WORLD BANK GROUP

Using Big Data to Expand Financial Services: Benefits and Risks

Facundo Abraham   Sergio L. Schmukler   José Tessada

Big data is transforming financial services around the world. Advances in data analytics and computational power are allowing firms to exploit data in an easier, faster, and more reliable manner, and at a larger scale. By using big data, financial firms and new entrants from other sectors are able to provide more and better financial services. Governments are also exploring ways to use big data collected by the financial sector more systematically to get a better picture of the financial system as a whole and the overall economy. Despite its benefits, the wider use of big data has raised concerns related to consumer privacy, data security, discrimination, data accuracy, and competition. Hence, policy makers have started to regulate and monitor the use of big data by financial institutions and to think about how to use big data for the benefit of all.

Big Data and the Financial Sector

The finance industry has always been driven by information. Providing finance is the act of offering money in exchange for a promise that it will be returned at a later date. Because financial institutions have imperfect knowledge about their customers, they need to acquire substantial information to assess customers' repayment capacity as accurately as possible. Therefore, collecting data on prospective borrowers is central to the provision of finance. These data typically include "hard data" (such as credit history, income, employment, education level, tax records, and financial statements) and "soft data" (such as opinions from loan officers, internal discussions, and economic prospects). Financial institutions need information not only about their borrowers, but about their lenders. Financial institutions lend and invest using money collected from third parties, which requires knowing their creditors' investment preferences and risk tolerance (including their financial goals, investment horizons, income, and future expenditures).

Once clients are using financial services, financial institutions collect data from that use. For example, financial institutions document all transactions clients make to keep track of their running balances. The amount of transactional data collected by financial institutions has grown exponentially as economic activity is becoming less cash-based and transactions are increasingly conducted through financial sector providers.

Financial institutions are also required to compile data to submit to regulators, supervisors, and other institutions, such as credit bureaus. Financial institutions are unlike any other businesses in two ways. First, because they intermediate other peoples' money, any losses end up being borne not only by the financiers but also possibly by the original creditors. Second, financial institutions can pose systemic risk. A crisis in one institution can lead to instability in the entire industry and the economy. For these reasons, the financial industry is tightly regulated and supervised. As a consequence, financial institutions need to periodically prepare comprehensive data about their activities and associated risks and submit this information to the authorities.

Financial institutions collect vast and detailed data on their clients' preferences, behavior, characteristics, and risks. These data have immense potential to help financial institutions better understand the environment in which they operate and, as a result, create new financial products and reach new segments of the population. But systematically using this huge body of data for market analysis is very cumbersome. With millions of clients and transactions per day, financial institutions have to process and analyze an enormous amount of data.

New methods to analyze data and higher computational power in the context of the so-called big data revolution are now enabling financial sector participants to collect and analyze their data more easily and quickly than ever. For example, to map the use of different financial services with the characteristics of customers that use them (such as their income, residence, or educational level), financial institutions can employ association rules. They can also use text mining tools to search for specific keywords in large, complex documents such as internal reports and firms' financial statements.

The big data revolution has also allowed financial institutions to augment their own private data with public data and proprietary data sold by other providers. For instance, financial institutions can mine data from social media profiles or public records to supplement their information on clients. Similarly, they can track news, tweets, blog posts, and other online publications to monitor market sentiment and predict market upturns and downturns. Several financial activities such as personal and business banking, asset and wealth management, and insurance are already using big data (figure 1).

