McKinsey Working Papers on Risk, Number 30

New credit-risk models for the unbanked

Tobias Baer
Tony Goland
Robert Schiff

March 2012

Contents

New credit-risk models for the unbanked

Introduction

Lending to lower-income households and small and informal enterprises is challenging. Many (though not all) of these customers have limited familiarity with formal financial services, which inhibits their ability to make good decisions about the responsible and appropriate use of credit. And lenders often have little to none of the data they might traditionally use to make sound lending decisions (for example, official proof of income and a credit history).

In this environment, most lenders have pursued one of two business models:

- **Traditional consumer finance**, where lenders typically compensate for higher default incidence with higher interest rates while supplementing revenues with penalty and other fees

- **Microcredit**, with its focus on informal, usually women-owned microenterprises, shorter-term loans (often of just a few months), labor-intensive but strong relationships between loan officers and borrowers, use of joint-liability groups (in which members guarantee their fellow borrowers' loans), continuous lending cycles with borrowers taking on a new loan as soon as the last one is paid, and default rates as low as 2 to 3 percent

Neither of these models is ideal for cost-effectively serving the diverse needs of economically active lower-income families and businesses on a truly sustainable basis. But there is an opportunity for lenders to chart another path, using increased computing power and new sources of information and data (including mobile-phone usage patterns, utility-bill payment history, and others) to build better risk models. With these assets, and with scrupulous attention to privacy laws and customer preferences, banks, retailers, utilities, and telecommunications providers can make responsible lending decisions in low-touch and low-cost ways.

At its best, this third path, beyond traditional consumer finance and microcredit, can help companies profitably serve vast unbanked populations, and can also help societies move toward an elusive goal: full financial inclusion. Today there are more than 2.5 billion people without access to formal financial services, and there are hundreds of millions of micro, small, and midsize enterprises with unmet financing needs.[1] Ensuring that these people and groups have access to a range of quality, affordable, and appropriate financial services is on the agenda of governments and businesses worldwide. New uses of data and information move us closer to this vision by enabling a more complete understanding of households' financial needs. With that understanding, providers can move beyond simple lending to achieve several aims:

- Help customers make good financial decisions (for example, through the use of targeted alerts)

- Offer the right noncredit products (various forms of savings and insurance, for instance)

- Conduct marketing and communications in ways that are more likely to resonate for distinct segments

These are exciting areas that are ripe for further innovation. But lenders must first master the tools, data, and information that underpin the new approach to lower-income lending. The efforts of some pioneering lenders are quite promising:

---

1   Alberto Chaia, Tony Goland, Robert Schiff, et al., "Half the world is unbanked," Financial Access Initiative and McKinsey & Company, October 2009; Alberto Chaia, Tony Goland, and Robert Schiff, "Counting the world's unbanked," *McKinsey Quarterly*, March 2010; and Tony Goland, Peer Stein, and Robert Schiff, "Two trillion and counting," International Finance Corporation and McKinsey & Company, October 2010 (updated version to be published in 2012).

new alternative data models have cut credit losses in experimental forays into lower-income segments by 20 to 50 percent and doubled their application approval rates.

New data, new uses

Lenders need to look into the future to determine whether to make a loan. It used to be that the most reliable way to predict the future was to review the past; for centuries, long-standing banking and credit relationships provided banks with a reasonable basis for extending credit.

In developed markets, this approach has evolved over the past few decades: data such as credit reports and salary history now help lenders to make the same kind of predictions but on a larger scale and in an increasingly automated way. These credit-scoring approaches typically assess three characteristics:

1. Identity, to reduce fraud

2. Ability to repay, based on income and current debt load

3. Willingness to repay, usually based on past credit performance

These methods are less effective, however, in emerging economies, and especially among lower-income segments. Because lower-income customers often have no access to formal financing, there is no record of past borrowing behavior. Debt capacity is hard to judge because most lower-income workers are often paid wages in cash and have little or no formal savings or registered assets that could be used as collateral. Many do not receive regular fixed wages, but rather are self-employed or depend on a portfolio of income-earning activities that is inconsistent by nature.

Other problems also crop up. Lending to lower-income consumers is fraught with controversy. Negative press about aggressive consumer-finance practices, overlending by microfinance institutions, and the aggressive lending practices seen in the recent subprime mortgage crisis in the United States all point to the substantial challenges of serving low-income people responsibly.

