American Economic Review 2012, 102(6): 2923–2954
http://dx.doi.org/10.1257/aer.102.6.2923

Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi†

By Xavier Giné, Jessica Goldberg, and Dean Yang*

We implemented a randomized field experiment in Malawi examining borrower responses to being fingerprinted when applying for loans. This intervention improved the lender's ability to implement dynamic repayment incentives, allowing it to withhold future loans from past defaulters while rewarding good borrowers with better loan terms. As predicted by a simple model, fingerprinting led to substantially higher repayment rates for borrowers with the highest ex ante default risk, but had no effect for the rest of the borrowers. We provide unique evidence that this improvement in repayment rates is accompanied by behaviors consistent with less adverse selection and lower moral hazard. (*JEL* D14, D82, G21, O12, O16)

Imperfections in credit markets are widely seen as key barriers to growth (King and Levine 1993). Among such imperfections, asymmetric information problems play a prominent role, as they limit the ability of borrowers to commit to carrying out their obligations under debt contracts. Borrowers cannot credibly reveal their borrower type (adverse selection), promise to exert sufficient effort on their enterprises (ex ante moral hazard), or promise to repay loans upon realization of enterprise profits, even when such profits are sufficient for repayment (ex post moral hazard).[1] Lenders seek to mitigate asymmetric information problems by imposing collateral requirements, engaging in costly screening of borrowers prior to approval, and, when a credit reporting system is available, sharing credit information with

* Giné: Development Economics Research Group, World Bank, and Bureau for Research and Economic Analysis of Development, 1818 H Street NW, Mail Stop MC 3–307, Washington, DC 20433 (e-mail: xgine@worldbank.org); Goldberg: Department of Economics, University of Maryland, 3115G Tydings Hall, College Park, MD 20742 (e-mail: goldberg@econ.umd.edu); Yang: Ford School of Public Policy and Department of Economics, University of Michigan and Bureau for Research and Economic Analysis of Development, 3316 Weill Hall, 735 S. State Street, Ann Arbor, MI 48109, and National Bureau of Economic Research (e-mail: deanyang@umich.edu). This paper was previously circulated under the title "Identification Strategy: A Field Experiment on Dynamic Incentives in Rural Credit Markets." Santhosh Srinivasan deserves the highest accolades for top-notch field work management and data collection. Lutamyo Mwamlima and Ehren Foss were key contributors to the success of the fieldwork, and Niall Keleher helped organize wrap-up data entry. We appreciate the vital support and assistance of Michael Carter, Charles Chikopa, Sander Donker, Lena Heron, Weston Kusani, David Rohrbach, Kondwani Shaba, Mark Visocky, and Eliza Waters. We received excellent comments from Abhijit Banerjee, Martina Björkman, Shawn Cole, Alan de Brauw, Quy-Toan Do, Greg Fischer, Raj Iyer, Dean Karlan, Craig McIntosh, Dilip Mookherjee, Jonathan Morduch, John Papp, Jean Philippe Platteau, Mark Rosenzweig, and participants at presentations at Bocconi University, University of Maryland, University of Michigan, University of Malawi, University of Namur, Oxford University, Syracuse University, UCLA, UC San Diego's 2009 microfinance conference, SITE 2009 at Stanford, NEUDC 2009 at Tufts, the third Impact Evaluation Network conference (2009) in Bogota, and BREAD 2010 in Montreal. This project was funded by the World Bank Research Committee, USAID's BASIS AMA CRSP, and USAID Malawi.
† To view additional materials, visit the article page at http://dx.doi.org/10.1257/aer.102.6.2923.

[1] For reviews of this literature, see Ghosh, Mookherjee, and Ray (2001) and Conning and Udry (2007).

other lenders. Microcredit institutions have addressed informational problems by relying on nontraditional mechanisms such as group liability. Microlenders, however, have recently come under attack, especially in India, because of allegations of overindebtedness of clients driven in part by rapid growth and increased competition. As a result, microlenders are seeking to participate in credit bureaus, much like traditional lenders.[2]

For a credit bureau to function effectively, however, it must be possible to uniquely identify individuals with reasonable certainty. Identification is necessary in order to retrieve a current loan applicant's past credit history from a credit database. Most developed countries have a unique identification system in the form of a social security number or government-issued photo identification. But in many of the world's poorer countries, large segments of the population lack formal identification documents, and even for those who have them, there is often no national system for uniquely identifying individuals in a database. In these countries, lenders accept different forms of identification, such as a passport, a health insurance policy number, or even a letter from the local village leader. Because documents can be falsified, and because individuals may simply use different types of identification when dealing with different lenders, it is extremely difficult to track a customer across multiple lenders, and it can even be difficult for lenders to identify defaulters within their own client base. Loan defaulters may avoid sanction for past default by simply applying for new loans under different identities.

Lenders respond by limiting the supply of credit, due to the inability to sanction unreliable borrowers and, conversely, to reward reliable borrowers with expanded credit. In rural areas, the result is that smallholder farmers are severely constrained in their ability to finance crucial inputs such as fertilizer and improved seeds, which limits production of both subsistence and cash crops.

Motivated by the benefits of a unique identification system, a number of efforts in the developing world are underway, many on a massive scale. For example, the Indian government has embarked on a vast effort to fingerprint and assign personal identification numbers that will replace all other forms of identification and enable citizens to access credit markets, public services, and subsidies on food, energy, and education that now suffer from major pilferage (Government of India 2005; *The Economist* 2011; Polgreen 2011).

Despite its importance, there is essentially no empirical evidence thus far on the impacts of improved personal identification in credit markets. A number of questions are of general interest. First, how do improvements in personal identification affect borrower and lender behavior and ultimately loan repayment rates? Second, how prevalent are adverse selection and moral hazard in the credit market? And finally, how does improved personal identification affect the operation of credit bureaus?

