THE INITIATIVE FOR SMALLHOLDER FINANCE

BRIEFING 11

THE RISE OF THE DATA SCIENTIST: How big data and data science are changing smallholder finance

February 2, 2016

THE INITIATIVE FOR SMALLHOLDER FINANCE

THE RISE OF THE DATA SCIENTIST:

How big data and data science are changing smallholder finance

The enormous gap between the supply and demand of formal credit for smallholder farmers is caused, in part, by the extreme lack of information available to lenders on potential borrowers. These information asymmetries create risks for financial institutions and limit their willingness to lend in the agriculture sector. Though less pronounced, this data challenge also exists in other low-income segments in developing countries, as well as in developed countries, where credit assessment of "thin file" borrowers is difficult.

To overcome these challenges, innovative lenders are incorporating new data sources and analytics into their credit assessment models. Lenders in the developing world have experimented mostly with mobile phone usage data, as well as data from e-wallets and mobile money platforms, to help them reach unbanked populations. These experiments have unlocked credit in underserved markets in the developing world, but so far, mostly for younger, tech-savvy populations in urban areas. The heavy reliance of these models on mobile phone data makes them better suited for urban lending contexts. The small value, high velocity loans also reflect the needs of this market – financing for trade inventory, unplanned expenses, and purchase of consumer goods.

While credit scoring models that use alternative data are still new in the agriculture sector, several trends suggest this market will soon be better served. The increasing penetration of smartphones and improved rural connectivity will increase the availability of digital data there. This is enhanced by the increasing digitization of informal data such as from saving and lending circles common in rural areas. Finally, data scientists are also conducting inductive analyses, along with deductive, to test the value of previously unused data for risk scoring.

Going forward, a focus on three key areas will facilitate the adaptation of current models to the particular needs of the agricultural sector:

- **More collaboration with non-traditional actors that have data on agricultural borrowers,** such as technical assistance providers, agricultural and climate research institutes, and value chain aggregators.
- **More digitization of data** from, for example, subsidy programs, censuses, and extension worker data could be useful for borrower verification and credit algorithms in the agriculture sector.
- **More context-appropriate investment** that is aligned to the longer product development and testing cycles in the agriculture sector: 3-4 year timelines are more appropriate than 1-2 year timelines.

ABOUT THIS BRIEFING

This briefing note seeks to supplement the growing body of literature on alternative applications of data in assessing credit risk with a discussion of the opportunity to extend these models to the agriculture sector. We profile the current state of alternative credit assessment models apart from and with respect to the agriculture sector, and lay out the challenges that prevent an easy transfer of urban-tested models to agricultural borrowers. We highlight efforts to develop new data sources for the agriculture sector, as well as trends that may support alternative data-driven credit assessment in these environments. The note concludes with recommendations for how lenders, donors, and other partners can push forward the development of such alternative credit assessment models.

Introduction: The lending challenge

The challenges of lending and financial service provision in low-income, data-scarce environments are driven in part by information asymmetries that create real or perceived risks for financial institutions.[1] These asymmetries can make it difficult for lenders to identify, understand, and assess their clients, as well as serve them with appropriate products. At the most basic level, financial institutions need to be able to answer two questions when making lending decisions:

- Is the borrower creditworthy? That is to say, how likely is the borrower to default on a loan?
- How much credit can the borrower take on without creating or increasing the risk of default?

Without the right information to accurately make these assessments, financial institutions may be led to lend to fundamentally riskier clients (known as "adverse selection" in economic terms) or to clients whose behavior may change once a loan is obtained (known as "moral hazard").

The risks created by these information asymmetries have historically resulted in low levels of financial inclusion for low-income populations. In rural and smallholder environments, where data on potential borrowers is even thinner and more difficult to access, these risks are compounded, and access to credit is even lower. Estimates suggest that only 3% of the global demand for smallholder financing is met by formal financial institutions.[2]

