizes the motivation for enhancing customer engagement in their collaborative paper with Morgan Stanley:

> "**Beyond the hard dollar savings, we think market share shift potential is the bigger carrot/stick:** As FIs face increasing threats from disruptive innovators and from each other, we think those that are able to effectively use data for improving internal decision making as well as driving better consumer engagement will be better positioned to not only preserve, but potentially accumulate market share from others. One of the most relevant avenues is through customer loyalty. Acquiring customers can be expensive but FIs often lose them frequently when competitors offer lower priced products. Making broad-based pricing cuts is often not a solution given profit ramification across the portfolio. Data can be an effective

tool to understand the customer's motivations and help improve retention by either offering targeted pricing changes or value added services which customer may value more than a few dollars of savings."[61]

Big Data can enhance the customer experience in a variety of ways, which all serve the goal of improving customer retention. Citigroup's head of operations and technology Don Callahan commented in 2013 that "We are focused on predictive analytics, learning not only people's needs but what they are likely to want in the future–so we can be ready with the right products and services."[62] Various datasets enable differing types of insights on the consumer that can be profitable in the short-term, and differentiating for the lender in the long-term. Below we outline several of the ways that companies are using Big Data to develop a deeper understanding of their consumers:

Retail Offers. American Express' program, Amex Offers, detects the consumer's location and pushes offers to them in real time. One use case is in making targeted restaurant recommendations based on observations of a given consumer's spending patterns and preferences.[63] CapitalOne offers a similar, coupon-based solution called Mobile Deals.[64]

Customer Service. CapitalOne uses historical call center data along with transactional data to predict the topic of customer service calls. Case studies suggest that the company can predict the topic of a support call within 100 milliseconds with 70% accuracy.[65]

Predictive Sales. CapitalOne was an early mover in data analytics, offering targeted credit pricing customer-by-customer as early as the 1990s. Today, CapitalOne uses Big Data to enhance its credit pricing and decision-making.[66]

First-in-wallet Analysis. Visa uses the power of its transaction database to help credit card issuers on its network understand their consumers' spending habits. In particular, Big Data can help an issuer understand if it is a consumer's primary card or is losing the competition for that wallet to a competitor.[67]

These mechanisms of deeper customer engagement based on applications of Big Data can each be differentiating for a card issuer. The insights may be generated at various points in the value chain based on which players have the greatest access to data or most developed teams. For example, while credit network operators like Visa and Mastercard will have the greatest access to transactional databases, issuers such as banks will have more access to consumer data, especially that which was provided on the application. As Big Data techniques become more ubiquitous, outsourced analytics services may become standard practice and a requirement to compete in the space. Meanwhile, proprietary research and development may help companies establish and protect competitive advantages in customer retention. Regardless of the method,

[61] Faucette et al., "Data Analytics for FIs: The Journey from Insight to Value."
[62] Woodie, "How Credit Card Companies Are Evolving with Big Data."
[63] Woodie.
[64] Woodie.
[65] Woodie.
[66] Woodie.
[67] Melissa McSherry.

gaining advantage in this area may prove critical to shaping the future competitive landscape of consumer lending.

Product Development

The ability to make credit decisions is one of the most attractive applications for Big Data in consumer lending, and carries the potential to change the face of social mobility in the United States while expanding the addressable market for credit card companies.

The opportunity in using Big Data for credit assessments encompasses both approval rates and credit offers. Big Data can help make better creditworthiness assessments faster. By either incorporating data in addition to what is available from credit bureaus, or assessing existing datasets using more advanced techniques, Big Data-enabled companies can make more targeted assessments of creditworthy applications. The impacts to profitability are two-fold: less high-risk debt will be issued to reduce losses, and more high-quality borrowers will receive credit to increase revenue. Reliable borrowers with thin files or low credit scores would otherwise be eliminated from the approval process, but with Big Data, the diamonds in the rough of lower credit bands can be served by the financial system.

In addition, credit decisions can be made much more quickly with Big Data. Credit decision data is often purchased on a "pull" basis – rather than owning large datasets, lenders buy information on applicants as needed. While these information requests introduce lag time, the technical requirements of Big Data suggest that incremental automation will fit seamlessly into existing systems. Providing faster credit decisions not only reduces organizational complexity for lenders, but streamlines the customer experience, decreases the time-to-revenue on a credit product, and avoids the churn caused by lengthy approval delays. As Big Data improves each credit decision, the speed at which it does so will dramatically improve lenders' ability to serve customers profitably.

The second decision pertaining to credit assessments is the size of the credit line to grant to an applicant. Credit limits on cards vary from the low hundreds to the hundreds of thousands based on credit scores. The high ends of credit cards are often in amounts the user would never spend, but as industry insider Ash Gupta points out, the goal of a high credit limit is to win the share of the consumer's wallet. The goal for a credit card company is to be the first card in each customer's wallet, rather than splitting the share of purchases with a competitor. The balance to strike is with avoiding a delinquency incident, since it would take a multitude of paying customers to amortize a default on a high credit line.[68] While Big Data can help identify customers who can be trusted with credit *at all*, it also can help providers decide *how much* credit to grant to each applicant in a way that optimizes for high spending and low default risk across the portfolio.

Capitalizing upon the promise of Big Data in making credit decisions is a different process based on the type of customers being targeted. As techniques become more and more advanced, the addressable market grows in lower and lower credit bands. In the market today, there are

[68] Ash Gupta (Former chief risk officer, American Express) interviewed by Marshall Lux and Guillaume Delepine, October 2018.

already lenders moving down the credit band stack to identify new potentially profitable customers.

Existing markets (credit scores 700+). HSBC uses Big Data to make more confident credit decisions in its existing markets. The bank emphasizes consumer trust and focuses on more advanced statistical methods using the same data, rather than incorporating more intrusive consumer datasets, even if it has technical consent via terms and conditions. HSBC has a stringent credit policy aimed at preventing money laundering.[69]

Middle income America (credit scores 550-700).[70] Fair Square Financial focuses on identifying viable prospects for pre-approved credit card offers in the 600-700 credit score range, with the median credit score for approved customers being around 660. Where traditional credit data is abundant, the company uses advanced algorithms to build credit risk models with more predictive power. Where traditional credit data is scarce (i.e. thin files), the company is testing the value of acquiring additional consumer data from services like LexisNexis and ID Analytics to inform lending decisions.[71]

Avant's mission is to provide responsible, reliable credit products to consumers who would traditionally have been underbanked. A key component of that approach is protecting the 40% of Americans with sub-700 FICO scores from predatory payday lenders. The company has facilitated $5.3 billion in loans in the United States, and runs a customer-friendly operation that offers qualified individuals access to credit within twenty-four hours. However, only 2-4% of applicants make it through the application process. In addition to using Big Data to detect fraud, Avant uses sophisticated algorithms and 3rd-party datasets to help its partners determine optimal credit outcomes. Though prohibited by the Fair Credit Reporting Act (FCRA) from using intrusive data like college transcripts, the Avant team uses approximately 800 variables in consumer datasets provided by various data providers to assess creditworthiness and affordability through its platform. In addition, bank account integration services like Yodlee and Plaid can automatically provide information on a consumer's direct deposits to validate reported income on credit applications before finalizing a lending decision.[72]

Thin files (no credit score). The thin files space is distinct from the credit rehabilitation space, for consumers who have had bad credit fidelity in the past. Silicon Valley-based Oportun focuses on providing small loans to borrowers without enough credit history to qualify for credit cards. 45% of loan recipients at Oportun did not previously have a credit score. The company compensates by asking customers for additional information at the time of application through questions such as: How long have you lived at your current address? What is your income? Do you own or lease your car? How long have you owned your cell phone? How long have you been at your current job? The answers to these questions feed Oportun's proprietary risk engine. A key component of Oportun's business model is ensuring their customers successfully repay their loan, so that oportun can provide their customers

[69] David Knott.
[70] This category contains credit bands generally referred to as Low Prime, Shallow Subprime, and Near Prime.
[71] Bhaskar Dole.
[72] Suk Shah.

multiple subsequent loans in larger amounts and lower interest rates. 50% of granted loans are to customers who have previously paid off a loan product with Oportun. These customers are trusted with larger loans than previously. As these customers demonstrate their capacity to manage debt, they can take out loans of up to $9,000, rather than sums as low as $300 for initial loans. These customers are likely to "graduate" to lower-priced loan options and different credit products, which Oportun does not currently offer. The company announced the acquisition of Springboard Auto, an auto lending company, in October 2018,[73] and is building out its own credit card offering.[74]

Across the credit band spectrum, advances in the quantity and diversity of data, along with more advanced processing techniques, are enabling lenders and lending platforms to make more confident credit decisions, control share of wallet, and extend credit services to previously underbanked populations. The potential for extending credit access to the underbanked in ways that do not compromise the profitability of the consumer lending sector is reason for excitement from policymakers and business leaders alike.