We report the results from a randomized field experiment that sheds light on the above questions. The experiment randomizes fingerprinting of loan applicants to test the impact of improved personal identification. The experiment was carried out in a context—rural Malawi—characterized by an imperfect identification system and limited access to credit.

[2] One of the recommendations of the Malegam Committee, set up in October 2010 after the crisis in Andhra Pradesh, India, was the establishment of a credit bureau and the adoption of a customer protection code.

According to the 2006 Doing Business Report, Malawi ranked 109 out of 129 countries in terms of private credit to GDP, a frequently used measure of financial development. Malawi also gets the lowest marks in the "depth of credit information index," which proxies for the amount and quality of information about borrowers available to lenders. Few rural Malawian households have access to loans for production purposes: only 11.7 percent report any production loans in the past 12 months, and among these loans only 40.3 percent are from formal lenders.[3]

In the experiment, farmers who applied for agricultural input loans to grow paprika were randomly assigned to either (i) a control group, or (ii) a treatment group where each member had a fingerprint collected as part of the loan application. A key advantage of fingerprints as a form of personal identification is that they are unique to and embodied in each person, so they cannot be forgotten, lost, or stolen. Improved borrower identification allows lenders to construct accurate credit histories and condition future lending on past repayment performance. Loan repayment could improve with fingerprinting, by making the lender's threats of future credit denial as well as promises of larger future loans more credible.

To frame the empirics, we develop a simple two-period model in the spirit of Stiglitz and Weiss (1983) that incorporates both adverse selection and moral hazard and show that dynamic incentives (that is, the ability to deny credit in the second period based on the first-period repayment performance) can reduce both types of asymmetric information problems and therefore raise repayment. Adverse selection problems can be mitigated because riskier individuals that would otherwise default may now take out smaller loans (or avoid borrowing altogether) to ensure access to credit in the future.[4] In addition, borrowers may have greater incentives to ensure that agricultural production is successful, either by exerting more effort or by diverting fewer resources away from production (lower moral hazard). Also, intuitively, the model predicts that the impact of dynamic incentives on borrowing, farmer actions during the production phase, and repayment will be largest for the riskiest individuals.

We find that fingerprinting led to substantially higher repayment rates for the subgroup of borrowers with the highest ex ante default risk.[5] In the context of the model, this result suggests that fingerprinting, by improving personal identification, enhanced the credibility of the lender's dynamic incentive. The impact of fingerprinting on repayment in the highest default risk subgroup (representing 20 percent of borrowers) is large: the average share of the loan repaid (2 months after the due date) was 66.7 percent in the control group, compared to 92.2 percent among fingerprinted borrowers.[6] In other words, for these farmers fingerprinting

[3] Figures are nationally representative and come from the 2004 Malawi Integrated Household Survey. Formal lenders include commercial banks, nongovernmental organizations, and microfinance institutions; informal lenders include moneylenders, family, and friends.

[4] In this paper we use the term "adverse selection" to mean ex ante selection effects deriving from borrowers' hidden information. We acknowledge that such selection may occur on the basis of either unobserved risk type (emphasized in the model) or unobserved anticipated effort (as highlighted by Karlan and Zinman 2009).

[5] To create the ex ante default risk measure, we regress loan repayment rates on borrowers' baseline characteristics in the control group, and then predict loan repayment for the entire sample (including the treatment group). This essentially creates a "credit score" for each borrower based on their ex ante (preborrowing) characteristics. We provide further details on this procedure in Section IV.

[6] The treatment effect implied by these figures is not regression-adjusted, but regression-based estimates are (as would be expected) very similar.

accounts for roughly three-quarters of the gap between repayment in the control group and full repayment. By contrast, fingerprinting had no impact on repayment for farmers with low ex ante default risk.

While we cannot separate the effects of moral hazard and adverse selection on repayment, we collect unique additional evidence that points to the presence of both informational problems. Fingerprinting leads farmers to choose smaller loan sizes. In the context of the theoretical model, this is consistent with a reduction in adverse selection. In addition, high-default-risk farmers who are fingerprinted also divert fewer inputs away from the contracted crop (paprika), which in the model represents a reduction in moral hazard. When we compare these benefits to estimated costs of implementation, we find that adoption of fingerprinting is cost effective, with a benefit-cost ratio of 2.34.[7]

The key contribution of the paper, in our view, is that it provides the first empirical evidence of the importance of personal identification for credit market efficiency. Imperfect personal identification is an information problem that has received little attention in the literature. Prior to this study, the extent to which identification of borrowers is a problem for formal lenders—in any sample population—was unknown. Our results indicate that alleviating this specific information asymmetry in rural Malawi has nonnegligible benefits for credit markets.

Our analysis is further distinguished by the nature of our outcome data. In addition to using the lender's administrative data to measure impacts on borrowing decisions and repayment, we also use a detailed follow-up survey to estimate impacts on several typically unobserved behaviors related to moral hazard. For example, we provide direct evidence of changes in production decisions and the use of borrowed funds stemming from improved identification.

This paper also has implications for the perceived benefits of a credit reporting system. Despite the absence of a credit bureau in Malawi, study participants were told that their fingerprints and associated credit histories could be shared with other lenders. Since fingerprinting led to positive changes in borrower behavior, the paper underscores the borrowers' belief that improved identification will allow the lender to condition credit decisions on past credit performance. This is important, because it suggests how borrowers may respond to the introduction of a credit bureau.