Some nontraditional lenders, however, are successfully getting around these problems. A handful of pioneering mobile operators, utilities, retailers, and direct-sales companies are using new approaches to tap into the new forms of data spun off from their core businesses.

These new approaches have their own challenges. Traditional credit scoring draws on a thin stream of data collected monthly from a small number of sources (for example, credit cards, savings accounts, pay stubs, and mortgages). The new nontraditional data, on the other hand, must sometimes be gathered from diverse sources, and the volume is often several times that of traditional sources. For example, each mobile account may generate hundreds or even thousands of calls and text messages per month, each carrying a rich data set that includes when the call was made, the location of the caller at the time of the call, who received the call, the type of information accessed via text messaging, and the types and number of payment transactions made through the device.

That poses difficulties for risk modelers. While some new technologies are throwing off reams of data, others are allowing us to collect, aggregate, and analyze them in ways never before possible. There are new data standards and

protocols, and new tools to bring together disparate data sets, matching and comparing them to generate insights. Many practitioners are not yet skilled in these and are unfamiliar with aggregating diverse and oblique data to derive meaningful insights. For example, an organization that wants to use data gathered from mobile operators, grocery stores, and utilities will probably need to have expertise in each of these sectors to determine which data are meaningful, what level of detail is optimal, and what combinations of data are most effective.

Gaining access to data can be difficult as well. In many cases, the data sets that lenders want will be owned by entities (telecommunications companies, utilities, or retailers, for instance) that may not want—or are not allowed—to share them. They may be disinclined to take the risk of offending their customers by sharing the information, and they may not have an immediate incentive to find ways to share it, even with their customers' consent. Regulatory requirements and privacy laws may prohibit lenders from gaining access to certain types of information. (See "The benefits of strong privacy frameworks".) Governments are likely to be particularly cautious about sharing identity and other information that they collect about citizens.

The benefits of strong privacy frameworks for lenders and borrowers

In markets where privacy laws are well established, lenders and their partners must take great care to ensure compliance. And they must ensure that the practices they establish do not have unintended negative repercussions on society, particularly when using data about individual behavior as discussed earlier (for example, if information about health-services usage were employed as a marker of riskiness).

In other markets, where policy is being actively shaped, lenders and their partners should work with regulators to develop privacy laws that protect individual rights while also empowering consumers to make their own data available to lenders as a way to improve financial access. There are powerful benefits to lenders and others from helping in the creation of privacy standards—not least the foreknowledge of events that will help them avoid being caught by surprise developments. Only the firms that act in bad faith will benefit from a "Wild West" scenario. Responsible firms have everything to gain from privacy regulation with clear rules.

Consider some of the potential effects of strong privacy rules. Use of those data related to "positive" behaviors, for example, can prove beneficial to all. Scoring models could reward potential borrowers for desirable behavior, such as when an individual purchases health products or a farmer regularly tracks the weather or price of seeds via a mobile device. Rewarding such behaviors can generate other positive outcomes. For example, a person who knows that purchasing health-enhancing products such as vitamins and household disinfectants could actually help his or her credit score may be more likely to do so.

Some might complain that people would "game" the system, calling the weather line just to make their credit scores rise. But is that really so bad? While gaming can be a major issue with some types of data, and has to be managed carefully, this example shows that it often simply reinforces positive behaviors. Once the farmer calls the weather line, he is likely to get that crucial weather warning that triggers him to bring in the harvest before the rain, and hence materially reduces the risk of a devastating loss. This is not much different from a driver who wears a seat belt only to avoid paying a fine; the law may have failed to change the driver's risk attitude, and the driver may have worn the seat belt for the wrong reason, but the result is the same, and the law has achieved its goal of fewer deaths.

Effective modeling

In our experience, organizations should tackle these challenges one by one and pursue three steps to develop effective credit-scoring strategies that will help them lend to economically active lower-income households and enterprises at scale:

1. Identify promising data sources

2. Secure access to appropriate data

3. Convert data into credit insights

Throughout this process, it is important for lenders that are mining their data to exercise careful judgment about what constitutes responsible lending to avoid harm to their reputation, a consumer backlash, or worse.