Growth Trends

While Big Data provides attractive opportunities for the consumer lending industry, its positive development is not inevitable. At play are macroeconomic factors, market forces, public policy, and the cyclic nature of technology investment. The prevailing trends in each of these areas will impact the extent to which the promise of Big Data is realized across the consumer lending industry.

Tailwinds

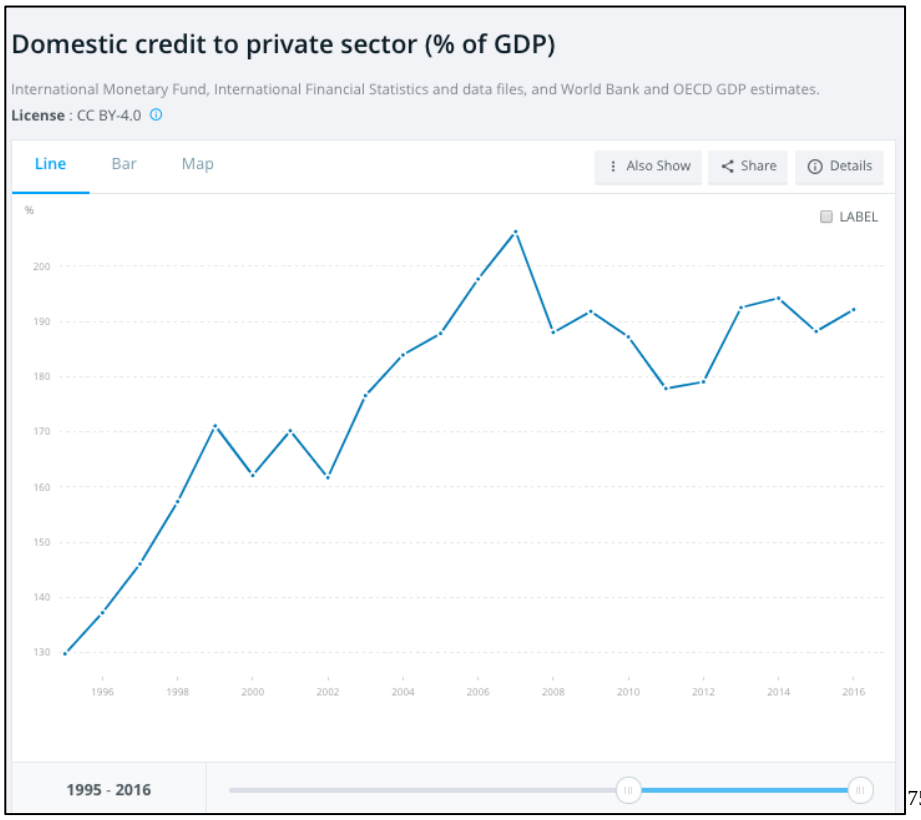
Macroeconomic factors and trends within the consumer lending market appear most likely to accelerate the rollout of Big Data in the short to medium term. While the future is not guaranteed, robust trends in these two areas suggest they will continue to fuel the growth of Big Data in the consumer lending market.

Macroeconomic factors. Continued economic growth and low unemployment underpin Big Data development by fueling the market for credit. With greater shares of the population gainfully employed, there are more qualified applicants for credit products, who can be identified more confidently using Big Data. In economic downturns, when algorithms have to be more restrictive, Big Data would most likely be used to *exclude* less creditworthy consumers, rather than *include* parts of the market that would not be served using existing methods.

The amount of credit disbursed to the market has historically trended upwards in the United States, but according to the World Bank's data, the amount of domestic credit provided to the private sector dropped after the 2008 financial crisis. In the next

[73] Michael Azzano, "Oportun Acquires Intellectual Property and Lending Platform of SpringboardAuto.Com," accessed December 20, 2018, https://globenewswire.com/news-release/2018/10/09/1618529/0/en/Oportun-Acquires-Intellectual-Property-and-Lending-Platform-of-SpringboardAuto-com.html.
[74] Raul Vazquez.

recession, it remains to be proven that algorithm-driven decision-making would react similarly. However, the current wave of economic growth would suggest increased lending in the near term, and the use of algorithms to accelerate the process.



[75]

Industry Trends. Two critical catalysts for future Big Data deployment are the availability of 3rd party data-enabling services, and the ability of consumer lending companies to find talent to perform the necessary analyses. Multiple interviewees cited the feeling of a burning platform that emphasizes that companies need to hurry to form partnerships and hire talent lest they lose market share to more agile players.

3rd party data services provide both tools and data for the tools to analyze – the strong growth of this sector is critical for the continued expansion of Big Data techniques in consumer lending. Companies that help ingest, warehouse, analyze, and curate consumer information will drive research and development into next-generation services that lenders cannot dedicate resources to design. 3rd party providers such as Argus may also serve as critical data pipelines, aggregating and anonymizing data from multiple providers so that customers can benefit from the analysis of broader datasets. From these

[75] The World Bank, "Domestic Credit to Private Sector (% of GDP) | Data," data.worldbank.org, December 26, 2018, https://data-worldbank-org.ezp-prod1.hul.harvard.edu/indicator/FS.AST.PRVT.GD.ZS?contextual=default&end=2016&locations=US&name_desc=true&start=1995&view=chart.

larger populations can be drawn better trained machines offering more precise insights to consumer lenders.

Accessing talent will equally drive the industry. As lenders and service providers transform their workforces to meet the changing demands of their data and analytics groups, skill sets will evolve from basic reporting (in tools such as Microsoft Excel) to more advanced algorithm design and deployment (in software such as R, Stata, and Python). While some parts of existing workforces can be retrained to work in these new environments, overall increases in data and analytics headcounts as well as the need for skills in newly invented areas will force lenders to hire externally. The industry will therefore benefit from a growing network of data science schools, which have already been shown to prepare talent for the rigors of machine learning careers. A key example is KPMG, which recruits from such groups, and was recently named a leader on the Forrester Wave for data analytics in professional services.[76]

We note that the presence of 3rd party data services and the ability to recruit sufficient talent are linked. In the present technology ecosystem, it is necessary to hire a data science team that includes expertise in data architecture, cloud implementations, cyber security, analytics processing, and risk model design. In the future, turnkey solutions by 3rd party providers may alleviate the burden on consumer lenders to hire talent that can manage the entire technology stack. With advancements to make technology more accessible, machine learning and artificial intelligence may become basic skills incorporated in broader roles at consumer lending companies, rather than independent career paths.

Consumer Trends. In the introduction we note that 16.7% of participants in a Consumer Financial Protection Bureau study decided not to apply because they thought they would be denied.[77] The number of credit applications received by a consumer lending firm is a raw input that drives the number of cards issued and overall amount of credit disbursed by the industry. With the advent of Big Data, there may lie substantial opportunity in convincing more consumers with non-traditionally attractive credit histories to apply for credit cards. Should it lead to increased net approvals, an increase in application volume would be a key driver in demonstrating the profitability use case for Big Data and promoting further investment by consumer lenders.