A related paper is Karlan and Zinman (2009) (hereafter KZ), who find experimental evidence of moral hazard and weaker evidence of adverse selection in urban South Africa. KZ introduce a dynamic incentive by making future interest rates conditional on current loan repayment. Our experiment differs from KZ's in several key ways. First, our experiment manipulates the *credibility* of dynamic incentives, while KZ's experiment *informs* borrowers of the *existence* of a dynamic incentive. Second, our follow-up survey provides insight into the specific behaviors that the intervention affects and that result in higher repayment. KZ, by contrast, relies only on the lender's administrative data and so cannot shed light on what borrower behaviors may have changed. Third, the timing of our intervention relative to the borrowing

[7] As we emphasize below, we have used quite conservative implementation cost estimates, often based on our own field implementation costs. The benefit-cost ratio could be even more attractive in a full-scale implementation that spreads fixed costs over a larger volume of borrowers, particularly in the context of a credit bureau with many participating lenders.

decision differs. In KZ, the dynamic incentive is announced *after* clients have agreed to borrow (and all loan terms have been finalized), so differences in repayment can only be due to moral hazard. In our case, the intervention that improves dynamic incentives is revealed *before* agents decide to borrow. This makes it possible to examine changes in the composition of borrowers and in loan size. In addition, we estimate the more relevant policy parameter because potential borrowers cannot be repeatedly surprised.

We informed the lender which clubs had been fingerprinted, so the lender could have changed its behavior toward treated and control clubs. For example, loan officers could have spent more time monitoring and enforcing repayment from control clubs, since treatment clubs were already subject to dynamic incentives. We provide evidence to the contrary: approval decisions and monitoring of clubs by loan officers did not differ across treated and control clubs. We therefore interpret our findings as emerging solely from borrowers' responses.

By documenting impacts on behaviors related to adverse selection and moral hazard, our findings contribute to a burgeoning empirical literature that tests claims made by contract theory and measures the prevalence of asymmetric information (see Chiappori and Salanié 2003 for a review). A number of recent papers provide empirical evidence of the existence and impacts of asymmetric information in credit markets, in both developed and developing countries. Ausubel (1999) uses a large-scale randomized trial of direct-mail preapproved solicitations from a major US credit card company and finds evidence of higher risk individuals selecting less favorable credit cards, consistent with adverse selection. Klonner and Rai (2009) exploit the introduction of a cap in bidding revolving savings and credit associations (ROSCAs) of South India and find higher repayment rates in earlier rounds attributable to changes in the composition of bidders, consistent with lower adverse selection. Visaria (2009) documents the positive impact of expedited legal proceedings on loan repayment among large Indian firms, even among loans that originated before the reform, consistent with a reduction in moral hazard. Giné and Klonner (2005) find that incomplete information about fishermen's ability in coastal India limits their access to credit for technology adoption. Edelberg (2004) also develops a model of adverse selection and moral hazard and finds evidence consistent with both informational problems in the United States.[8]

The paper is also related to a framed experiment conducted by Giné et al. (2010) in Peru that shows that dynamic incentives can be important. In addition, there is a theoretical and empirical literature on the impact of credit bureaus that are also related to this paper. The exchange of information about borrowers should theoretically reduce adverse selection (Pagano and Jappelli 1993) and moral hazard (Padilla and Pagano 2000). Empirically, de Janvry, McIntosh, and Sadoulet (2010) study the introduction of a credit bureau in Guatemala and find that it did contribute to efficiency in the credit market. The paper is also related to the literature on the recent rise in personal bankruptcies in the United States (Livshits, MacGee, and Tertilt 2010).

[8] Ligon (1998) implements empirical tests of the extent to which consumption allocations can be best described by permanent income, full information, or private information models, and finds that the private information model is most consistent with the data in two out of three International Crops Research Institute for the Semi-Arid Tropics villages in India. Paulson, Townsend, and Karaivanov (2006) estimate structurally competing models of credit markets in Thailand and find moral hazard to be important.

The remainder of this paper is organized as follows. Section I describes the experimental design and survey data and Section II presents the intuition of a simple model of loan repayment. Section III describes the regression specifications, and Section IV presents the empirical results. Section V provides additional discussion and robustness checks. Section VI presents the benefit-cost analysis of introducing biometric technology, and Section VII concludes.

I. Experimental Design and Survey Data

The experiment was carried out as part of the Biometric and Financial Innovations in Rural Malawi (BFIRM) project, a cooperative effort among Cheetah Paprika Limited (CP), the Malawi Rural Finance Corporation (MRFC), the University of Michigan, and the World Bank. CP is a privately owned agribusiness company established in 1995 that offers extension services and high-quality inputs to smallholder farmers via an outgrower paprika scheme. MRFC is a government-owned microfinance institution and provided financing for the in-kind loan package for one-half to one acre of paprika. Loaned funds were not disbursed in cash, but rather took the form of a credit at an agricultural input supplier for the financed production inputs. For further details on CP, MRFC, and the loan particulars, please see online Appendix A.

At the time of the study, the vast majority of farmers in the sample had no access to formal-sector credit. In our baseline survey, only 6.7 percent of farmers had any formal loans in the previous year. Among these few farmers with formal-sector credit, MRFC was the largest single lender, providing 34 percent of loans (more than twice the share of the next largest lender).[9] Farmers therefore had a strong interest in maintaining good credit history with MRFC so as to maintain access to what would likely be their primary source of formal credit in the future.

In the absence of fingerprinting, farmer identification relies on the personal knowledge of loan officers. Loan officers do build up knowledge of borrowers over time, which allows MRFC to implement some dynamic incentives: it does attempt to withhold loans from past defaulters, and to reward reliable borrowers with increased loan amounts at lower interest rates. The identification "technology" based on personal loan officer knowledge, however, is regarded as imperfect by top management at MRFC, who view the existing dynamic incentives as weak.[10] Loan officers are sometimes promoted and rotated to other localities. Among the 11 loan officers who were responsible for our study participants, the median number of years at the branch is only 2, while the median number of years working for the lender is 13.[11] In the absence of an independent mechanism for identifying borrowers, the institutional memory is lost when the loan officer is transferred to another location. Even when loan officers remain in a given location over time, the large number of borrowers

[9] Across study areas, access to formal credit varies from 4 percent to 10 percent. In Dedza, the region with highest access to formal loans, MRFC provides almost half of these formal loans.