This brief discusses how big data is transforming the provision of financial services. In particular, it discusses how financial intermediaries use big data both to adapt services they offer to existing clients and to incorporate new clients into the financial system. The brief does not cover how big data is affecting other important aspects of the financial sector, such as algorithmic trading, the prevention of fraud, or artificial intelligence chatbots.
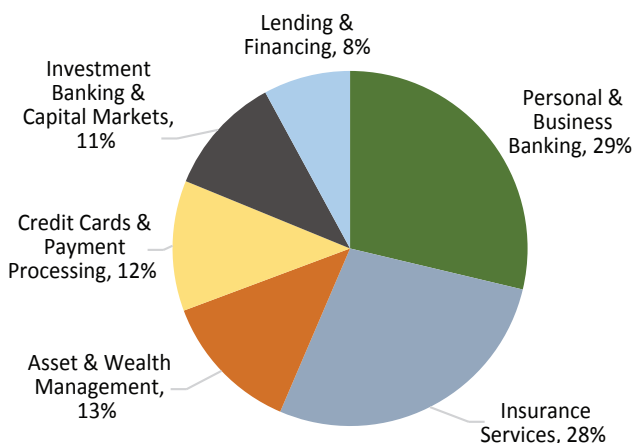
Potential Benefits of Big Data

Traditionally, to know and understand the needs of their customers, financial institutions use personalized relationships of clients with officers at a branch. However, this mode can have several limitations, such as when clients are new or when officers

WORLD BANK GROUP
Research & Development Center in Chile

Ministry of Economy and Finance
Republic of Korea

WORLD BANK GROUP
Global Knowledge & Research Hub in Malaysia

Figure 1. Investments in Big Data by Financial Sector Industry, 2018

Lending & Financing, 8%

Investment Banking & Capital Markets, 11%

Personal & Business Banking, 29%

Credit Cards & Payment Processing, 12%

Asset & Wealth Management, 13%

Insurance Services, 28%

A wide range of industries within the financial sector use big data.
Source: SNS Telecom & IT (2018).

are relocated to another branch. Large banks, with millions of clients, can find it particularly difficult to maintain personalized relations with clients.

Nowadays, financial institutions can leverage big data to offer personalized services to millions of customers. For example, Banorte, a Mexican bank with more than 13 million clients, is studying clients' banking behavior to know which products best suit each client, create a personalized experience when customers use Banorte's app, and enhance customers' direct interactions with bank personnel. Barclays has used customer sentiment in social media to detect and correct shortcomings in its app. Garanti, a leading Turkish bank, combines purchase data with GPS location data to identify customers' favorite stores and notify them when they are close to a preferred store that has a special offer (Forbes 2015).

Financial institutions are also using big data to conduct targeted advertising and promote cross-selling of their products. Instead of launching "one-size-fits-all" marketing campaigns, big data allows financial institutions to use a wide range of individual characteristics to segment clients into different groups and provide tailored offers according to their preferences. Some institutions are even using big data to anticipate customers' future needs. For example, a credit card company in the Republic of Korea uses transaction records of doctor visits and product purchases to determine whether a user is expecting a child and, as a result, provide tailored recommendations for credit cards and discounts on child care services (Korea Joongang Daily 2019).

Financial institutions can also use big data to expand their customer base. One of the main reasons for denying credit is lack of or inadequate credit history, which prevents financial institutions from properly assessing prospective customers' creditworthiness. Big data can mitigate this issue by enabling financial institutions to profile prospective borrowers based on the repayment of services other than loans, such as past payments of rent, utilities, or mobile phone plans. This information is not new; financial institutions already collected payment data from online payment of bills. But new technologies are making it easier to organize these data and report them to credit bureaus and registries. In some instances, new fintech companies, such as Destacame in Chile and Mexico, are the ones centralizing and verifying payment data. They then provide these data to financial institutions to assess prospective borrowers and extend loans.

Other financial institutions are taking advantage of new data to build alternative credit scores for low-income individuals. Examples include fintech companies like Lenddo and Tala, with presence in Africa, Asia, and Latin America. These companies typically offer an app that individuals can download on their phone and use to apply for loans. When they submit a loan application, the app scans the digital footprint of a mobile phone (contacts, social network profile, geographical patterns) and instantaneously decides whether to extend the loan. Many of these alternative credit providers are located in East Asia, such as Ayannah (the Philippines); CredoLab (Indonesia, Malaysia, and Singapore, among other countries); and Kakao Bank (Korea) (Fintech News 2017; The Economist 2019). Automated credit scoring not only can help expand credit, but also saves time and might reduce some of the human biases that can emerge in the evaluation process.