Headwinds

Despite the presence of multiple drivers of continued Big Data development, there remains the possibility that the deployment of the technology is a bubble that sees a deceleration in the coming cycles. Disruption to the tailwinds listed above, a shock event leading to industry

[76] Jens Rassloff, "KPMG Named a Leader in Microsoft Dynamics 365 Services by Forrester Research | KPMG | GLOBAL," KPMG, January 31, 2018, https://home.kpmg/xx/en/home/insights/2018/01/kpmg-named-a-leader-in-forrester-research.html.
[77] Board of Governors of the Federal Reserve System, "Report on the Economic Well-Being of U.S. Households in 2016 - May 2017."

disillusionment with the technology, or restrictive public policy could each play a role in undermining the Big Data revolution.

Technological disillusionment. We state in our introduction that Big Data fell off of the Gartner hype curve because it was *too ubiquitous.* If, however, the hype cycle for Big Data were to face more troubling disruption, there could be negative impacts on investment and adoption. Such a disruption may come from a major recession, or another edge case that "breaks" the algorithms (we note that the firms we interviewed take steps to prevent both situations). Without the enthusiasm from industry analysts, shareholders, and major consulting firms, executive suites of consumer lending firms may be less eager to invest in cutting-edge technologies if they have been shown to backfire in the past. Similarly, the venture capital required to fund startups – either upstart lending houses or intellectual property-based service providers – would be likely to erode. Without these two sources of investment, the expansion of Big Data applications would most likely slow down.

Restrictive public policy. While public policy already does and will continue to play a role in consumer lending, the way that it is applied may have significant impacts on the development of Big Data technologies. We provide further analysis of the frameworks and specific policies that should be applied in later sections, but in this section we point out the tradeoff between the deployment of Big Data and other social goals.

While Big Data can be revolutionary in expanding credit access to lower rungs of the economy, it also carries the risk that lenders and data providers violate consumers' privacy, then fail to protect their customers' confidential information sufficiently and expose data to hackers. Further, unethical or excessively risk-loving credit models may expose the economy to greater downside risk than traditional models.

Mitigating the potential for Big Data to harm consumers may require policy. In the European Union, data residency laws designed to protect consumers have complicated the handling of data for major lenders who aggregate datasets across borders to generate better insights. Compliance requires its own staff and internal corporate governance, which adds to the cost of deployment for Big Data systems.[78]

Europe's challenges with its General Data Protection Regulation (GDPR) demonstrate the risk that public policy impairs consumer lenders' deployment capabilities in the area of Big Data. But this is not the only way that regulation may impact companies' ability to deploy next-generation systems. As we highlight in later sections, collaborating with Big Data experts on the private sector side will be critical to ensuring that policy designed for other, valid social objectives does not impair the ability of the consumer lending sector to deliver on its promise of myriad social benefits.

Stakeholder Outcomes

[78] David Knott.

The Big Data revolution in consumer lending brings different opportunities and threats. Effective policymaking will depend upon a balanced understanding of the impacts faced by various players.

Lenders and Lending Platforms

The myriad benefits to lenders equipped with Big Data solutions have been summarized above as increased revenue, higher operational efficiency, decreased credit risk, and reduced transaction and application fraud. For lenders with advanced data and analytics teams, or who purchase a developed suite of services from 3rd party providers, the best outcome is for Big Data to continue its expansion around the industry with as little regulation as possible.

A risk to the lending industry, however, is the consolidation of insights among the large-scale lenders that have the excess profits to invest in developing or purchase Big Data solutions and incorporate them into their business operations. Smaller lenders, including community banks, may not have the means to compete against data-enabled industry behemoths. Marshall Lux has written previously on the struggles of community banks to compete with big banks since the passage of the Dodd-Frank Act.[79] Community banks' traditional advantage as lenders is their ability to meet face-to-face with the customer, gather information through social interactions that cannot be captured in an efficient application form. In the world of Big Data, the risk to community banks is that the large lenders can make decisions faster and more effectively using their algorithms. This would not only give large lenders an advantage in assessing risk, but also let the large banks steal share as they become the most convenient source of credit in downstream markets.

Community banks will by definition never have the volume of data that the large lenders can muster, and may resort to digging into their profit margins to buy third party datasets and services just to compete with the larger banks. Absent a cooperative cross-community-bank approach, however, the basic economics would suggest a trend toward consolidation wherein community and local lenders either join forces or go out of business. Community banks may also find themselves left with deeper subprime applicants and are forced to take on greater risk when lending, taking on a role increasingly similar to payday lenders, who make minimally informed decisions and compensate by raising the price of their credit. Further, as large lenders deploy Big Data solutions to root out fraud, non-Big Data enabled lenders may become perceived as softer targets, leading to losses both from fraud and the erosion of public trust in the safety of their lending services.

The path to continued relevance for the community banks is to actively pursue partnerships and make agility their competitive advantage. As smaller organizations, they may be able to innovate more quickly, enter into lower-risk partnerships with Big Data information providers and infrastructure partners, and enjoy a shorter test-and-learn cycle. To compete with the large banks, the community banks must leverage this agility to gain a level of Big Data scale that their own customer bases do not intrinsically provide.

Fintech Startup Ecosystem

[79] Marshall Lux and Robert Greene, "The State and Fate of Community Banking," February 2015, 32.

We present above numerous examples of the new investment in and applications for Big Data that have come out of the fintech startup ecosystem. Big Data has created opportunities for startups that have competitive moats from the beginning. Companies with new algorithms or risk models, such as Oportun,[80] have defensible intellectual property. Companies that serve near-prime customers, such as Fair Square Financial,[81] have an advantage in their ability to target an underserved population that other lenders prefer to avoid. Companies that are less hesitant to incorporate alternative datasets, can make decisions informed by more factors than companies that emphasize compliance and consumer privacy, like HSBC.[82] As with all technology waves, disruption is not without victims, and traditional companies may face the deleterious effects of competing against these new players.

US Economy

The impacts of the consumer credit market on the US economy are not fully understood, but a few basic principles can help inform predictions of how Big Data may impact the relationship. To understand the potential impact of Big Data, it is important to understand both components:

> **Consumer lending and the US economy.** The *Wall Street Journal* reported in 2017 that amidst a period of strong growth and low unemployment, credit card lenders were observing higher delinquency rates.[83] This finding would suggest that the relationship between economic performance and the credit card market is non-linear. Richard Fairbank, CEO of CapitalOne, remarked further that it is "not unusual for the credit cycle to diverge from the economic cycle."[84]
>
> Delinquencies can become problematic when paired with an economic downturn, when they can compound the financial burdens provided by unemployment and more restrictive lending overall. *MarketWatch* proposes a three-pillared approach to understanding how credit card usage can interact with declines in the overall economy. First, the financial sector's willingness to lend is a key driver of economic activity. Second, credit supply expansion increases household spending and thereby gross domestic product. Third, in harder economic times, credit dries up, further restricting spending and accelerating the decline in economic activity.[85]

[80] Raul Vazquez.
[81] Bhaskar Dole.
[82] David Knott.
[83] Aaron Back, "A Surprise Bump in Bad Card Loans - WSJ," September 20, 2017, https://www.wsj.com/articles/a-surprise-bump-in-bad-card-loans-1505899800.
[84] Hillary Hoffower, "Americans Are Using Their Credit Cards More than Ever — and Some Experts Think That's a Good Thing," Business Insider, accessed December 26, 2018, https://www.businessinsider.com/credit-card-use-increasing-in-us-economy-2018-5.
[85] Amir Sufi Mian Atif, "Credit Expansion Is the Real Cause of Booms and Busts," MarketWatch, accessed December 26, 2018, http://www.marketwatch.com/story/credit-expansion-is-the-real-cause-of-booms-and-busts-2018-03-08.

Algorithmic lending and credit's economic effects. Algorithms accelerate decision-making, and if they are incorporated into automated governance systems, may push decisions to scale faster than today's rules-based analytics can. Compared to rules-based systems, machine learning solutions may make room for more nuance, and avoid making large-scale decisions based on arbitrary triggers. However, this depends on redundant algorithm design. If willingness to lend as defined by algorithms declines immediately, the effects of a financial downturn may be felt more severely. If the impacts compound over time, algorithms may play a role in exacerbating and accelerating times of financial hardship.