New data from six sources

A. Telecommunications providers

In many emerging markets, data about payments received or sent via mobile phone are excellent proxies for income as well as ability to repay. Some users receive wages or social payments from the government in savings accounts accessed via mobile devices, and an increasing number of lower-income consumers can deposit funds into accounts that are accessed using mobile devices. Customers pay bills, make purchases, and send and receive remittances with such devices; all of these actions generate data that can be predictive of the customer's ability to repay loans.

Of course, mobile users also make calls, send text messages, and access information using their mobile devices. Subject to privacy laws, of course, some of this data can be used as a proxy for the likelihood of repayment, since probability of default can correlate with communication habits. For example, a farmer who regularly accesses information about the weather using a mobile phone might be expected to have higher crop yields, which would bolster his ability to meet his credit obligations. Customers who call large numbers of people with regular patterns may be more reliable than someone with few connections.

B. Utilities

Utilities collect information about usage to calculate bills, and they maintain payment records for customers that indicate not only how much is consumed but also how bills are paid, as well as whether and how often bills have been paid late or not at all. This data can serve as an excellent proxy for willingness and ability to repay, particularly because utility customers are typically billed monthly or at regular intervals that resemble repayment cycles used by many lenders.

C. Wholesale suppliers

Wholesale suppliers maintain similar payment histories for their small-business customers. Such data can be used not only as a proxy for ability to repay but also to estimate revenue. The data can reveal the demand patterns that drive revenues for small businesses over the course of a year, and a business's payment patterns can function the way credit reports do in more traditional credit contexts.

D. Retailers

Retailers can also be a rich source of data as POS devices and retailer loyalty programs become more prevalent throughout emerging markets. Data about customer purchases can enable providers to estimate income levels. Privacy laws permitting, for example, purchases of items such as vitamins and household disinfectants, which suggest a certain level of health consciousness, can provide valuable clues to potential borrowers' propensities for risk taking, which can then be used to estimate probability of default.

E. Government

Government agencies collect vast data sets about citizens to inform decision making and administer public policy, including social programs. Many developing countries are building systems that resemble the US Social Security identification system. These produce databases of demographic information—including date of birth, place of residence, and type of employment—that can be useful for estimating default risk, when its use is permitted. Governments may also maintain census records that can be used to predict risk, including the average income of individuals living in particular areas. And agencies that administer payments and other benefits usually maintain records of disbursements that could prove useful to lenders for estimating income levels.

In 2009, for example, the government of India enlisted Nandan M. Nilekani, co-founder of outsourcing giant Infosys, to launch an ambitious effort to provide an ID card to each of India's 1.2 billion citizens. The card will come with a unique identification number similar to the Social Security number assigned to US citizens. It will also include an embedded microchip with biometric data (for example, fingerprints) used to verify identity. The government hopes to collect a wide variety of data on the card, from facial scans to financial-account information, much of which would also be accessible via a central repository. The program was launched in 2011 and aims to provide 600 million IDs within the first five years of operation.

F. Financial institutions' own previously overlooked data

Lenders can also look internally for previously ignored data. One successful lender found that it actually already had paper records of interactions with some small businesses it wanted to target; it had overlooked the records because they were not in a digital format. A team manually entered the data into an electronic system, enabling the provider to conduct the analysis required to identify qualified borrowers. Other providers have taken similar steps, such as conducting interviews with customers of a new partner organization (for example, a wholesale supplier) to collect necessary data.

Sometimes the answer may be even simpler. Many large banks still do not take into account information as simple as the balances and transaction patterns of their own customers' checking and savings accounts when assessing creditworthiness.

1. Identify promising data sources

Lenders should look for data that can be used as reliable proxies for identity (for example, to reduce fraud), ability to repay (for instance, income or current debt load), and willingness to repay (for example, past credit experience). Six data sources in particular should be assessed: telecommunications providers, utilities, wholesale suppliers, retailers, government, and financial institutions' own previously overlooked data.

Consider mobile phones, which have become ubiquitous. By the beginning of 2009, emerging markets accounted for approximately 75 percent of the world's four billion mobile phones.[2] Each of these mobile-phone accounts provides a particularly rich potential source of data. Virtually every detail about each call, text, and request for information a customer makes is captured and stored by mobile operators.

For example, prepaid-minute purchase patterns can indicate a steady or uneven cash flow, and the timing and frequency of calls and text messages can indicate whether someone is working a steady job (for example, fewer calls between 9 a.m. and 5 p.m. may indicate that someone is working during those hours). Another example: the proliferation of data from mobile payments can provide credit underwriters with rich transactional information for generating credit insights.