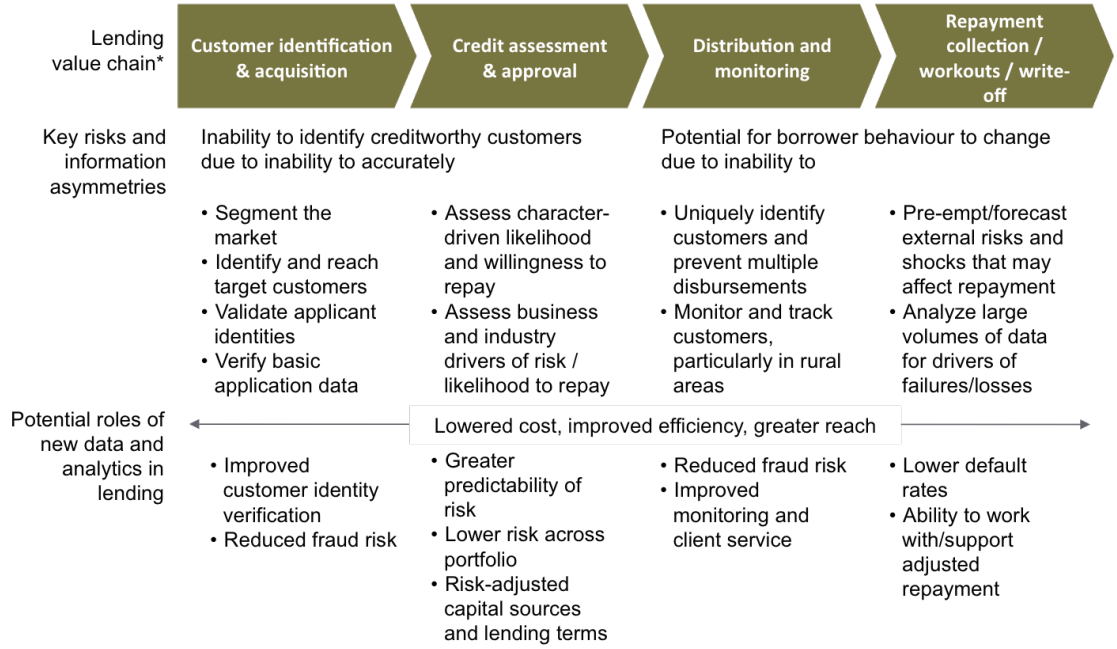
The data opportunity

Two recent innovations have the potential to dramatically reduce these information challenges and the lending risks associated with them, which could help open up pools of capital for previously underserved populations:

- The ability to collect and collate new data sources ("alternative data") both digital and non-digital
- The ability to analyze and apply this data through new methodologies ("data science")

While new data and analytics cannot address all of the constraints to smallholder financing, these tools can help lenders overcome key information constraints in the lending process. Namely, they can be used to significantly improve identity verification and credit assessment processes, as well as enable improved disbursement and post-credit servicing, as shown in Figure 1.

Figure 1: How data and analytics can help lenders overcome key information constraints



*Lending value chain modified from Financial Services value chain developed by ISF/MCF for Value Chain Financing for SHF study
Sources: The Omidyar Network, "Big Data Small Credit: The Digital Revolution and its Impact on Emerging Market Consumers", 2015

By reducing informational asymmetries and, in turn, lowering lending risk, these innovations could play a major role in unlocking lending. A recent estimate by the Omidyar Network suggests that innovations that enable cost-effective data collection and analytics for underserved customers could help lenders "reach between 635 million and 1 billion emerging-market consumers who are currently unable to access formal credit."[3]

Furthermore, given the lower costs and inherent scale of these new approaches, these customers could also be reached with greater efficiency and at a significantly lower cost than traditional credit assessment methods allow. A study commissioned by CGAP estimated that using alternative data and new analytics could help lenders reduce the cost of delivering microloans by 20%-30%, with the bulk of the savings coming from lowered underwriting, application, risk, and collection costs.[4]

Despite widespread excitement about their potential impact, we must remember that applications of alternative data to credit scoring are still new; most have been developed in only the last five years. However, we, along with other industry observers, expect to see enormous evolution in the field over the coming years as models are refined to account for the different needs, data contexts, and drivers of risk within developed and developing country markets.

Applications in the developed world

In the developed world, alternative data-based models focus on serving underbanked populations, borrowers with so-called "thin files."[5] Leveraging the vast digital footprint of today's consumer, these models use alternative data to supplement traditional credit data.

This combination of traditional and alternative data for credit assessment is being deployed by firms like Earnest, Affirm and UpStart in the US; Paidy in Japan; Lendable, Klarna, and Kreditech in Europe. These innovators almost all target consumers with a strong digital footprint that can provide additional data for risk assessment.