It remains unclear whether regulators are well-positioned to regulate Big Data as it is used for lending decisions. The algorithms are proprietary and can be opaque to regulators, who may not have the level of exposure to Big Data systems as their industry counterparts. On the other hand, the entrepreneurial wave for Big Data in the fintech space has led to the funding of companies by founders who have not worked through a recession in the space. The transition of lending decision-making from industry veterans to technologists to algorithms may present a risk to the economy if not managed effectively.

Individual Consumers

There are three primary considerations for the individual consumer, who in the world of Big Data will be required to abdicate increasing amounts of personal information: an increased data burden, unequal burdens based on various credit bands, and the risk of expanding inequality through biased credit decisions.

The key assumption underlying any assessment of the benefits and harms of Big Data for individual consumers is that credit is beneficial to an individual's personal finances. On this question, W. Bruce Wydick of the University of San Francisco found that "empirical results show that upward class structure mobility increases substantially with access to credit, and also suggest that the combined effect of innate entrepreneurial ability and credit access has a greater impact on upward class structure mobility than the interaction between formal schooling and credit access."[86] In 2017, *TechCrunch* argued that access to credit is a key driver of social mobility that helps individuals secure the assets they need to continue moving up in the financial system, transition from jobs to careers, access services like insurance and healthcare, and contribute to the local tax bases that can invigorate their local communities.[87] Therefore, we will proceed in this analysis under the assumption that credit access is beneficial to consumers.

Big Data as an enabler of expanded credit access. As discussed in prior sections, Big Data is an enabling technology that allows lenders to make confident lending decisions to consumers ordinarily ignored based on their low FICO scores. The other side of the coin is that borrowers who would otherwise be approved based on their scores may no longer

[86] W. Bruce Wydick, "Credit Access, Human Capital and Class Structure Mobility," *Journal of Development Studies* 35, no. 6 (August 1999): 131–52, https://doi.org/10.1080/00220389908422605.
[87] Max Levchin, "Honest Financing as a Path to Economic Mobility and Social Justice," *TechCrunch* (blog), July 5, 2017.

qualify for credit they normally would have if they had provided less data. However, on net, the impacts of both these effects will be to provide credit more often to borrowers who are well-positioned to manage debt effectively. In the long term the expected outcome, all things being equal, would be to reduce delinquencies and interest payments, providing better credit products to consumers. While not all lenders will have the goal of expanding credit access to the underbanked, the net effect of Big Data will most likely be to provide higher-quality credit products to consumers in the financial position to manage them well.

The cost of accessing credit. The improved access to credit is not free for consumers, who will have their personal information collected and analyzed in ever-more-sophisticated ways. As we note above, data collected from 3rd party sources will include personal information not previously available during the traditional credit card application process. In addition to the multitude of data points, Big Data will also make more precise comparisons between prospective borrowers, using predictive algorithms that can make consumers uncomfortable. Target suffered from a spate of bad press in 2012 when it was found that its predictive algorithms were detecting women's pregnancies and offering them coupons for maternity products.[88] Even when companies adhere to privacy laws, data-drawn insights can still be perceived as intrusive by consumers.

Further, the burden of providing personal data will fall more heavily on people in lower credit bands. While a working professional with a high credit score may earn approval based on nothing but the information in their application, the owner of a subprime FICO score and/or thin file is more likely to be assessed using a combination of their application and a host of other observations from external datasets. While these customers would be granted access to credit otherwise unavailable to them, the hidden cost of that credit is the unequal distribution of the burden of data intrusion.

For all parties involved, however, the rise of Big Data presents privacy risks in that there are more transfers of personal information. Traditional datasets are compromising enough when hacked. A hack of the credit reporting firm Equifax compromised the Social Security numbers, birth dates, and home addresses of 143 million Americans in September 2017.[89] A previous hack of Experian exposed 15 million Americans' information.[90] The increased importance and breadth of large credit data brokers in the era of Big Data will drive increasing storage and transfer of personal information. To protect consumers, it is vital that the information be secured and curated to prevent its exfiltration and exploitation by hackers and other financial criminals.

---

[88] "How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did," accessed December 26, 2018, https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#6e539ca06668.

[89] Brian Fung, "Data of 143 Million Americans Exposed in Hack of Credit Reporting Agency Equifax," Washington Post, accessed December 26, 2018, https://www.washingtonpost.com/business/technology/equifax-hack-hits-credit-histories-of-up-to-143-million-americans/2017/09/07/a4ae6f82-941a-11e7-b9bc-b2f7903bab0d_story.html.

[90] Sam Thielman, "Experian Hack Exposes 15 Million People's Personal Information," *The Guardian*, October 2, 2015, sec. Business, https://www.theguardian.com/business/2015/oct/01/experian-hack-t-mobile-credit-checks-personal-information.

The risk of further marginalization of economically disadvantaged borrowers. As noted in the discussion of Big Data as a credit assessment tool, Big Data is expanding access to credit, but not uniformly. Here we present two potential drivers of further marginalization of traditionally underbanked groups that may result from poor algorithm or business process design as the lending industry leans ever more heavily upon algorithmic decision-making.

As Big Data comes to govern lending decisions, one decision operators will face will be to decide whether to optimize their credit pools for revenue or profit. To optimize for revenue would be to expand credit access to as many as possible, potentially offering more restrictive credit limits and higher interest rates to protect from the extra risk involved with subprime lending. As Ryan Brellenthin, former credit risk associate at Avant and current David M. Rubenstein fellow at the Harvard Kennedy School's Center for Public Leadership, points out, this is desirable for those with the goal of providing more equitable credit access to marginalized groups in society. On the other hand, if lenders optimize for profit, they may be inclined to use Big Data to restrict their credit pools, increasing their creditworthiness thresholds to reduce delinquencies as much as possible.[91] The result would be a pool shrinking toward the most creditworthy, most financially stable parts of the applicant pool.

One way to counteract this trend would be to promote free information and give people more access to their credit information, through services such as Credit Karma, which provides free credit scores with insights relevant to the individual consumer. This would help populations at risk of further marginalization to have more control over their financial health and position themselves for successful entry or re-entry into credit markets.

The risk of unintentional discrimination against protected populations. Another potential moral hazard comes from the risk of racial or quasi-racial selection. Multiple regression analyses are designed to assess datasets with large numbers of features on each consumer, then determine which are the most predictive of paying back a loan based on statistical association, not a cause-and-effect story as in traditional analysis. Carol Evans, Associate Director of the Division of Consumer and Community Affairs for the Board of Governors of the Federal Reserve System, points out that "the more speculative the nexus with creditworthiness, the higher the fair lending risk."[92] This phenomenon was first pointed out in 1997,[93] but may increase in relevance in a world of Big Data. Evans points out that increased use of data has also led to examples of regressive policies across

[91] Ryan Brellenthin (Harvard Kennedy School Center for Public Leadership), interviewed by Guillaume Delepine, Cambridge, Massachusetts, September 2018

[92] Carol Evans, "Keeping Fintech Fair: Thinking About Fair Lending and UDAP Risks," Consumer Compliance Outlook, Second Issue 2017, https://www.consumercomplianceoutlook.org/2017/second-issue/keeping-fintech-fair-thinking-about-fair-lending-and-udap-risks/.

[93] "Cf. 1994 Policy Statement 59 Fed. Reg. at 18269," *OCC Bulletin*, no. 97–24 (May 20, 1997): Appendix at 1-2, 10-11.

sectors:[94] criminal justice (racially biased predictions of recidivism rates),[95] internet advertising (gender biased advertising for lucrative jobs),[96] and delivery services (exclusion of minority neighborhoods).[97] Detecting and preventing such insidious forms of discrimination will likely require a collaborative effort to make sure that data are representative of the true population, are applied in ways that are clear indicators of creditworthiness, and are analyzed in the context of historical societal trends.