Other technologies are also generating considerable raw data. Basic customer life-cycle management (CLM) applications are becoming increasingly commonplace throughout emerging markets, enabling businesses to collect information about the frequency and character of their interactions with customers. Point-of-service (POS) devices are used with increasing frequency by retailers of all kinds to gather transaction data. Retailer loyalty cards can provide important insights into consumers' income and even family structure (for example, buying diapers or school supplies is a good indication of children at home). And governments are developing improved identification and tracking systems for their citizens, to improve delivery of government services, among other things.

"New data from six sources" explains the new sets of data available from each sector.

2. Secure access to appropriate data

The simplest way to gain access to data is simply to pay for it. This can work well in some circumstances, but it is not ideal insofar as it increases costs. And for regulatory or other reasons (retailers' POS data for example is not typically available for sale), paying for access may not be possible. A better solution may be to strike partnerships with organizations to gain access in ways that provide benefits to all parties.

Many companies that do not provide financial services can benefit from partnering with organizations that do. Mobile providers can significantly improve their value proposition to customers by making credit services accessible through mobile phones. Not only will this help increase customer loyalty, it can also lead to higher revenues by driving more phone use. Mobile operators can also ask for their bank partners' help to better assess the credit risk for moving customers from prepaid to postpaid. If regulation inhibits mobile operators from sharing data freely, another avenue is to prompt mobile customers to seek loans of their own volition. For example, mobile providers can assess their customers' risk profiles and send credit solicitations to those who might qualify for loans. Customers can then decide for themselves if they want to share their information with the lender as part of the application process.

Lenders can strike partnerships with utilities and wholesalers to achieve the same ends. (Or to put it another way, utilities and wholesalers can build a lending business themselves, with varying degrees of help from a bank.) Lenders can also partner with retailers to accelerate the adoption of POS and other technologies that collect valuable customer

---

data. Partnerships of this kind are already forming in several parts of the world. For example, retailers can offer credit to make in-store purchases, which may drive greater purchase volumes and deepen relationships with consumers. The technologies used to provide credit can then double as transaction-tracking devices. Credit providers can also support the development of reward cards for consumers that will enable retailers to collect information about customers' purchasing habits.

Partnerships with governments can be particularly beneficial, since government agencies collect so much data about citizens. Governments can benefit from the opportunity to further financial inclusion. Already, many governments promote credit bureaus as a way to both enable smarter lending and protect consumers from overextending themselves.

Governments are rightly cautious about sharing citizens' information. See "The benefits of strong privacy frameworks for lenders and borrowers" (p. 3) for further discussion.

3. Convert data into credit insights

Many consumer lenders have advanced credit-risk modeling capabilities. But incorporating these kinds of new data will require some big changes in people, technologies, and approach. Three key areas for change are talent, IT, and the collaboration between risk and marketing teams.

Talent. Even the strongest lenders will likely have to develop new skills to create risk models that capture the predictive potential of available data. Providers should involve people from several areas; primarily, this means statisticians with the technical skills to analyze risk variables, and it could mean adding to the army of PhDs that many banks already have on staff. However, even the slightest reduction in loss rates will justify these hires. For institutions without advanced skills, the benefits are commensurately larger.

Other resources that lenders should tap into are experts in the business areas from which the data are being drawn, potentially including those from partner organizations. Marketing professionals and sales leaders should be called on to contribute valuable customer insights and to identify ways in which the more sophisticated risk profiles can also be used to improve CLM and vice versa.

IT. Lenders will also need more computing horsepower. Typical credit-bureau data, including credit-line utilization, delinquency status, and credit inquiries, are based on a small number of financial events. But the new data, such as mobile records, might arrive by the gigabyte, even for a small number of prospective customers. Such data sets can easily overwhelm the software that many institutions use for statistical analysis. Lenders will need to invest in heavier-duty processing power and software. Recent developments related to cloud computing make this possible at a more reasonable cost than before.

Collaboration between risk and marketing teams. Lenders should also develop a road map for working with nontraditional data. Many credit-risk teams have been doing essentially the same thing for decades. They have developed long lists of standard transformations—such as calculating growth rates or using a dummy variable to flag customers who have become overdue at least once in the past. But when confronted with a table of the 300 payments, 2,000 phone calls, and 4,000 text messages a prospective customer made in the past year, credit-risk teams will need a thoughtful plan to convert data into insights.