Big data can also promote access to finance for small and medium enterprises (SMEs). Financial institutions have difficulties assessing the risk of SMEs because they may not have been in business long and public information about them is limited. Large online retailers, such as Alibaba, Amazon, eBay, and Mercado Libre, are using big data to gather and analyze information on firms participating in their platforms (such as sales and customer ratings) and then providing loans (The Economist 2015). Alibaba offers credit to SMEs through its own digital bank called MyBank, which compiles data on monthly sales on Alibaba and then preapproves firms for loans. When a firm operating in Alibaba requires a loan, it simply completes an online application that is processed in only three minutes. Loans are proportionate to the volume of sales by the firm, are unsecured, and carry lower rates than traditional banks. As of 2018, MyBank serviced roughly half of all SMEs in China, 6 million SMEs (Chataing and Kushnir 2018).

Big data is also helping underserved groups by enhancing financial networks. In several emerging economies, mobile money still relies on networks of agents that allow users to deposit and withdraw cash. If poorly managed, agents in these networks can run out of money or not be located close enough to their customers. Zoona, a financial services provider in Zambia, uses simulations, data from field staff, and Google Maps to determine the optimal location of agents as well as potential demand (Mastercard Foundation and IFC 2017). DBS Bank, a Singaporean bank, uses data on ATM withdrawals and deposits to forecast ATM cash demands and optimize its ATM network (Fitzgerald 2014).

Some financial institutions, such as the BBVA Data & Analytics Center and the JP Morgan Chase Institute, are using their unique proprietary data to produce new knowledge that could be used to better understand consumers and provide better and new services. For example, one study examines income and consumption patterns of 2.5 million U.S. account holders (and 135 related million transactions) to draw conclusions about households' earnings and spending volatility (JP Morgan Chase Institute 2015).

Risks in the Use of Big Data

While big data has the potential to transform the way financial services are provided, successfully using big data presents some important challenges that, if not addressed, could have adverse effects on consumers and society. These concerns are not exclusive to the financial sector (Forbes 2017; Carrière-Swallow and Haksar 2019).

Using big data in financial institutions involves constructing large data sets that include private information (such as income, details about assets held, credit card numbers, and spending habits) submitted by individuals, willingly or de facto. Financial institutions can augment their own data by hiring third parties to collect consumer data on their behalf. Financial institutions could

be tempted to sell their proprietary data to interested third parties, such as marketing and retail companies, that are interested in understanding spending habits of different demographic and social groups (Asrow and Xu 2018). In this context, data privacy is an important concern. Financial institutions need to be transparent and promptly disclose what information they have on consumers and what are they doing with it. In some circumstances, consumers could be granted the power to limit financial institutions from selling specific information.

Once data are collected, financial institutions need to safeguard the information. In the wrong hands, it could be used to commit identity theft, scams, and extortions. Even as financial institutions have increased spending on IT security, the number of data breaches has grown. Around one-quarter of financial service companies reported a data breach in 2018—and half of them said that it was not the first time (Thales 2018). Financial institutions need to take steps to enhance data security by, for example, strengthening data access monitoring, data masking, and data encryption.

Cloud computing also poses benefits and risks. Cloud computing refers to storing, managing, and processing data in the Internet through third-party servers rather than in a firm's own servers. It reduces costs and increases flexibility. Despite these benefits, financial institutions might prefer to keep data in-house, arguing that storing it online would make it more vulnerable to hacking. However, hacking of in-house servers is just as likely through an Internet connection or other means. Because cloud computing services are specialized in data storage, they might even have more expertise in dealing with data security issues than a financial institution.

Another challenge of relying on big data is that it can sometimes lead to biases if used without the appropriate care (so-called algorithm bias) (FTC 2016). In the process of analyzing big data through algorithms, human bias can be introduced in multiple stages: from choosing the variables in the model and their relative importance to how the sample is collected and results interpreted. Furthermore, models can unintentionally introduce biases even when they are not programmed to do so. For example, involuntary bias can occur when models are calibrated using historical data that was generated under conditions of discrimination (such as loan applications where applicants were discriminated based on race) (Barocas and Selbst 2016).