Ideal borrowers for these models are financially responsible but struggle to access credit at an affordable cost, if at all, due to their limited credit histories. These ideal candidates are considered "future prime" borrowers: young professionals and recent graduates whose credit scores do not fairly reflect their creditworthiness simply because they have not had enough time to build their credit histories.

These types of models rely heavily on digital data – online transaction and e-commerce histories, utility bill payment histories, social media networks and activity – along with education and employment histories and prospects. When applied in addition to traditional credit reference data, this data helps enable highly accurate, predictive assessments of risk that drive the kinds of lending most in demand by this market segment. Student loan refinancing and short-term payday or personal loans are the primary focus of lenders like Vouch Financial, Zest Finance, and Lendup. E-commerce loans for consumer goods commonly sought by this target segment (for example, home furniture, cars, and wedding dresses) are offered by lenders such as One Road Lending, Neo Finance, Kreditech, I-Do Lending, Klarna, Paidy, and Bristlecone Lending.

A third category of models provides credit assessments and loans for merchants and small businesses. Lenders like Kabbage, OnDeck Capital, CAN Capital, Square Capital, and PayPal Working Capital leverage digital data generated by the businesses that are asking for a loan. The data come from that business's transaction platforms and social media activity, and lower the risk of lending to small and growing enterprises.

Applications in the developing world

In contrast to the focus on the underbanked in developed markets, the alternative credit assessment models being tested in developing markets aim to address the credit gaps of the unbanked. To do so, lenders have had to create business models that engage new partners, and leverage different infrastructure platforms or data sources. They have had some early successes, particularly with models that target urban clients seeking relatively small loans.

Innovators and their business models

In developing markets, alternative data-based lending models are being built by various configurations of traditional financial institutions and new actors that see an opportunity to reach traditionally underserved market segments. This contrasts somewhat with the

profile of innovators in developed markets. There, innovation in credit assessment is commonly driven by serial tech entrepreneurs who have previous experience in data analytics, social media, e-commerce, or finance, and who are able to access venture capital to finance their experimentation with new lending models. Figure 2 on the following page illustrates some emerging archetypes of business models in developing countries, along with examples.

The banks and MFIs that dominate traditional lending rely on credit reference bureaus, collateral requirements, and social capital to reduce risk. However, new actors that generate and analyze data about the habits and behaviors of borrowers are redefining this lending landscape.

The opportunity to evaluate customers using new data has drawn in MNOs and telcos (with prominent companies, like Safaricom and Airtel, often leading the charge), as well as third party data analytics firms, like Cignifi, Demyst Data, and First Access. New organizations that interact directly with consumers such as off-grid energy companies, education providers and input suppliers, are also increasingly engaging with lending models as the data they can provide on consumers could add significant value to credit assessment algorithms.

Data sources and platforms

Currently, the majority of alternative credit assessment models in the developing world use mobile utilization data and e-wallet or mobile money data as their primary data sources. Some models also incorporate utility bill and social media data as secondary sources, and a few, mostly in China, use online trade platform data as a primary data source. Figure 3 on the following page illustrates the dominance of mobile-based alternative lending models.

Borrowers and products

Alternative data-driven lending models are heavily concentrated in urban environments, among a predominantly young, technologically savvy, educated, financially stable demographic – those termed the "early adopters."[6]

As illustrated above, the heavy reliance of these models on mobile utilization and mobile money data makes them well suited for urban environments, where the depth and breadth of mobile phone (and particularly smartphone) use creates strong digital footprints that can be analyzed.

The bias of lenders towards this market segment is also reflected in the types of financial products offered. Borrowers in these contexts tend to need relatively small (low value) loans with relatively fast turnover that can be used to finance trade inventory, attend to unplanned or emergency expenses, and purchase small consumer goods. This particular product, or use case, has resulted in the label "big data small credit" being used in other studies.[7]

Broadening and deepening: Is it possible?