[10] While we do not have systematic evidence on past defaulters taking out new loans under false identities, an accumulation of anecdotes had convinced top management at MRFC and other institutions that this was a major obstacle in their effort to expand access to credit.

[11] Because soft information about borrowers is important, one may be surprised by the high loan officer turnover rate. MRFC, like other lenders, rotates credit officers for many reasons. For example, rotation is thought to improve morale and help minimize corruption. Promotion of successful individuals within the organization also leads to replacement of loan officers at the local level and some loss of soft information on borrowers.

can lead them to make mistakes in identification. In this project, loan officers issued an average of 104 loans, and also handled other loan customers not associated with the project. Loan officers may also rely for identification on local informants, local leaders, and other borrowing group members, but such methods are also imperfect because of the possibility of collusion against the lender among fellow villagers.

The timeline of the experiment is presented in Appendix Figure 1. Our study sample consists of 214 clubs with 3,206 farmers in Dedza, Mchinji, Dowa, and Kasungu districts. Farmer clubs in the study were randomly assigned to be fingerprinted (the treatment group) or not (the control group), with an equal probability of being in either group. Randomization of treatment status was carried out after stratifying by locality and week of club visit.[12] Each loan officer is assigned to one locality. The stratification by locality and week of club visit thus ensured stratification by loan officer as well (i.e., each loan officer was responsible for roughly the same number of treatment and control clubs).

Club visits began with private administration of the baseline survey to individual farmers, and were followed by a training session. Both treatment and control groups were given a presentation on the importance of credit history in ensuring future access to credit. The training emphasized that defaulters would face exclusion from future borrowing, while borrowers in good standing could be rewarded with larger loans in the future. Then, in treatment clubs only, individual participants' fingerprints were collected. Our project staff explained how their fingerprint uniquely identified them for credit reporting to all major Malawian rural lenders, and that future credit providers would be able to access the applicant's credit history simply by checking his or her fingerprint.[13] Online Appendix A provides the script used during the training. See online Appendix B for further technical details on the biometric technology used.

After fingerprints were collected, a demonstration program was used to show participants that the computer was now able to identify an individual with only a fingerprint. One farmer was chosen at random to have his right thumb rescanned, and the club was shown that the person's name and demographic information (entered earlier alongside the original fingerprint scan) was retrieved by the computer program. During these demonstration sessions all farmers whose fingerprints were rescanned were correctly identified. The control group was not fingerprinted, but as mentioned previously, also received the same training emphasizing the importance of one's credit history and how it influences one's future credit access.[14]

The baseline survey administered prior to the training and the collection of fingerprints included questions on individual demographics (education, household size, religion), income-generating activities, and assets, including detailed information on crop production and crop choice, livestock and other assets, risk preferences, past and current borrowing activities, and past variability of income. Summary statistics

[12] In other words, each unique combination of locality and week of initial club visit constituted a stratification cell, within which clubs were evenly divided randomly between treatment and control (or as close as possible to evenly divided, when there was an odd number of clubs in the stratification cell). There are 11 localities in the study, each of which was covered by one loan officer. The full sample of 214 clubs (3,206 farmers) was spread across 31 stratification (location-week) cells.

[13] Our team of enumerators encountered essentially no opposition to fingerprint collection.

[14] Because we provided education on the importance of credit history to our control group as well, we can estimate neither the impact of fingerprinting without such education, nor the impact of the credit history education alone.

from the baseline survey are presented in Table 1, and variable definitions are provided in online Appendix C.[15]

After the completion of the survey, credit history training, and fingerprinting of the treatment group, the names and locations of the members that applied for loans along with their treatment status were handed over to MRFC loan officers so that they could screen and approve the clubs according to their protocols. Among other standard factors, MRFC conditions lending on the club's successful completion of 16 hours of training. MRFC approved loans for 2,063 out of 3,206 customers (in 121 out of 214 clubs). Of the customers approved for loans, some failed to raise the required down payment and others opted not to borrow for other reasons. The sample of borrowers consists of 1,147 loan customers from 85 clubs, in 21 stratification (location-week) cells.[16] Loan packages had an average value of MK16,913 (US$117).[17]

Within a group, take-up of the loan was an individual decision, but the subset of farmers who took up the loan was told that they were jointly liable for each others' loans. In practice, however, joint liability at this lender was not enforced.[18] MRFC applies sanctions primarily on individual defaulters and not on other (nondefaulting) members of a borrowing group. In other words, an individual who repaid a previous loan could obtain a new loan even if other borrowers in the same group had failed to repay a past loan, as long as defaulters from the group were removed before the group applied for new loans.

During the months of July and August, farmers harvested the paprika crop and sold it to CP at predefined collection points. CP then transferred the proceeds from the sale to MRFC, who then deducted the loan repayment and credited the remaining postrepayment proceeds to an individual farmer's savings account. This garnishing of the proceeds for loan repayment essentially allows MRFC to "seize" the paprika crop when farmers sell to CP (and for most farmers it is the only sales outlet).[19] Farmers could also make loan repayments directly to MRFC at their branch locations or during credit officer visits to their villages; this occurred, for example, among the small number of farmers who sold to paprika buyers other than CP. This channel of repayment is less desirable to MRFC because it is riskier.

We also implemented a follow-up survey of farmers in August 2008, once crops had been sold and income received. The sample size of this follow-up survey is 1,226 in total (borrowers plus nonborrowers), among whom 520 were borrowers.[20]

[15] To ensure that survey answers were not influenced by knowledge of the experiment or the respondent's treatment status, survey data were collected prior to the credit history education and fingerprinting intervention.