A good start is for companies to turn to internal consumer-insights specialists, often drawn from the sales and marketing organization. A good retail expert might point out that cash-strapped people buy small, round-numbered amounts of gasoline, while larger, odd-numbered purchases are made by people unconcerned about the cost of filling up. Similarly,

Credit-risk innovators

A growing number of lenders, in both developed markets and especially in the emerging markets that are home to many unbanked customers, are using nontraditional sources of data to their advantage.

■ Brazilian wholesaler Grupo Martins started a joint venture with a Brazilian bank to lend to small consumer-goods retailers. The joint venture subsequently became a stand-alone bank, Tribanco. Within a couple of years, Tribanco contributed about 50 percent of Grupo Martins' consolidated net profit; the experiment proved transformational for the company.

■ An Asian consumer lender started by assembling a team of marketing professionals that created profiles based on sets of risk attributes that are characteristic of particular groups or types of individuals, such as "upwardly mobile migrants" (a desirable group) and "daredevils" (people to avoid). The marketing team then worked with statisticians to analyze widely available data such as ID and phone numbers to find creative ways to generate insights about customers, such as their migration history. This was helpful because it had been determined that some migration paths were associated with significantly lower credit risk than others.

■ The same Asian lender also found that delinquencies on mobile-phone bills were 60 percent more predictive of eventual small-loan defaults than were delinquencies on loans from other banks. Even the choice of payment plan for the phone bill, a seemingly unimportant variable, was found to be just as predictive as the second-best variable available from the credit bureau.

■ A Latin American consumer lender found that analyzing recharge behavior—the pattern in which a consumer "tops up" her prepaid mobile phone account—more than doubled its predictive power compared with the use of simple demographic data alone.

■ Another lender found that the contractual terms and resulting cash flows between small businesses and their key suppliers was extraordinarily predictive. With a Gini coefficient of 35, this factor compared favorably with the best variables offered by developed-market credit bureaus (these variables typically range from 25 to 45).

■ Alibaba, the Chinese B2B Internet company, has created a credit score it calls "Ali-loan," derived from the transactions conducted on its portal. It then sells this score to lenders. With a bad-loan rate of just 0.35 percent, this data source is something that makes even developed-market lenders envious.

■ The giant Brazilian food producer JBS launched Banco JBS in 2008 to provide financing to farmers and other suppliers.

All have tapped the power of nontraditional data to provide financial services to creditworthy lower-income households and enterprises that would have been much harder or even impossible to serve under more traditional approaches.

people who buy gas on holidays, when prices are higher, are exhibiting less frugal behavior than people who economize by buying on the days before a big holiday weekend. Between two customers who earn exactly the same amount of money, the latter behavior signals greater capacity to take on a loan.

Greater communication between the risk and marketing teams may also invert the identification of new segments to serve. In addition to the traditional approach of the marketing team identifying the customers to serve and asking the risk team to build credit models accordingly, the credit-risk team can now flag for the marketing group new data sources that increase the feasibility of serving those customer segments that were not previously a priority.

"Credit-risk innovators" (on p. 8) shows how some leading players have brought these ideas together.

□□□

Institutions that aspire to serve lower-income communities should ask themselves three questions:

- **Strategy and capabilities.** What is my data-enabled risk-management strategy, and do I have the right capabilities to deliver against it?

- **Competition for partners.** Who are the right partners, and is my company at risk of losing out if a competitor beats us to the best partners (for example, mobile operators, retailers, wholesalers, or utilities)?

- **Better outcomes for borrowers.** How can we use new skills in alternative credit-risk assessment to ensure that we are lending in appropriate ways and supporting consumers to make good financial decisions (for example, through the use of targeted alerts, offering the right products beyond credit)?

Financial institutions and other organizations with answers to these questions have an opportunity for significant social impact and will create meaningful value for their enterprises.

Tobias Baer, based in McKinsey's Taipei office, leads our global credit risk analytics team. **Tony Goland** is a director in the Washington, DC, office. **Robert Schiff** is an associate principal in the New York office.

Contact for distribution: Francine Martin
Phone: +1 (514) 939-6940
E-mail: Francine_Martin@mckinsey.com

McKinsey Working Papers on Risk

McKinsey Working Papers on Risk