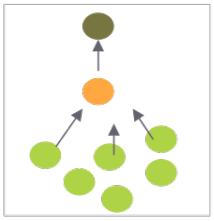
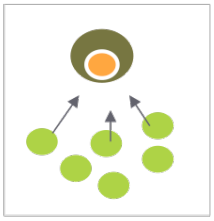
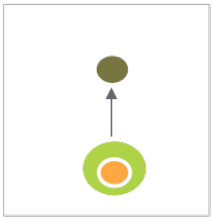
The question remains, to what extent will these new credit assessment models be able to reach underserved populations, particularly rural farmers? Despite the large number of such models being tested in developing countries, few are actively targeting agricultural borrowers. Gro Ventures, Farm Drive, and the Grameen Foundation have built models explicitly for the agriculture sector; and a few others, such as EFL and Arifu, are in pilot programs to adapt their models and test their relevance with farmers.

While these examples above are currently the exception, the broadening of mobile money services into rural areas suggests that alternative credit assessment models should also be able to penetrate these markets. However, the different financing needs and data contexts in rural areas mean urban models cannot simply be transposed. To reach these "late adopters", lenders will have to develop new lending approaches and data sources that better reflect the needs and risks of agricultural borrowers.

In particular, three features of the agricultural market pose challenges for lenders and the development of credit assessment models:

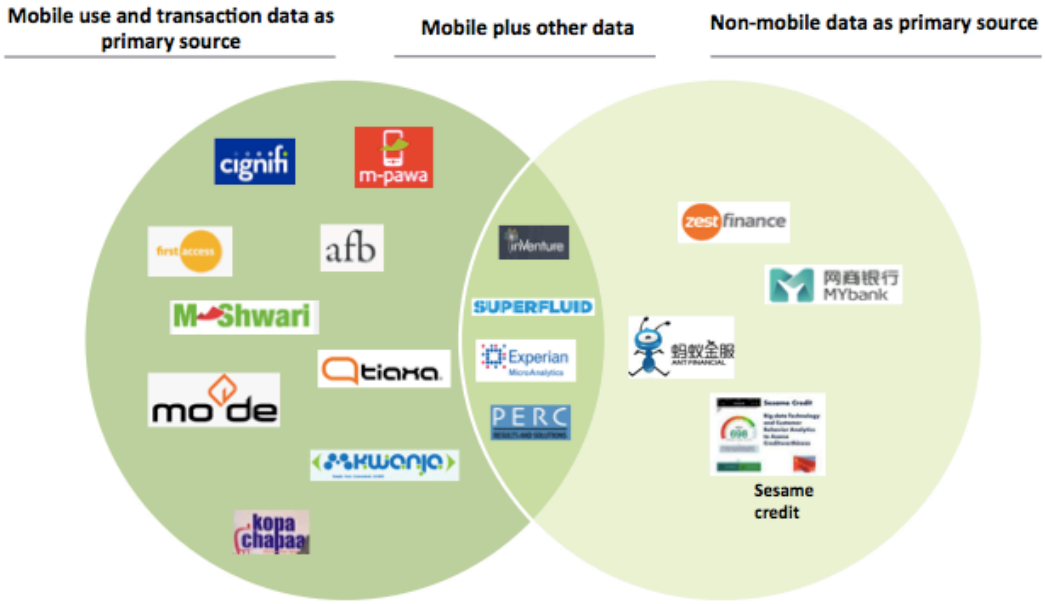
1. The need for larger loans of longer tenor
2. The weaker and often unreliable digital footprints of rural populations
3. The varied drivers of default risk for smallholder farmers

Figure 2: Archetypes of alternative data-based lending models in developing countries

				
Description	Analytics provider sources and aggregates data, conducts analysis, and sells it as an ongoing service or once-off product to a lender/ decision-maker	Lender has in-house analytics capability and models, but accesses or buys data generated by other players in the ecosystem	Data generator recognises the potential of data collected, and builds own analytics capability or close partnership to then serve a decision-maker	Data collection, analytics, and lending decisions housed within same entity. Lender may start analysing previously collected data for decision-making purposes or data generator begins lending
Prevalence	Very high	High	Low	Low
Examples	• First Access • Lenddo • Superfluid Labs • EFL • Demyst Data • Cignifi • Tiaxa • Experian Micro-analytics / PERC • Farm Drive	• M-Shwari • M-Pawa • Branch • AFB • Kopa Chapaa • Gro Ventures	• Arifu • Agrilife*	• Ant Finance • InVenture • Grameen Foundation