[16] While a natural question at this point is whether selection into borrowing was affected by treatment status, treatment and control groups did not differ in their rates of MRFC loan approval or the fraction of farmers who ended up with a loan. Furthermore, treated and untreated borrowers do not differ systematically on the basis of baseline characteristics. These points will be discussed in detail in the results section below.

[17] All conversions of Malawi kwacha to US dollars in this paper assume an exchange rate of MK145/US$, the average exchange rate at the time of the experiment.

[18] See Giné and Yang (2009) for another example of limited enforcement of joint liability loans.

[19] Proceeds from other types of crops of course cannot be seized in this way to secure loan repayment because MRFC does not have analogous garnishing arrangements with other crop buyers.

[20] The 520 borrowers are spread across 17 stratification (location-week) cells. The follow-up sample is smaller than the sample of baseline borrowers because for budget reasons we could not visit each borrowing household at their place of residence. Instead, we invited study participants to come to a central location at a certain date and time to be administered the follow-up interview. Not all farmers attended the meeting where the follow-up survey was administered, but as we discuss below in Section VC (see online Appendix Table 3), there is no evidence of selective attrition related to treatment status. For the full sample as well as the borrower subsample, in no regression is fingerprinting or fingerprinting interacted with predicted repayment statistically significantly associated with attrition from the survey.

Table 1—Summary Statistics

	Mean	SD	10th percentile	Median	90th percentile	Observations
Basic characteristics						
Male	0.80	0.40	0	1	1	1,147
Married	0.94	0.24	1	1	1	1,147
Age	39.96	13.25	24	38	59	1,147
Years of education	5.35	3.50	0	5	10	1,147
Risk taker	0.56	0.50	0	1	1	1,147
Days of hunger last year	6.05	11.05	0	0	30	1,147
Late paying previous loan	0.13	0.33	0	0	1	1,147
Income SD	27,568.34	46,296.41	3,111.27	15,556.35	57,841.34	1,147
Years of experience growing paprika	2.22	2.36	0	2	5	1,147
Previous default	0.02	0.14	0	0	0	1,147
No previous loans	0.74	0.44	0	1	1	1,147
Predicted repayment	0.79	0.26	0.33	0.90	1.02	1,147
Take-up						
Approved	0.99	0.08	1	1	1	1,147
Any loan	1.00	0.00	1	1	1	1,147
Total borrowed (MK)	16,912.60	39,08.03	13,782	16,100	20,136.07	1,147
Land use						
Fraction of land used for maize	0.43	0.16	0.28	0.40	0.63	520
Fraction of land used for soya/beans	0.15	0.16	0.00	0.11	0.38	520
Fraction of land used for groundnuts	0.13	0.12	0.00	0.11	0.29	520
Fraction of land used for tobacco	0.08	0.12	0.00	0.00	0.27	520
Fraction of land used for paprika	0.19	0.13	0.00	0.18	0.36	520
Fraction of land used for tomatoes	0.01	0.03	0.00	0.00	0.00	520
Fraction of land used for leafy vegetables	0.00	0.02	0.00	0.00	0.00	520
Fraction of land used for cabbage	0.00	0.01	0.00	0.00	0.00	520
Fraction of land used for all cash crops	0.57	0.16	0.38	0.60	0.72	520
Inputs						
Seeds (MK, paprika)	247.06	348.47	0	0	560	520
Fertilizer (MK, paprika)	7,499.85	7,730.05	0	5,683	18,200	520
Chemicals (MK, paprika)	671.31	1,613.13	0	0	2,500	520
Man-days (MK, paprika)	665.98	1,732.99	0	0	2,400	520
All paid inputs (MK, paprika)	9,084.19	8,940.13	0	8,000	19,990	520
KG manure, paprika	90.84	313.71	0	0	250	520
Times weeding, paprika	1.94	1.18	0	2	3	520
Outputs						
KG maize	1,251.30	1,024.36	360	1,080	2,160	520
KG soya/beans	83.14	136.86	0	40	200	520
KG groundnuts	313.89	659.34	0	143	750	520
KG tobacco	165.47	615.33	0	0	400	520
KG paprika	188.14	396.82	0	100	364	520
KG tomatoes	30.56	126.29	0	0	0	520
KG leafy vegetables	29.94	133.24	0	0	0	520
KG cabbage	12.02	103.79	0	0	0	520
Revenue and profits						
Market sales (MK)	65,004.30	76,718.29	9,800	44,000	137,100	520
Profits (market sales + value of unsold crop − cost of inputs, MK)	117,779.20	303,100.80	33,359	95,135	261,145	520
Value of unsold harvest (Regional prices, MK)	80,296.97	288,102.70	24,645	70,300	180,060	520
Repayment						
Balance, September 30	2,912.91	6,405.77	0	0	13,981	1,147
Fraction paid by September 30	0.84	0.33	0	1	1	1,147
Fully paid by September 30	0.74	0.44	0	1	1	1,147
Balance, eventual	2,080.86	5,663.98	0	0	9,282	1,147
Fraction paid, eventual	0.89	0.29	0	1	1	1,147
Fully paid, eventual	0.79	0.41	0	1	1	1,147

The formal loan maturity (payment) date was September 30, 2008. Some additional payments were made after the formal due date; MRFC reports that there is typically no additional loan repayment two months past the due date for agricultural loans. In the empirical analysis we obtain our dependent variables from the August 2008 survey data as well as administrative data from MRFC on loan take-up, amount borrowed, and repayment.