The Fair Credit Reporting Act (FCRA) limits lenders' ability to factor certain consumer traits into credit decisions. In general, society tends to react against discrimination based on innate traits like race, but to be more accepting of discrimination based on behaviors. With the mass ingestion and processing of consumer data, there may be a risk that algorithms make decisions based on behaviors that associate strongly with race, but for which the causal link to creditworthiness is unclear. Given the history of economic marginalization in the United States, there could be any number of non-regulated consumer data points that could serve as proxies for race, and re-entrench the economic disparities upon racial lines that we present in the introduction to this paper. We do not intend to suggest that any lender, among those we interviewed or otherwise, would willingly enact such a policy, but that vigilance for this phenomenon could prevent lenders from injecting unintentional racial biases into their operations.

Individual consumers are likely to remain the most vulnerable stakeholders in the Big Data ecosystem from consumer lending. While the private sector seeks to profit off of end consumers by offering better services at higher margins, it will be the job of the public sector to protect individual rights and interests.

Policy Assessment

In this section, we propose a framework for governing Big Data along with short-term policy ramifications. The Federal Trade Commission and Consumer Financial Protection Bureau are already taking steps to prepare for the evolution of the industry and its impacts on consumers, and our recommendations reflect step evolutions from their current activities. As with any evolving technology, government will need to adapt quickly to new technological developments in its efforts to protect consumers.

Framework for Big Data Governance

The business and policy landscape for governing Big Data and consumer lending is likely to evolve rapidly over the coming years. The specific policy demands of the moment will depend

[94] Kate Crawford, "Opinion | Artificial Intelligence's White Guy Problem," *The New York Times*, January 20, 2018, sec. Opinion, https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.
[95] Alexandra Chouldechova, "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments," *ARXIV*, no. 1703 (February 28, 2017), http://arxiv.org/abs/1703.00056.
[96] Amit Datta, Michael Carl Tschantz, and Anupam Datta, "Automated Experiments on Ad Privacy Settings," *Proceedings on Privacy Enhancing Technologies* 2015, no. 1 (April 1, 2015): 92–112, https://doi.org/10.1515/popets-2015-0007.
[97] Crawford, "Opinion | Artificial Intelligence's White Guy Problem."

on the pace of evolution in lending technology, macroeconomic factors, and the range of abuses and breaches that surface in the coming business cycles. While we propose short-term policy solutions in later sections, here we suggest a more stable framework for approaching the transition from traditional to algorithmic lending.

Prepare to govern. The public hearings of Mark Zuckerberg and other Facebook officials before Congress fueled public perceptions, especially in Silicon Valley, of the government having a limited understanding of emerging technologies and therefore being unfit to put in place even commonsense regulations. We understand that there are multitudes of data- and business-literate government workers, who have an acute understanding of how Big Data will impact consumer lending. However, just as the private sector has had to revamp its workforces to meet the demands of a more Big Data-driven value chain, so must regulators. Understanding what algorithms can and cannot do, and at which point to step in and govern, will be critical for regulators to craft effective policy.

Singapore may present a valuable case study for American policymakers. As part of Singapore's Digital Government Blueprint (DGB), the country has set the goal of educating 20,000 of its government workers in Big Data by 2023, appointing chief digital strategy officers to various agencies.[98] Such progressive policies appear to integrate with a growing Big Data ecosystem that will include a $1b USD Facebook datacenter and joint research institute with Alibaba.[99]

Encourage and enhance innovation. The enormous potential for Big Data to improve the health of the financial system, company profits, and consumer welfare should be valued highly in the minds of policymakers. As new advances in statistics, machine learning, and artificial intelligence continue to proliferate, the benefits will continue to increase. To ensure that society captures these benefits, regulators must be wary of instituting policies that make research and development more difficult for consumer lending companies. Policies that restrict access to data, the deployment of new technologies at scale, or certain statistical techniques may carry short-term benefits, but could hamper innovation. Policymakers will need to consider whether their social objectives can only be achieved through innovation-hampering policy choices, or if social goals can be met in other ways.

Policymakers should continually seek to enhance innovation by adding emphasis to transparency. As Ash Gupta pointed out, there can be a risk that Big Data creates opaqueness in a system such that it will be unclear how pricing or creditworthiness decisions are made. In a search engine, this is a tolerable trade secret, but in financial markets, it is critical to ensure the financial safety of protected classes. Regulators must make sure that any Big Data development is guided by higher transparency so that those

[98] Princy Lalawat, "How Is Singapore Gearing Up for Big Data and AI Revolution," *Analytics Insight* (blog), June 16, 2018, https://www.analyticsinsight.net/how-is-singapore-gearing-up-for-big-data-and-ai-revolution/.
[99] "Singapore's Big Ambitions for Big Data in 2019," accessed January 27, 2019, https://www.edb.gov.sg/en/news-and-resources/insights/innovation/singapore-s-big-ambitions-for-big-data-in-2019.html.

forming credit rankings can give the public a chance to correct any problematic trends.[100] In this goal, regulators will not always find themselves at odds with the private sector, who also benefits from transparency. Gupta cites a potential scenario in which a business might see that a larger loan size is correlated to lower delinquency rates. Assuming a causal relationship would lead the lender to give out larger loans, but if the mechanism is really that the lender simply focuses more support resources on the larger loans, the outcomes could be particularly detrimental to the company.[101] If policymakers can identify win-win opportunities to build more transparency into Big Data lending models, they will ease the process of crafting useful regulation.

Collaboration with the private sector will need to continually evolve. A good starting point is the Federal Trade Commission's series, Hearings on Competition and Consumer Protection, in which the FTC holds public discussions with industry leaders to understand the latest trends. Their event in November 2018 featured 27 speakers from government, academia, and industry to share ideas on panels arranged around the latest topics at the intersection of business and Big Data technology.[102] Collaboration through extensive hearings will ensure that government and the private sector share the same understanding of the role of Big Data in consumer lending decisions, and lead the policymaking process to a socially optimal result.

Monitor the algorithms. Algorithms may be easier to govern than humans because there is a documented record of their decision-making structures. To show that an algorithm is in compliance with a set of regulations would be a simpler process than the need for comparative testimonies associated with assessing the ways that human corporate agents make decisions. We suggest two guidelines for policymakers: to monitor outcomes, not techniques; and to make sure to keep humans in the decision-making loop to the extent feasible.

In her testimony to the FTC, Melissa McSherry of Visa explained the benefits of regulating *outcomes* rather than *techniques.*[103] We suggest policymakers pursue regulation that applies existing policy's values to the world of Big Data, such as making lending decisions fairly and averting fraud with the existing levels of privacy invasion. However, to regulate the techniques themselves, such as prohibitions on types of machine learning, would inhibit the ability of the private sector to innovate new methods of bringing socially beneficial statistical tools to market.

We also propose that humans remain in the loop in making credit decisions. Many of the algorithms in place today have yet to face the edge cases sure to arise in a recession or credit crisis. Should edge cases lead to unpredictable behavior from the algorithms, vigilant monitoring can prevent these behaviors from leading to business and societal

[100] Ash Gupta (Former chief risk officer, American Express) interviewed by Marshall Lux and Guillaume Delepine, October 2018.
[101] Ibid.
[102] Andrew I Gavil et al., "Tuesday, November 13, 2018 Welcome and Introductory Remarks," November 13, 2018, 10.
[103] Melissa McSherry.

disruption. We recommend that regulatory agencies recruit business operators to the task of monitoring consumer lending algorithms, as their day-to-day experience and investments at stake in the algorithms' performance align the incentives of business and government.

In deploying these monitoring systems, we recognize that the resources are not available to examine every algorithm. Therefore, establishing thoughtful rules that allow companies to self-govern takes on increased importance in ensuring fair financial play at industry scale without requiring the establishment of a massive regulatory agency.