Credit decision-maker/lender
Analytics provider
Data generator/collector

Figure 3: The dominance of mobile-based alternative lending models



The first challenge is one of product development. Agricultural borrowers tend to need higher value but lower velocity loans to accommodate large pre-season purchases and long lags in production, harvest, and marketing cycles. As such, the nano- to microloans available through new lending platforms do not meet the majority of farmer credit needs**,** though they can be useful for short-term income or cash flow smoothing. Whereas typical loans on mobile platforms range from $20-$70, farmers often require loans of $500 - $1000.[8]

Secondly, the weak digital footprints in rural contexts limit lender abilities to identify borrowers, validate their identity, and importantly, trace their activity patterns. This, in turn, makes it difficult to build credit histories if the models used rely heavily on mobile data. Digital footprints in rural areas are limited by the relatively lower levels of mobile phone access there (across sub-Saharan Africa 69% of urban dwellers have a mobile phone compared to 53% of the rural population[9]), but also because rural contexts impose different use patterns:

- **Low access to power for charging means that mobile phones are often turned off.** This results in utilization and transaction rates that are much lower than those in urban areas, and makes it difficult to gather historical usage or transaction patterns for analysis.
- **Limited ownership of handsets means they are often shared within a household or even among multiple households**, which complicates efforts to determine individualized use patterns and, in turn, to assess the creditworthiness of an individual.
- **Due to spotty network coverage, users are often unreachable**. Along with contributing to their lower utilization rates, as mentioned above, this also makes monitoring, retention, and client service difficult.
- **Frequent deactivation of SIM cards, due to inactivity, results in high rates of phone number churn**, which further complicates efforts to identify, track, and retain customers.

The larger size of loans required in the agriculture sector, combined with the lower predictive power of mobile utilization data, points to the need for lenders to understand, test, and incorporate different forms of data for the rural market. They will not be able easily adopt the portfolio approach that has proved successful in urban environments. There, using estimates of the risk of default across the portfolio and maintaining a minimum exclusion threshold for high-risk borrowers, lenders are able to disburse low value loans. As the risk of default is spread across a large group of borrowers, lenders can reduce the need for great amounts of individual-specific risk assessment. This approach, however, becomes problematic as loan sizes increase, as in the agriculture sector. To assess credit risk in that market, then, lenders need to develop detailed profiles of individual borrowers.

To obtain a more individualized client profile, innovators are looking beyond mobile data, and have started experimenting with data and analytics that generate a deeper understanding of the character of a potential borrower. Two types of data sources, in particular, have received significant interest: psychometric data and a combination of social media and social network data. The hypothesis behind both approaches is that character assessments can provide lenders with insight into both the willingness and predisposition of a borrower to repay his loan. As such, these models seek to protect lenders against the selection of borrowers who are unlikely to repay, as well as from those whose behavior may change once a loan is disbursed. While current applications of both approaches have largely focused on the urban market, the relevance of this data for larger loan sizes, combined with their low reliance on MNO data, points to their strong potential in rural contexts. Some considerations and examples are illustrated in Figure 4 on the following page.

For agricultural loans, a farmer's ability to pay may be driven by the production and price risks she faces. Factors such as weather, soil health, and input quality can affect production and yields while market prices, warehousing, and buyer linkages can affect sales. Innovators must understand the relative importance of these potential drivers, and incorporate data into their models that can adequately assess them. While not prevalent yet, some models have started to test the utility of various geographic and climatic data, as well as value chain and market data, for this purpose. Figure 5 on the following page provides some examples.

Figure 4: Using psychometric and social media/social network data to understand borrowers

Data category	Example indicators	Potential sources	Advantages	Disadvantages	Examples
Psychometric and behavioural	• Goal setting • Persistence through failure • Continual learning • Planning/ budgeting	• In-person interviews • Mobile-based surveys • Learning platform engagement patterns • Money management platform engagement and use patterns • Quiz performance	• Works for everyone • High value in context with low digital data • Deep character profile • Interview is a familiar stage in borrowing for most clients	• Time intensive – each interview takes 15-20 minutes vs. 20 seconds to download massive digital datasets • May require a threshold literacy level that may not exist in some rural contexts	• Entrepreneurial Finance Lab • Arifu • Visual DNA • Inventure
Social media/ social network	• Breadth of network • Frequency of connection • Depth of connection • Time and frequency of social media use	• App installed on smartphone • Access to public profiles	• Device-based therefore negates need for telco/ MNO data, which may be weak, limited, expensive, etc. • Powerful for identity verification (fraud prevention) and for minimizing adverse selection – key life events that could increase risk are easy to identify, e.g., redundancy, house fire, etc.	• Less predictive of credit default risk • Driven by smartphone penetration, which may be low (though growing) in rural contexts • Data may be considered sensitive or private	• Lenddo • Superfluid labs • Branch

Source: Initiative for Smallholder Finance analysis, organization websites, and select interviews.