Balance of Baseline Characteristics across Treatment versus Control Groups.— To confirm that the randomization across treatments achieved balance in terms of pretreatment characteristics, online Appendix Table 1 presents the means of several baseline variables for the control group as reported prior to treatment, alongside the difference vis-à-vis the treatment group (mean in treatment group minus mean in control group). We also report statistical significance levels of the difference in treatment-control means. These tests are presented for both the full baseline sample and the loan recipient sample.

Overall, we find balance between the two groups in both the full baseline sample and the loan recipient sample. In the full baseline sample, the difference in means for the treatment and control groups is not significant for any of the 11 baseline variables. In the loan recipient sample, for 10 out of these 11 baseline variables, the difference in means between treatment and control groups is not statistically significantly different from 0 at conventional levels, and so we cannot reject the hypothesis that the means are identical across treatment groups. For only one variable, the indicator for the study participant being male, is the difference statistically significant (at the 10 percent level): the fraction male in the treatment group is 6.6 percentage points lower than in the control group.[21]

II. A Simple Model of Borrower Behavior

Fingerprinting improves the personal identification of borrowers and thus increases the credibility of dynamic incentives used by the lender. To study how dynamic incentives affect borrower behavior, online Appendix D develops a simple model that incorporates both adverse selection and moral hazard. Here we provide an intuitive discussion of the model.

We assume that prospective borrowers have no liquid assets and decide how much to borrow for cash crop inputs, so the amount invested in production cannot exceed the loan amount. We introduce adverse selection by assuming that borrowers differ in the probability that production is successful, while moral hazard is modeled by allowing borrowers to divert the loan amount instead of investing it in production.[22] Consistent with the credit contract offered in the context of the experiment, we model a lender that offers a loan amount that can take on two values (depending on the number of fertilizer bags borrowed) and a gross interest rate. We also assume

[21] It turns out, however, that the regression results to come are not substantially affected by the inclusion in the regressions of the "male" indicator and other control variables (results not shown).

[22] Given the arrangement to buy the cash crop (paprika) in the experiment, we assume that the lender can only seize cash crop production but not the proceeds from diverted inputs. To be clear, the production of paprika does not reduce moral hazard because paprika faces less *production* risk than other crops, but rather because it is less risky for *the lender*, given the lender's ability to confiscate paprika output for repayment of the loan.

that when the smaller amount is borrowed, production can cover loan repayment even if it fails.

When personal identification of clients is not possible, borrowers can obtain a new loan even if they have defaulted in the past simply by using a different identity. As a result, lenders are forced to offer the same one-season contract every period, as they cannot tailor the terms of the contract to individual credit histories.

By contrast, when personal identification is possible, the lender can use dynamic incentives, conditioning future credit on past repayment performance. In this situation, borrowers face a trade-off between diverting inputs away from cash crop production but jeopardizing chances of a loan in the future versus ensuring repayment of the current loan and therefore securing a loan in the future. In addition, by choosing the smaller loan amount they obtain lower net income in the first period in return for securing a loan in the future.

With this setup, the model predicts that dynamic incentives will have different effects on the optimal choices of borrowers depending on their probability of success. In particular, borrowers with relatively low probability of success are most affected by the introduction of dynamic incentives. They choose the higher loan amount and divert it all without dynamic incentives, but borrow the lower amount and invest it in cash crop production when dynamic incentives are introduced. Borrowers with the highest probabilities of success are the least affected: even without dynamic incentives, they never divert inputs and always choose the higher loan amount. Finally, borrowers with intermediate values of the probability of success will, upon introduction of dynamic incentives, change either the diversion or the loan size decisions (depending on parameter values and functional forms).

The model provides a reasonable structure for framing the empirical results to come. Its key advantage is a close adherence to the context of the experiment, in which the main simplifying assumptions (e.g., binary loan size and the lender's inability to seize non–cash crop output) are reasonable. That said, our model may not be the only one that could be used to understand borrower behaviors in this and other contexts; other models may provide a different interpretation of the results. Therefore, our empirical results should be interpreted in the context of this specific model.

III. Regression Specification

Because the treatment is assigned randomly at the club level, its impact on the various outcomes of interest (say, repayment) can be estimated via the following regression equation:

$$Y_{ijs} = \alpha + \beta T_{js} + \gamma_s + \varepsilon_{ijs}, \tag{1}$$

where Y_{ijs} = repayment outcome for individual i in club j in stratification cell s (e.g., equal to 1 if repaying in full and on time, and 0 otherwise), T_{js} is the treatment indicator (1 if fingerprinted and 0 if not), and γ_s is a fixed effect for stratification cell s. ε_{ijs} is a mean-zero error term. Treatment assignment at the club level creates spatial and other correlation among farmers within the same club, so standard errors must be clustered at the club level (Moulton 1986). Inclusion of the stratification cell fixed effects can reduce standard errors by absorbing residual variation.[23] The coefficient β on the treatment indicator is the average treatment effect (ATE) of fingerprinting on the dependent variable.[24]

The point that β in equation (1) is an average treatment effect is important, because we also devote attention to treatment effect heterogeneity. In particular, we are interested in the interaction between the randomized treatment and a measure of the ex ante probability of repayment. Examining this dimension of heterogeneity is a test of the theoretical model's prediction that the impact of dynamic incentives on repayment is negatively related to the ex ante repayment rate (what the repayment rate would have been in the absence of dynamic incentives): borrowers who, without the dynamic incentive, would have had lower repayment will see their repayment rates rise more when the dynamic incentive is introduced.[25] To test this question, we estimate regression equations of the following form:

$$Y_{ijs} = \alpha + \rho(T_{js} \times D_{ijs}) + \beta T_{js} + \chi D_{ijs} + \gamma_s + \varepsilon_{ijs}; \tag{2}$$

D_{ijs} is a variable representing the individual's predicted likelihood of repayment. The coefficient ρ on the interaction term $T_{js} \times D_{ijs}$ reveals the extent to which the impact of the treatment on repayment varies according to the borrower's predicted repayment. The main effect of predicted repayment, D_{ijs}, is included in the regression as well.