Financial institutions might not only use big data to offer products that consumers might demand, but could also tailor prices to extract the maximum profit from each consumer. For example, algorithm-based lending can result in higher interest rates being charged to minorities after determining that these groups rarely do cost comparisons and live in areas where the supply of financial services, and thus competition, is limited (Bartlett et al. 2019). Some financial institutions even use big data to identify vulnerable groups in need of quick cash and offer them financially risky products, such as payday loans (Business Insider 2013).

More data do not necessarily lead to better analyses. Data sets can be so large that is hard to know where to start and what to look for; the relevant information might be hidden underneath a mound of useless data (O'Donnell 2017). Analytical tools used in big data can be good at finding correlations between variables but not so good at distinguishing correlation from actual causation. Traditional techniques to infer causation, such as randomized control trials and natural experiments, may also be needed (Athey 2017). Furthermore, big data is not always completely accurate. For example, a survey in the United States found that more than

two-thirds of respondents claimed that their consumer information was only 0–50 percent correct (Lucker, Hogan, and Bischoff 2017). Different factors can explain these errors, including the cost of maintaining up-to-date information and individuals consciously providing wrong information to protect their privacy.

While reducing information asymmetries between consumers and financial institutions, big data could potentially create information asymmetries across financial institutions. Large financial institutions start with a big base of consumers. With more data, large financial institutions can conduct better analyses and improve their services more easily than small institutions, widening their data advantage and reinforcing their dominant position. This feedback loop, which can lead to higher entry costs and monopolistic positions, can be amplified by the existence of economies of scale in big data (OECD 2016). Big data could also widen asymmetries between users of financial services. For instance, large firms tend to be older and have more available data than small firms. By making relatively better forecasts, big data could reduce the risk premiums for large firms relatively more than for smaller firms, widening their relative cost of capital and impairing competition (Begenau, Farboodi, and Veldkamp 2018). Concentration might also arise from other financial sector activities beyond the provision of finance, such as in algorithmic trading.

Policy Discussion

Policy makers around the world have started to discuss ways to meet the new challenges that accompany big data and to regulate its use (EBA 2016; U.S. Treasury 2018). Several countries have issued regulations to enhance consumers' rights over their own data and have increased penalties for data misuse (Greig 2019). Policy makers have also started to discuss whether big data can be a source of market power and whether steps to promote a more efficient distribution of data are needed (OECD 2016). Some governments have gone even further, regulating the use of automatic algorithms (Bloomberg Law 2019).

Beyond regulating big data, policy makers could be interested in collecting data from financial institutions. In the hands of financial institutions, data are used to identify new business opportunities that might result in better financial services. But if the government could collect, centralize, and asses data dispersed among financial institutions, it could use that information for policy making, such as identifying and mitigating gaps in financial markets. A recent survey (Central Banking 2018) shows that about 60 percent of central banks are using big data for multiple purposes, including forecasting, stress testing, and combating money laundering (figure 2). Uses include analyzing credit card spending data to gain real-time information on current economic activity. Compared to traditional retail survey data that are available only with considerable lags and are not collected for all locations, big data could prove useful to study localized and short-lived shocks (Aladangady et al. 2019).

Moreover, partially disclosing some of these data could increase data access, level the playing field, and encourage competition. Korea is an example of a country that is moving forward to open data for digital development and competition in the financial sector. In 2020, Korea is expected to launch CreDB, an open data system that will grant fintech firms, financial companies, and the academia access to de-identified credit data. Korea is also creating a platform for data exchange and a new agency to ensure the safe use and intermediation of big data (FSC 2019).