Prohibit Unfair or Deceptive Acts or Practices (UDAP). The Federal Trade Commission Act contains stipulations outlawing Unfair or Deceptive Acts or Practices, defined as those that mislead consumers and cause them harm. Carol Evans, Associate Director in the Division of Consumer and Community Affairs for the Board of Governors of the Federal Reserve System, explains that:

> "Deceptive acts or practices are representations, omissions, or practices that are likely to mislead a consumer acting reasonably under the circumstances and are material (i.e., are likely to affect the consumer's conduct or decision with respect to a product or service). Unfair acts or practices are those that cause or are likely to cause substantial injury to consumers that consumers cannot reasonably avoid."[104]

There is already case history in regulation of credit cards marketed toward sub-prime borrowers. In 2008, the FDIC announced a $114m settlement with CompuCredit Corporation, who was found to have misrepresented the fees and available credit its borrowers would receive.[105] As credit line decisions are further automated, it will be critical that consumers are not misled by "black box" algorithms. The burden falls on both private sector lenders and public sector regulators to make sure that marketing to populations targeted by Big Data-enabled products does not mislead consumers.

Protect privacy. Though the topic has been much discussed and is not the focus of this work, we remind policymakers of the importance of protecting consumer privacy as Big Data algorithms grow in sophistication and breadth of application. As algorithms seek to ingest a wider range of consumer data points, it will be crucial to monitor whether those data points are business critical or whether the tradeoffs with privacy they necessitate are worthwhile.

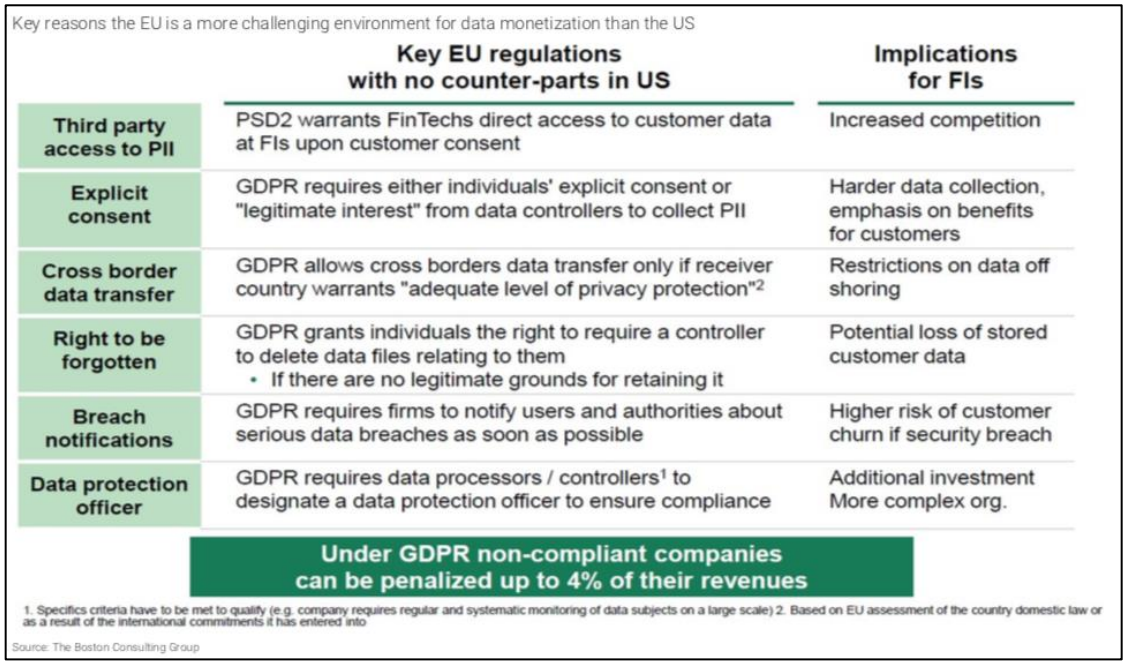
Policymakers face a new challenge in regulating Big Data as it rises to prominence in the consumer lending industry. Adherence to these four principles and vigilance for additional ways to protect the interests of the various stakeholders involved will help prevent policy pitfalls and misunderstandings between the public and private sector.

---

[104] Evans, "Keeping Fintech Fair."
[105] David Barr, "FDIC: Press Releases - PR-142-2008 12/19/2008," *FDIC.Gov* (blog), December 19, 2008, https://www.fdic.gov/news/news/press/2008/pr08142.html.

Recommendations for Evolving Policy

In applying the data policy framework, there is certain to be a measure of trial and error in converting society's values into a set of policies for the industry to follow. The European Union has been more aggressive than the United States in regulating corporations using data to make decisions. The General Data Protection Regulation applies to companies that do business in the European Union. The Boston Consulting Group summarizes the policy implications of GDPR as they apply to financial institutions. They lend perspective to the "Encourage and enhance innovation" pillar of the policymaking framework we propose and are presented in the figure below.



Key reasons the EU is a more challenging environment for data monetization than the US

	Key EU regulations with no counter-parts in US	Implications for FIs
Third party access to PII	PSD2 warrants FinTechs direct access to customer data at FIs upon customer consent	Increased competition
Explicit consent	GDPR requires either individuals' explicit consent or "legitimate interest" from data controllers to collect PII	Harder data collection, emphasis on benefits for customers
Cross border data transfer	GDPR allows cross borders data transfer only if receiver country warrants "adequate level of privacy protection"[2]	Restrictions on data off shoring
Right to be forgotten	GDPR grants individuals the right to require a controller to delete data files relating to them • If there are no legitimate grounds for retaining it	Potential loss of stored customer data
Breach notifications	GDPR requires firms to notify users and authorities about serious data breaches as soon as possible	Higher risk of customer churn if security breach
Data protection officer	GDPR requires data processors / controllers[1] to designate a data protection officer to ensure compliance	Additional investment More complex org.

Under GDPR non-compliant companies can be penalized up to 4% of their revenues

1. Specifics criteria have to be met to qualify (e.g. company requires regular and systematic monitoring of data subjects on a large scale) 2. Based on EU assessment of the country domestic law or as a result of the international commitments it has entered into

Source: The Boston Consulting Group

106

In applying the lessons of Europe's foray into data protection, American policymakers will have an opportunity to improve upon the lessons learned in Europe. While the GDPR policies have complicated operations for financial institutions, they have also led to new best practices in data governance that have enhanced companies' ability to track and manage data across multiple storage facilities. Ongoing collaboration with the financial institutions will help to enact policies that can protect consumers without overly crippling the operations of credit providers.

We recognize that the process of governing consumer lending in the age of Big Data will require an ongoing process with multiple rounds of input from consumers and the private sector, but we suggest a few concrete steps that the agencies most likely to be responsible for the area can take to prepare. In particular, we make suggestions for the Consumer Financial Protection Bureau and Federal Trade Commission.

Consumer Financial Protection Bureau. The CFPB holds the charter for the regulatory aspects of consumer lending. Its office of Supervision, Enforcement, & Fair Lending is

[106] Faucette et al., "Data Analytics for FIs: The Journey from Insight to Value."

led by policy associate director Erik Blankenstein and associate director Christopher D'Angelo.[107] We recommend the leadership continue to maintain close ties with the FTC to find opportunities for collaboration in consumer education and enforcement. The CFPB has more direct capabilities in monitoring lenders and enforcing the law, and therefore may need to take a more active role than the FTC after any discovered incidents that require intervention. In this regard, assistant director Kristen Donoghue's office is likely to play a key role.[108]

The Research, Markets, & Regulations team led by policy associate director Tom Pahl and associate director David Silberman can play a key role in helping the CFPB prepare for the changes coming to the consumer lending industry. In particular, the agents likely to be involved are William Wade-Gery, assistant director of card, payment, & deposit markets, and John McNamara, assistant director of consumer lending, reporting, & collection markets.[109] We would recommend that the Research, Markets, & Regulations team work closely with the FTC, and consider hiring trained data scientists to ensure that the agency can access an unbiased description of what new algorithms can and cannot do.