To implement equation (2) examining heterogeneity in the effect of fingerprinting, we construct an index of predicted repayment. This involves creating what is essentially a "credit score" for each borrower in the sample on the basis of the relationship between baseline characteristics (some of which may not be observable to the lender) and repayment among borrowers in the *control* (nonfingerprinted) group. Limiting the sample to borrowers in the control group ($N = 563$), we run a regression of repayment (fraction of loan repaid by the September 30, 2008 due date) on various farmer- and club-level baseline characteristics. Conceptually, the resulting index will be purged of any bias introduced by the effects of fingerprinting on repayment because it is constructed using coefficients from a regression predicting repayment for only the control (nonfingerprinted) farmers.

Table 2 presents results from this exercise. Statistically significant results in column 1, which only includes farmer-level (individual) variables on the right-hand side, indicate that older farmers and those who do not self-identify as risk takers have better repayment performance on the loan. Inclusion of a complete set of fixed effects for (locality) $\times$ (week of initial club visit) interactions raises the R^2 substantially

[23] Recall that stratification cells are defined by unique combinations of locality and week of initial club visit. By definition there are as close as possible to equal numbers of treatment and control clubs in each cell.

[24] Because we had perfect compliance with fingerprinting in the treatment group (and no fingerprinting in the control group), this happens to be a rare situation where β is also the average treatment effect on the treated (ATT).

[25] While in the model the single dimension of borrower heterogeneity is the probability of success, p, we have no way to estimate this directly for our full borrowing sample. Note that the repayment rate is monotonic in p, making it a good proxy for p. While in principle one could apply the procedure in online Appendix E with crop output as the dependent variable, in practice this would limit us because crop output is only observed in the smaller subsample of borrowers ($N = 520$). The repayment rate, on the other hand, comes from administrative data and so is available for the entire borrowing sample.

(from 0.05 in column 1 to 0.46 in column 2). The explanatory power of the regression is marginally improved further in column 3 (to an R^2 of 0.48) when age and education are specified as categorical variables (instead of being entered linearly).

We then take the coefficient estimates from column 3 of the table and predict the fraction of loan repaid for the *entire* sample (both control and treatment observations). This variable, which we call "predicted repayment," is useful for analytical purposes because it is a single index that incorporates a wide array of baseline information (at the individual and locality level) correlated with repayment outcomes.[26]

To investigate heterogeneity in the treatment effect, this index is either interacted linearly with the treatment indicator (as in equation (2)), or it is converted into indicators for quintiles of the distribution of predicted repayment in the absence of fingerprinting and then interacted with treatment. For this analysis to be valid, it must be true that randomization leads to balance with respect to predicted repayment across treatment and control groups. This is indeed what we find.[27] In all regression results where the treatment indicator is interacted with predicted repayment, we report bootstrapped standard errors because the predicted repayment variable is a generated regressor.[28]

IV. Empirical Results: Impacts of Fingerprinting

This section presents our experimental evidence on the impacts of fingerprinting on a variety of interrelated outcomes. We examine impacts on loan approval and borrowing decisions, on repayment outcomes, and on intermediate farmer actions and outcomes that may ultimately affect repayment.

Tables 3 through 7 will present regression results from estimation of equations (1) and (2) in a similar format. In each table, each column will present regression results for a given dependent variable. Panel A will present the coefficient on treatment (fingerprint) status from estimation of equation (1).

Then, to examine heterogeneity in the effect of fingerprinting, panels B and C will present results from estimation of versions of equation (2) where fingerprinting is interacted linearly with predicted repayment (panel B) or with dummy variables for quintiles of predicted repayment (panel C). In both panels B and C, the respective main effects of the predicted repayment variables are also included in the regression (but for brevity the coefficients on the predicted repayment main effects will not be presented). In panel C, the main effect of fingerprinting is not included in the regression, to allow each of the five quintile indicators to be interacted with the indicator for fingerprinting in the regression. Therefore, in panel C the coefficient on each fingerprint-quintile interaction should be interpreted as the impact of

[26] In the loan-recipient subsample, predicted repayment has a mean of 0.79, with standard deviation 0.26. As expected, predicted repayment is highly skewed, with median predicted repayment of 0.90.

[27] In regressions of the treatment indicator on the continuous predicted repayment variable and indicators for stratification cells, the coefficient on predicted repayment is always far from statistical significance at conventional levels in all samples used in this paper. In regressions of the treatment indicator on indicators for each quintile of repayment, the coefficients on the quintile dummies are individually and jointly insignificantly different from zero in all subsamples.

[28] For coefficients in regressions in the form of equation (2), we calculate standard errors from 200 bootstrap replications. In each replication, we resample borrowing clubs from our original data (which preserves the original club-level clustering), compute predicted repayment based on the new sample, and rerun the regression in question using the new value of predicted repayment for that replication. See Efron and Tibshirani (1993) for details.