However, obtaining the needed information presents its own challenges. Financial institutions might be reluctant to disclose information that they use to gain a comparative edge. Forcing

Figure 2. Use of Big Data by Central Banks
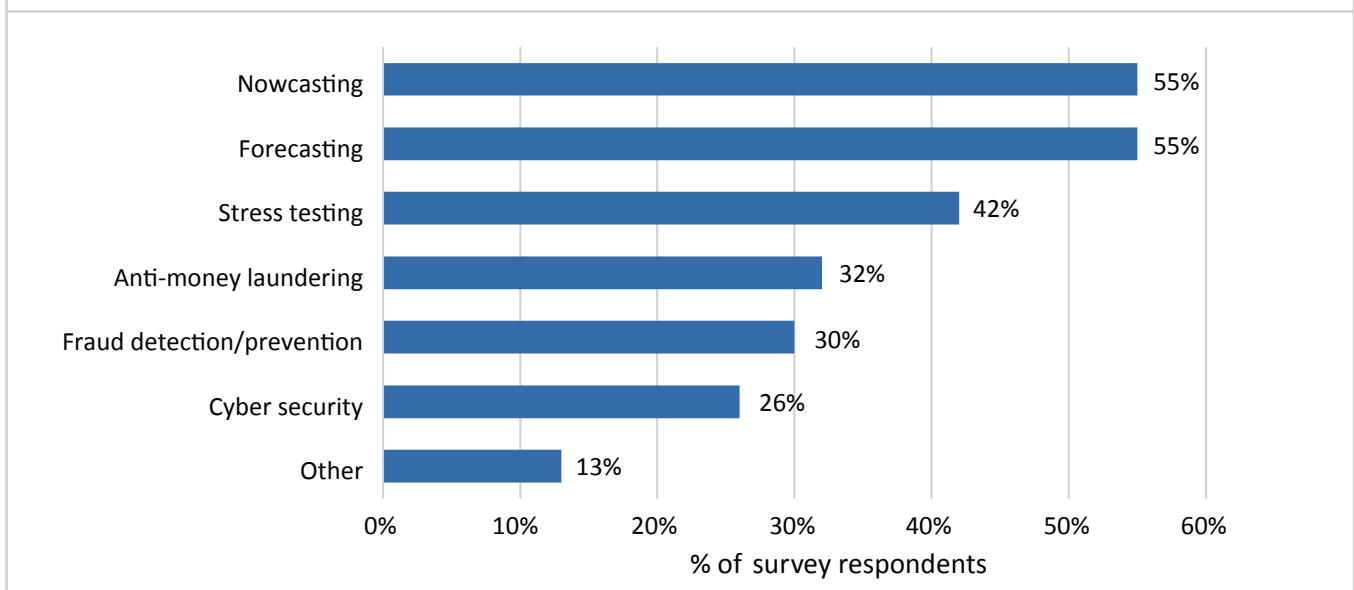
Central banks are using big data in a variety of ways to enhance their operations.
Source: Central Banking (2018).

financial institutions to disclose information could also have the unintended consequence of discouraging them from investing in technologies to collect and analyze data. Consumers might also raise concerns about the government gaining access to their personal data.

If the government manages to get big data from the financial sector, the next challenge would be to integrate these data with other data that the government collects, such as data from the real sector, to obtain a more complete picture. For example, policy makers might obtain data from the financial sector about firms that use export credit. Combining these data with other data (such as exports by firm) would enable governments to examine whether

export credit is helping exporters. Research institutions, such as academic organizations and think tanks, and international organizations could help policy makers in these efforts to organize and analyze data.

Overall, countries would benefit greatly from developing an information system that allows financial and economic information collected by different agencies to be matched and used more systematically. While building this system might be challenging and take time, the developmental benefits might well be worth the costs. A more thorough understanding of the financial sector can prove of immense value in the quest for building a more developed, competitive, and stable financial sector.

References

Aladangady, A., S. Aron-Dine, W. Dunn, L. Feiveson, P. Lengermannm, and C. Sahm, 2019. "From Transactions Data to Economic Statistics: Constructing Real-time, High-frequency, Geographic Measures of Consumer Spending." NBER Working Paper 26253, National Bureau of Economic Research.