The CFPB also maintains a service for individual support at the federal level for complaints about credit reporting agencies.[110] As credit reporting becomes more and more algorithmically moderated, it is likely that errors and complaints about reporting will become more systematic than individual. The individual complaint line will continue to serve a useful function in pressuring credit reporting agencies to be responsive to individual needs. In addition to this function, we recommend that the CFPB monitor the complaints quarterly to try to identify trends in credit reporting problems. Along with an effective enforcement mechanism, this avenue may become the federal government's best option for detecting unfair credit assessment practices and aggregating consumer complaints.

Federal Trade Commission. The FTC does not directly manage the banks that issue credit cards, but has a role in regulating credit rating agencies (CRAs) based on the authority given to it under the Fair Credit Reporting Act (FCRA). As recently as 2015, the FTC issued a follow-up study on the accuracy of credit scores after multiple consumers complained that they had been scored unfairly using incorrect information.[111]

---

[107] "Bureau Structure," Consumer Financial Protection Bureau, accessed December 27, 2018, https://www.consumerfinance.gov/about-us/the-bureau/bureau-structure/.
[108] "Supervision, Enforcement, Fair Lending Division," Consumer Financial Protection Bureau, accessed December 27, 2018, https://www.consumerfinance.gov/about-us/the-bureau/bureau-structure/supervision-enforcement-fair-lending/.
[109] "Research, Markets, Regulations Division," Consumer Financial Protection Bureau, accessed December 27, 2018, https://www.consumerfinance.gov/about-us/the-bureau/bureau-structure/research-markets-regulation/.
[110] "CFPB Now Taking Complaints on Credit Reporting," Consumer Financial Protection Bureau, accessed December 27, 2018, https://www.consumerfinance.gov/about-us/newsroom/consumer-financial-protection-bureau-now-taking-complaints-on-credit-reporting/.
[111] "FTC Issues Follow-Up Study on Credit Report Accuracy," Federal Trade Commission, January 21, 2015, https://www.ftc.gov/news-events/press-releases/2015/01/ftc-issues-follow-study-credit-report-accuracy.

The critical first step for the FTC is to understand its role in regulating credit card companies in the age of Big Data. As credit rating agencies fade from being the definitive authority on consumer credit to one of a set of data inputs to credit assessment models, they will no longer be the only target for FCRA compliance. Instead, the banks, which make the majority of credit decisions, will come under FTC oversight. Before taking any further action, it is critical that the FTC understand, document, and publicize its role in consumer lending so that all parties can know in which situations to expect the commission to intervene. There are two primary ways in which the FTC has begun to stake out its role and should continue: providing intelligence to the CFPB and educating consumers.

The FTC provides annual reports to the Consumer Financial Protection Bureau (CFPB) describing its activities in enforcing the Equal Credit Opportunity Act (ECOA) for all entities that are in the commission's purview: financial institutions that are not banks, thrifts, or federal credit unions. As noted above, the FTC has held multiple hearings to prepare itself for the challenges of regulating institutions in its purview after Big Data reaches fruition.[112] Continuing to engage the private sector, even when the learning curve appears to slow, will be critical to ensuring a steady flow of ideas and shared mutual understanding of the best policy going forward.

The FTC also has in its purview customer education. We recommend that it continue with its approach of recruiting lenders themselves to educate their consumers.[113] In the short term, these activities can help lenders establish brands as technologically forward-thinking.

Conclusion

The Big Data revolution in consumer lending presents tremendous potential for business and social benefit. By fighting fraud, producing better financial products, and making more accurate credit assessments, Big Data can help keep value in the hands of consumers and lenders instead of letting it erode to fraudsters and credit delinquency. Policymakers need to be advised that regulation that stifles the growth of Big Data in the consumer lending space may place at risk the tremendous potential social benefit. However, there is reason for policymakers to be alert to ways to ensure that algorithms continue to provide benefits to society by limiting the potential harmful externalities: socially marginalizing algorithms, increasingly intrusive data policies, algorithmic fragility in recession periods, and the privacy risks from cyber security. It will be critical in the short term to work closely with the private sector while enhancing the skills base in the public sector to prepare to react to potential data-related shocks to the consumer lending system. With a well prepared regulatory body in place, there is reason for optimism to surround the wave of Big Data in consumer lending.

[112] James Reilly Dolan, "Federal Trade Commission Activities Under the Equal Employment Opportunity Act and Regulation B: Report to the Consumer Financial Protection Bureau," February 6, 2005.
[113] Federal Trade Commission, "Big Data: A Tool for Inclusion or Exclusion," September 15, 2014, https://www.ftc.gov/system/files/documents/public_events/313371/bigdata-transcript-9_15_14.pdf.

Works Cited

"American Express Creates New Technology Hub in Silicon Valley | American Express." Accessed December 20, 2018. https://about.americanexpress.com/press-release/american-express-creates-new-technology-hub-silicon-valley.

Azzano, Michael. "Oportun Acquires Intellectual Property and Lending Platform of SpringboardAuto.Com." Accessed December 20, 2018. https://globenewswire.com/news-release/2018/10/09/1618529/0/en/Oportun-Acquires-Intellectual-Property-and-Lending-Platform-of-SpringboardAuto-com.html.

Back, Aaron. "A Surprise Bump in Bad Card Loans - WSJ," September 20, 2017. https://www.wsj.com/articles/a-surprise-bump-in-bad-card-loans-1505899800.

Barr, David. "FDIC: Press Releases - PR-142-2008 12/19/2008." *FDIC.Gov* (blog), December 19, 2008. https://www.fdic.gov/news/news/press/2008/pr08142.html.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace. "Consumer-Lending Discrimination in the Era of FinTech," October 2018, 44.

Board of Governors of the Federal Reserve System. "Report on the Economic Well-Being of U.S. Households in 2016 - May 2017." FederalReserve.gov, June 14, 2017. https://www.federalreserve.gov/publications/2017-economic-well-being-of-us-households-in-2016-banking-credit.htm.

"Bureau Structure." Consumer Financial Protection Bureau. Accessed December 27, 2018. https://www.consumerfinance.gov/about-us/the-bureau/bureau-structure/.

"Cf. 1994 Policy Statement 59 Fed. Reg. at 18269." *OCC Bulletin*, no. 97–24 (May 20, 1997): Appendix at 1-2, 10-11.

"CFPB Now Taking Complaints on Credit Reporting." Consumer Financial Protection Bureau. Accessed December 27, 2018. https://www.consumerfinance.gov/about-us/newsroom/consumer-financial-protection-bureau-now-taking-complaints-on-credit-reporting/.

Chen, Min, Shiwen Mao, and Yunhao Liu. "Big Data: A Survey." *Mobile Networks and Applications* 19, no. 2 (April 2014): 171–209. https://doi.org/10.1007/s11036-013-0489-0.

Chouldechova, Alexandra. "Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments." *ARXIV*, no. 1703 (February 28, 2017). http://arxiv.org/abs/1703.00056.

Consumer Financial Protection Bureau. "The Consumer Credit Card Market 2017." Consumer Financial Protection Bureau, December 2017. https://www.consumerfinance.gov/about-us/newsroom/cfpb-releases-report-state-credit-card-market/.

Crawford, Kate. "Opinion | Artificial Intelligence's White Guy Problem." *The New York Times*, January 20, 2018, sec. Opinion. https://www.nytimes.com/2016/06/26/opinion/sunday/artificial-intelligences-white-guy-problem.html.

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. "Automated Experiments on Ad Privacy Settings." *Proceedings on Privacy Enhancing Technologies* 2015, no. 1 (April 1, 2015): 92–112. https://doi.org/10.1515/popets-2015-0007.

Dolan, James Reilly. "Federal Trade Commission Activities Under the Equal Employment Opportunity Act and Regulation B: Report to the Consumer Financial Protection Bureau," February 6, 2005.

Evans, Carol. “Keeping Fintech Fair: Thinking About Fair Lending and UDAP Risks.” Consumer Compliance Outlook, Second Issue 2017. https://www.consumercomplianceoutlook.org/2017/second-issue/keeping-fintech-fair-thinking-about-fair-lending-and-udap-risks/.