Figure 5: Using geographic, climatic, value chain, and market data to assess borrowers

Data category	Example indicators	Potential sources	Advantages	Disadvantages	Examples
Geographic (physical and human)	• Location / agro-ecology • Farm size/herd size/crops • Soil health/ moisture • Climate patterns • Age and sex distribution • Socio-economic distribution	• Satellite data • Agriculture research institutes • GPS/geo-location applications	• Enables verification/ validation of application data without deployment of staff to visit • Enables assessment of exposure to climatic risks • Enables estimation of yield potential, which can underpin assessment of production risk	• Weather, agro-ecology, and satellite data often found at high levels of aggregation, and need to be disaggregated to the level of individual farms	• Telephone Farmers • Planet Labs • Digital Globe • Aware • Acre Africa • Getchee • IPRS • The Climate Corporation • Geotraceability
Value chain / market	• Input quality/ supply • Output/yields • Market prices	• Input suppliers • TA providers • Buyers/traders • Coops/Aggregators • Value chain support services	• Provides detailed insight into production and marketing potential and challenges, which can guide more accurate assessment of risk • Supplier/offtaker data enables identity verification	• Value chains are so unique and the risks, costs, opportunities of each so different that most need to be mapped individually – time and resource intensive	• Agricultural Loan Evaluation System • Agrilife • Hurudza • MFarm • Soko Shambani • Scanntech • Gro Ventures • Icow • FarmDrive

Source: Initiative for Smallholder Finance analysis, organization websites, and select interviews.

New trends that point to broader and deeper reach

As noted in the previous section, applications of alternative data for credit scoring in the agriculture sector are still in their early stages. Several trends, however, point to the potential for these tools to be refined to serve and significantly boost access to credit for agricultural borrowers:

1. **The increasing penetration of smartphones in rural areas, as handsets become more affordable and connectivity improves, should improve the quality and quantity of digital data on smallholder farmers.** The GSMA forecasts that smartphone penetration in sub-Saharan Africa will grow five-fold, from 4% to 20%, between 2012 and 2017.[10] This growth will be driven by a combination of declining handset costs and international and corporate initiatives to overcome infrastructure and affordability challenges. Global partnerships like the Alliance for Affordable Internet, Internet.org, and the GSMA's Digital Inclusion program complement industry investments to ramp up 3G and 4G capacity, and to develop access schemes for excluded markets.

2. **Data scientists are testing inductive and deductive approaches to understand the predictive power of different data types and sources, as well as potential correlations among them.** Some models are testing new or *particular* data sources to determine if these help resolve *particular* information gaps that limit lending. EFL, for example, is using psychometric testing to understand borrower character; Lenddo has taken a different approach to the same problem, prioritizing social media data. Other actors are working backwards, exploring huge data sets that have already been generated to test for correlations within them that could facilitate decision-making. The Grameen Foundation, for example, has pilot projects in Colombia and Uganda that are analyzing customer data, generated by several microfinance partners, for predictive patterns that could then be calibrated into credit risk assessments.

3. **Large informal data sources are increasingly being digitized, expanding the breadth and depth of data available as inputs in credit decision-making.** Applications that enable mobile or online management of informal savings and lending groups, such as eMoneyPool, ChamaPesa, and EcoCash Savings Club, are increasing transparency among group members, reducing the risk of misappropriation, and creating digital records of savings and borrowing patterns, which brings users closer to formal credit histories and formal lending. As these informal groups (variously called SACCOs, chama's, chit funds, or susu's) are the predominant provider of financial services in the agriculture sector, the digitization of their data could be a valuable resource for increasing formal credit to farmers. It is important to note that the digitization of paper-based data is expensive and labor-intensive. In order for these sources to maintain their value, the underlying data collection mechanisms must be digital; otherwise, the result will be a single digital data point that is difficult to update.