Asrow, Kaitlin, and Joanne Xu. 2018. *The 2018 California Consumer Privacy Act: Understanding Its Implications and Ambiguities.* Fintech Edge Special Report. Federal Reserve Board of San Francisco

Athey, S. 2017. "Beyond Prediction: Using Big Data for Policy Problems." *Science* 355 (6324): 483–85.

Barocas, S., and A. D. Selbst. 2016. "Big Data's Disparate Impact." *California Law Review* 104 (3): 671–732.

Bartlett, R., A. Morse, R. Stanton, and N. Wallace. 2019. "Consumer-lending Discrimination in the FinTech Era." NBER Working Paper 25943, National Bureau of Economic Research.

Begenau, J., M. Farboodi, and L. Veldkamp. 2018. "Big Data in Finance and the Growth of Large Firms." *Journal of Monetary Economics* 97 (C): 71–87.

Bloomberg Law. 2019. "AI Regulations Aim at Eliminating Bias." May 31.

Business Insider. 2013. "How Marketers Use Big Data to Prey on the Poor." December 19.

Carrière-Swallow, Y., and V. Haksar. 2019. "The Economics and Implications of Data: An Integrated Perspective." Departmental Paper No. 19/16, International Monetary Fund.

Central Banking. 2018. *Big Data in Central Banks Focus Report* 2018.

Chataing, M-S., and K. Kushnir. 2018. "How Is Ant Financial Closing the SME Finance Gap in China." *SME Finance Forum Blog,* June 7.

EBA (European Banking Authority). 2016. "On the Use of Big Data by Financial Institutions." Joint Committee Discussion Paper JC/2016/86.

Economist, The. 2015. "Cool, Man." May 7.

Economist, The. 2019. "South Korea is Trying to Make Banking Fun." May 2.

Fintech News. 2017. "Alternative Credit Scoring Providers in Southeast Asia You Have to Know." June 9.

Fitzgerald, M. 2014. "DBS Bank Pumps Up the Volume on its Technology." Interview with David Gledhill, DBS Bank. *MIT Sloan Management Review*, January 6.

Forbes. 2015. "Turkey's Garanti Bank Captures Elusive Millennials." June 15.

---------. 2017. "3 Massive Big Data Problems Everyone Should Know About." June 15.

FSC (Financial Services Commission). 2019. "FSC Announces Plans to Establish Financial Big Data Infrastructure." June 3.

FTC (U.S. Federal Trade Commission). 2016. *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues.*

Greig, J. 2019. "How More Countries Plan to Pass Stringent Privacy Laws in 2019." *TechRepublic,* June 25.

JP Morgan Chase Institute. 2015. *Weathering Volatility: Big Data on the Financial Ups and Downs of U.S. Individuals.*

Korea Joongang Daily. 2019. "Big Data Means Big Money." May 20.

Lucker, J., S. K. Hogan, and T. Bischoff. 2017. "Predictably Inaccurate: The Prevalence and Perils of Bad Big Data." *Deloitte Review* 21 (July): 8–25.

Mastercard Foundation and IFC (International Finance Corporation). 2017. *Handbook: Data Analytics and Digital Financial Services.* Partnership for Financial Inclusion.

O'Donnell, B. 2017. "We All Thought Having More Data Was Better. We Were Wrong." *Vox,* April 12.

OECD (Organisation for Economic Co-operation and Development). 2016. "Big Data: Bringing Competition Policy to the Digital Era." Background Note by the Secretariat. DAF/COMP 14.

SNS Telecom & IT. 2018. *Big Data in the Financial Services Industry: 2018–2030—Opportunities, Challenges, Strategies & Forecasts.*

Thales. 2018. 2018 *Thales Data Threat Report.*

U.S. Department of the Treasury. 2018. *A Financial System that Creates Economic Opportunities Nonbank Financials, Fintech, and Innovation.*