Experian Information Solutions. “Experian Global Fraud Report 2018.” Experian.com, 1/18. https://www.experian.com/assets/decision-analytics/reports/global-fraud-report-2018.pdf.

Faucette, James, Vasundhara Govil, Danyal Hussain, Sushil Malhotra, Mohammed Badi, Berk Hizir, and Alenka Grealish. “Data Analytics for FIs: The Journey from Insight to Value.” The Boston Consulting Group & Morgan Stanley, October 25, 2016.

Federal Trade Commission. “Big Data: A Tool for Inclusion or Exclusion,” September 15, 2014. https://www.ftc.gov/system/files/documents/public_events/313371/bigdata-transcript-9_15_14.pdf.

“FTC Issues Follow-Up Study on Credit Report Accuracy.” Federal Trade Commission, January 21, 2015. https://www.ftc.gov/news-events/press-releases/2015/01/ftc-issues-follow-study-credit-report-accuracy.

Fung, Brian. “Data of 143 Million Americans Exposed in Hack of Credit Reporting Agency Equifax.” Washington Post. Accessed December 26, 2018. https://www.washingtonpost.com/business/technology/equifax-hack-hits-credit-histories-of-up-to-143-million-americans/2017/09/07/a4ae6f82-941a-11e7-b9bc-b2f7903bab0d_story.html.

Garg, Amit, Davide Grande, Gloria Macias-Lizaso Miranda, Christoph Sporleder, and Eckhart Windhagen. “Analytics in Banking: Time to Realize the Value.” McKinsey, April 2017. https://www.mckinsey.com/industries/financial-services/our-insights/analytics-in-banking-time-to-realize-the-value.

Gavil, Andrew I, Michael Kearns, John P Dickerson, and Michael D Abràmoff. “Tuesday, November 13, 2018 Welcome and Introductory Remarks,” November 13, 2018, 10.

Hoffower, Hillary. “Americans Are Using Their Credit Cards More than Ever — and Some Experts Think That’s a Good Thing.” Business Insider. Accessed December 26, 2018. https://www.businessinsider.com/credit-card-use-increasing-in-us-economy-2018-5.

Lalawat, Princy. “How Is Singapore Gearing Up for Big Data and AI Revolution.” *Analytics Insight* (blog), June 16, 2018. https://www.analyticsinsight.net/how-is-singapore-gearing-up-for-big-data-and-ai-revolution/.

“How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did.” Accessed December 26, 2018. https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#6e539ca06668.

Jeffs, Mark, and Christy Williams. “Fiserv and Mastercard Help Financial Institutions Approve More Genuine Transactions Without Increasing Risk.” Accessed November 7, 2018. https://www.businesswire.com/news/home/20180306005515/en/Fiserv-Mastercard-Financial-Institutions-Approve-Genuine-Transactions.

Levchin, Max. “Honest Financing as a Path to Economic Mobility and Social Justice.” *TechCrunch* (blog), July 5, 2017.

Lux, Marshall, and Robert Greene. “The State and Fate of Community Banking,” February 2015, 32.

“MasterCard Acquires 5one Marketing Limited.” MasterCard Social Newsroom. Accessed December 20, 2018. https://newsroom.mastercard.com/press-releases/mastercard-acquires-5one-marketing-limited/.

"Mastercard Acquires NuData Security Inc: Fraud Prevention." NuData Security, March 29, 2017. https://nudatasecurity.com/press/mastercard-enhances-security-internet-things-acquisition-nudata-security-inc/.

"MasterCard Announces Acquisition of Applied Predictive Technologies." MasterCard Social Newsroom. Accessed December 20, 2018. https://newsroom.mastercard.com/press-releases/mastercard-announces-acquisition-of-applied-predictive-technologies/.

"Mastercard Enhances Artificial Intelligence Capability with the Acquisition of Brighterion, Inc." MasterCard Social Newsroom. Accessed December 20, 2018. https://newsroom.mastercard.com/press-releases/mastercard-enhances-artificial-intelligence-capability-with-the-acquisition-of-brighterion-inc-2/.

Mian, Amir Sufi, Atif. "Credit Expansion Is the Real Cause of Booms and Busts." MarketWatch. Accessed December 26, 2018. http://www.marketwatch.com/story/credit-expansion-is-the-real-cause-of-booms-and-busts-2018-03-08.

Rassloff, Jens. "KPMG Named a Leader in Microsoft Dynamics 365 Services by Forrester Research | KPMG | GLOBAL." KPMG, January 31, 2018. https://home.kpmg/xx/en/home/insights/2018/01/kpmg-named-a-leader-in-forrester-research.html.

"Research, Markets, Regulations Division." Consumer Financial Protection Bureau. Accessed December 27, 2018. https://www.consumerfinance.gov/about-us/the-bureau/bureau-structure/research-markets-regulation/.

Robitzski, Dan. "The LAPD's Terrifying Palantir-Powered Policing Algorithm Was Just Uncovered and Yes It's Basically 'Minority Report' – Futurism." News. Futurism, May 10, 2018. https://futurism.com/lapd-documents-show-their-policing-algorithms-continue-to-target-minorities-and-past-offenders.

"Singapore's Big Ambitions for Big Data in 2019." Accessed January 27, 2019. https://www.edb.gov.sg/en/news-and-resources/insights/innovation/singapore-s-big-ambitions-for-big-data-in-2019.html.

Smith, Adam. "Amazon Shut Down Recruiting AI for Sexist Results | News & Opinion | PCMag.Com," October 10, 2018. https://www.pcmag.com/news/364315/amazon-shut-down-recruiting-ai-for-sexist-results.

Souza, Rob, Rob Trollinger, Cornelius Kaestner, and Jan Jamrich. "How to Get Started with Big Data." The Boston Consulting Group, May 29, 2013. https://www.bcg.com/publications/2013/advanced-analytics-how-get-started-big-data.aspx.

Sullivan, Bob. "State of Credit: 2017." Company Blog. Experian, January 11, 2018. https://www.experian.com/blogs/ask-experian/state-of-credit/.

"Supervision, Enforcement, Fair Lending Division." Consumer Financial Protection Bureau. Accessed December 27, 2018. https://www.consumerfinance.gov/about-us/the-bureau/bureau-structure/supervision-enforcement-fair-lending/.

The World Bank. "Domestic Credit to Private Sector (% of GDP) | Data." data.worldbank.org, December 26, 2018. https://data-worldbank-org.ezp-prod1.hul.harvard.edu/indicator/FS.AST.PRVT.GD.ZS?contextual=default&end=2016&locations=US&name_desc=true&start=1995&view=chart.

Thielman, Sam. "Experian Hack Exposes 15 Million People's Personal Information." *The Guardian*, October 2, 2015, sec. Business.

https://www.theguardian.com/business/2015/oct/01/experian-hack-t-mobile-credit-checks-personal-information.
US Census Bureau. “Population Clock,” November 7, 2018. https://www.census.gov/popclock/.
Wood, Laura. “Big Data Investments in the Financial Services Industry Will Account for Nearly $9 Billion in 2018 Alone.” PR News Wire, August 1, 2018. https://www.prnewswire.com/news-releases/big-data-investments-in-the-financial-services-industry-will-account-for-nearly-9-billion-in-2018-alone-300690188.html.
Woodie, Alex. “How Credit Card Companies Are Evolving with Big Data.” Datanami, May 3, 2016. https://www.datanami.com/2016/05/03/credit-card-companies-evolving-big-data/.
———. “Why Gartner Dropped Big Data Off the Hype Curve.” Blog. Datanami, August 26, 2015. https://www.datanami.com/2015/08/26/why-gartner-dropped-big-data-off-the-hype-curve/.
Wydick, W. Bruce. “Credit Access, Human Capital and Class Structure Mobility.” *Journal of Development Studies* 35, no. 6 (August 1999): 131–52. https://doi.org/10.1080/00220389908422605.
Yuhanna, Noel. “The Forrester Wave™: Big Data Fabric, Q2 2018.” Denodo, Q2 2018. https://www.denodo.com/en/page/forrester-wave-big-data-fabric-q2-2018.