Extending and deepening: What is needed from here

The use of alternative data and new analytics has tremendous potential to impact agricultural credit scoring and, in turn, unlock lending in the sector, but we are still far from realizing that potential. Given the nascent nature of the field, we echo the calls being made by others for more investments in data science, more experimentation, more refinement of business models, and more collaboration among actors. This should be done within a context of strengthened consumer protection and careful observation by policymakers.[11] We also want to emphasize the need for a particular focus on the deeply underserved agricultural market.

The findings from this briefing point to three areas within the field where more can be done to extend the potential of alternative data-driven credit scoring to smallholder borrowers:

- **More collaboration with non-traditional actors.** The unique needs and weaker digital footprints in rural sectors suggest that lenders will need to access new types of data, which will likely come from sources they have previously not engaged with. Technical assistance providers, for example, may have large 'known' data sets on the identities and assets of smallholder farmers that could be tested deductively for their value in risk assessment. These providers may also have 'unknown' data about the behavior

and character of farmers that could be analyzed inductively to inform lending practices.

- **More digitization of data.** The digitization of informal financial data provides an example that could be replicated in other relevant spheres. If digitized, government data from subsidy schemes, censuses, and extension workers, for example, could be used to refine borrower verification and credit algorithms in the agriculture sector.

- **More context-appropriate investment.** In building credit assessment models for farmers, innovators may face product development and testing cycles that are significantly longer than those of urban models – timelines that should be considered by funders. Agricultural loans must be made in line with harvest cycles, meaning that each step of the innovation process could take up to a year, compared to a few weeks or months in other lending contexts. In agriculture, the six- to nine-month planning and development phase is followed by the first harvest cycle when data to inform lending algorithms can begin to be collected. In the second year, these algorithms can be piloted and calibrated; in the third, the models can be tested and used predictively; and only in the fourth year is any real scale built into the model. This timeline is somewhat simplified, but serves to illustrate that the 12-18 month funding cycles used for mobile data-based innovations may not be aligned to support innovation of agricultural products or services.

Conclusion

The rapid growth and development of credit assessment models that use alternative data has been impressive. As recently as five years ago, lenders relied primarily on credit history, income information, and relationship-driven assessments to determine the creditworthiness of borrowers. The idea of leveraging the growing volume of digital data and new analytic methodologies to enable more accurate risk assessments of unreached clients was only just emerging.

In the short period since then, we have seen this idea take off, with as many as 50 models in developed and developing countries exploring its potential to change traditional approaches to lending. In the developing world, where access to credit has long been a challenge, this opportunity is drawing non-traditional actors into the financial services ecosystem; MNOs, telecommunication companies, data analytics firms, and utility providers are increasingly active in efforts to build predictive risk algorithms and expand lending.

Already these models are having success in urban contexts with "early adopters". Mobile-based platforms have been unlocking huge volumes of lending– over $280M of loans have been disbursed through M-Shwari alone since November 2012.[12] Much remains to be done, but the power of these models to transform lending possibilities for millions of previously unserved consumers is undisputed, and the trends are positive.[13]

However, the course by which these models will move beyond urban areas to reach smallholder farmers and other actors in the agriculture sector remains uncharted. The weaker digital footprints and different financing needs of consumers in rural areas mean that lenders cannot simply transpose successful urban models. They will have to first improve their data on the sector. This will require engaging a wider range of actors that can provide data uniquely relevant to agricultural contexts and borrowers. Valuable partners could include education platforms, technical assistance providers, informal savings institutions, agricultural research institutes, weather and climate centers, value chain aggregators and support platforms, among others

Despite these challenges, all trends suggest that these models will transform the lending landscape even more dramatically over the next five years. The amount of relevant data on smallholders will grow as smartphone access increases in rural areas, and as data from a variety of formal and informal sources are gathered and digitized. Complimenting this, innovators are eager to extend their models into unreached segments, and pioneers like Gro Ventures, Farm Drive, Grameen Foundation, EFL, and Arifu have already begun to test and refine their data and analytics models for "late adopters."

While we cannot expect alternative data and new analytics to be a panacea for rural financial inclusion, we can, and do, expect that they will transform lending for this sector. With further experimentation, these tools will enable stronger client validation, improve the accuracy of risk assessments, lower lending costs, and improve client monitoring. We hope these advances lead to significant increases in access to credit for long-underserved rural and agricultural borrowers.

NOTES

[1] The other drivers of low levels of financial inclusion have been well documented elsewhere. See for example ISF Briefing Note 1 "Local Bank Financing for smallholder farmers: A $9 billion drop in the ocean"

[2] Estimates cited in ISF Briefing Note 1 indicate that the total amount of debt financing supplied by local banks to smallholder farmers in the developing world is approximately $9 billion. This represents less than 3% of the estimated total demand for smallholder financing, which excluding China, is $300 billion, and $450 billion globally.

[3] The Omidyar Network, "Big Data Small Credit: The Digital Revolution and its Impact on Emerging Market Consumers", 2015 (page 7)

[4] CGAP, "Projecting Impact of Non Traditional Data and Advanced Analytics on Delivery Costs", December 2014;

CGAP, "Big Data for Financial Inclusion: Is Boring Better", 2014

[5] "Thin file" consumers have limited credit histories or insufficient information on their credit reference reports.

[6] The Omidyar Network, "Big Data Small Credit: The Digital Revolution and its Impact on Emerging Market Consumers", 2015

[7] Ibid

[8] A cow, for example, could cost $800. Evidence from industry experts in Kenya suggests that even loans in the $500 range are unpopular with farmers; they would prefer larger ticket sizes.

[9] Gallup, "Mobile Phone Access Varies widely in sub-Saharan Africa", September 2011

[10] GSMA Discussion Paper "Smartphones and Mobile Money: The Next Generation of Digital Financial Inclusion", July 2014

[11] The Omidyar Network, "Big Data Small Credit: The Digital Revolution and its Impact on Emerging Market Consumers", 2015 (pages 26-30)

[12] CGAP and FSD Kenya, "How M-Shwari works: The story so far", April 2015

[13] For example, annual comparisons of M-Shwari users show significant reductions in the bias towards an early-adopter client base that is young, urban, male, already-banked, and above the poverty line. Ibid.

ACKNOWLEDGEMENTS

This report builds on the knowledge, experience, and research of many experts in the field. The findings and analysis in these pages would not have been possible without the following individuals who shared data, insights, and perspectives:

- Greg Chen, **World Bank/CGAP**
- Abdigani Diriye, **SuperFluid Labs**
- Daniel Goldfarb, **Lendable**
- Craig Heintzman, **Arifu**
- Lisa Kienzle, **Grameen Foundation**
- Tim Kotin, **SuperFluid Labs**
- Luke Shankland, **EFL**
- Leesa Shrader, **AgriFin**
- Jeff Stewart, **Lenddo**
- Michael Tsan, **Dalberg**

ABOUT THE INITIATIVE FOR SMALLHOLDER FINANCE

The Initiative for Smallholder Finance (ISF) is a multi-donor and investor platform for the development of financial services for the smallholder farmer market. It was launched in May 2013 with the intention of making marked progress toward closing the gap between the $450 billion in smallholder financing demand and the current $10-20 billion supply. The ISF's primary role is to act as a "design catalyst." The emphasis is on mobilizing additional financing for smallholders and seeding replication of innovative models in new markets.

The ISF is housed in the **Global Development Incubator** (GDI), a non-profit, public charity designed to support innovative organizations and initiatives that have the opportunity to create large-scale social change.

The authors would like to acknowledge and thank the sponsors of the ISF – the Citi Foundation, Ford Foundation, Gates Foundation, KFW Development Bank, MasterCard Foundation, Skoll Foundation, Small Foundation, and USAID – for their support and financing. Additionally, we would like to thank the Aspen Network of Development Entrepreneurs, Business Action for Africa, Business Fights Poverty, CGAP, IDH Sustainable Trade, One Acre Fund, Root Capital, and TechnoServe, for their advisory support.

AUTHORS:

Fauzia Jamal from the Initiative for Smallholder Finance conducted this research from October 2015 to January 2016. Matt Shakhovskoy, Jack Luft, and Sara Wallace provided advisory and editorial support.

www.initiativeforsmallholderfinance.org