Table 2—Auxiliary Regression Predicting Loan Repayment

Dependent variable	Fraction paid by September 30 (1)	Fraction paid by September 30 (2)	Fraction paid by September 30 (3)
Male	0.080	0.061	0.058
	(0.073)	(0.048)	(0.048)
Married	−0.071	−0.091	−0.101
	(0.060)	(0.044)**	(0.046)**
Age	0.004	0.001	
	(0.001)***	(0.001)	
Years of education	−0.005	−0.003	
	(0.005)	(0.004)	
Risk taker	−0.078	0.008	0.013
	(0.041)*	(0.031)	(0.031)
Days of hunger in previous season	0.001	−0.000	−0.001
	(0.002)	(0.001)	(0.001)
Late paying previous loan	−0.058	−0.084	−0.084
	(0.071)	(0.046)*	(0.047)*
Standard deviation of past income	−0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Years of experience growing paprika	0.005	0.007	0.007
	(0.013)	(0.011)	(0.011)
Previous default	0.088	0.128	0.097
	(0.163)	(0.079)	(0.078)
No previous loan	−0.012	0.015	0.013
	(0.062)	(0.032)	(0.034)
Constant	0.729	0.949	0.982
	(0.114)***	(0.072)***	(0.090)***
Locality × week of initial club visit fixed effects	—	Yes	Yes
Dummy variables for five-year age groups	—	—	Yes
Dummy variables for each year of education	—	—	Yes
Observations	563	563	563
R^2	0.05	0.46	0.48

Notes: Sample is nonfingerprinted loan recipients from the September 2008 baseline survey. All standard errors are clustered at the club level. Robust standard errors in parentheses.
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

fingerprinting on borrowers in that quintile, compared to control group borrowers in that same quintile.

Finally, in Tables 3 through 7 the means of the dependent variable in a given column, for the overall sample as well for each quintile of predicted repayment separately, are reported at the bottom of each table.

A. *Loan Approval*, *Take-up*, *and Amount Borrowed*

The first key question to ask is whether fingerprinted farmers were more likely to have their loans approved by the lender, or were more likely to take out loans, compared to the control group. This question is important because the degree of selectivity in the borrower pool induced by fingerprinting affects interpretation of any effects on repayment and other outcomes.

Although loan officers were told which clubs had been fingerprinted in September 2007 when loan applications were due, they do not appear to have retained or used this information. Since biometric technology can be seen as a substitute for loan officer effort, one would expect loan officers to have better knowledge about nonfingerprinted clubs. This is not what we find, however. Loan officers' knowledge about clubs (identity of club officers, number of loans) is not related to treatment status, and in fact loan officers do not appear to know the treatment status of clubs. Borrower reports of contact with loan officers are also uncorrelated with treatment. (For further details on this analysis, see online Appendix E.) Given that loan officers do not appear to have responded to the treatment, we interpret impacts of the treatment as emerging solely from borrowers' responses to being fingerprinted.

Because loan officers did not take treatment status into account, it is not surprising that fingerprinting had no effect on loan approval. We also find no effect on loan take-up by borrowers, perhaps because clubs were formed with the expectation of credit availability and fingerprinting did not act as a strong enough deterrent to borrowing to affect farmers' decisions at the extensive margin. Columns 1 and 2 of Table 3 present results from estimation of equations (1) and (2) for the full baseline sample where the dependent variables are, respectively, an indicator for the lender's approving the loan for the given farmer (mean 0.63), and an indicator for the farmer ultimately taking out the loan (mean 0.35).[29]

There is no evidence that the rate of loan approval or take-up differs substantially across the treatment and control groups on average: the coefficient on fingerprinting is not statistically different from zero in either columns 1 or 2, panel A.

There is also no indication of selectivity in the resulting borrowing pool across subgroups of borrowers with different levels of predicted repayment. The coefficient on the interaction of fingerprinting with predicted repayment is not statistically significantly different from zero in either columns 1 or 2 of panel B. When looking at interactions with quintiles of predicted repayment (panel C), while the fingerprint quintile 2 interaction is positive and significantly different from zero at the 10 percent level in the loan approval regression, none of the interaction terms with fingerprinting are significantly different from zero in the loan take-up regression.

It does appear that, conditional on borrowing, fingerprinted borrowers took out smaller loans. In column 3 of Table 3, the dependent variable is the total amount borrowed in Malawi kwacha (MK). Panel A indicates that loans of fingerprinted borrowers were MK693 smaller than loans in the control group on average, a difference that is significant at the 10 percent level.

The patterns of coefficients in panel C are suggestive that this effect is confined to borrowers in the lowest quintile of expected repayment. Differences between fingerprinted and nonfingerprinted borrowers are small and not significant in quintiles two and above, but in quintile one, where fingerprinted borrowers take out loans that are smaller by MK2,657 (roughly US$18) than those in the corresponding quintile in the control group, the difference is marginally significant (the t-statistic is 1.55).

[29] Not all farmers who were approved for the loan ended up taking out the loan. Anecdotal evidence indicates that a substantial fraction of nontake-up among approved borrowers resulted when borrowers failed to raise the required deposit (amounting to 15 percent of the loan amount).

Table 3—Impact of Fingerprinting on Borrowing

	All respondents		Loan recipients
	Approved	Any loan	Total borrowed (MK)
Dependent variable	(1)	(2)	(3)
Panel A			
Fingerprint	0.045	0.056	−692.743*
	(0.054)	(0.045)	(381.745)
Panel B			
Fingerprint	0.215	0.118	−2,872.348
	(0.161)	(0.146)	(2,438.851)
Predicted repayment × fingerprint	−0.220	−0.081	2,693.752
	(0.196)	(0.169)	(2,630.912)
Panel C			
Fingerprint × quintile 1	0.099	0.081	−2,657.315
	(0.116)	(0.113)	(1,716.684)
Fingerprint × quintile 2	0.191**	0.113	−357.168
	(0.096)	(0.087)	(856.156)
Fingerprint × quintile 3	−0.022	0.057	−585.469
	(0.083)	(0.073)	(562.841)
Fingerprint × quintile 4	0.004	−0.032	−198.714
	(0.088)	(0.083)	(569.119)
Fingerprint × quintile 5	−0.009	0.044	−234.098
	(0.088)	(0.089)	(765.383)
Observations	3,206	3,206	1,147
Mean of dependent variable	0.63	0.35	16,912.60
Quintile 1	0.58	0.29	17,992.53
Quintile 2	0.64	0.36	17,870.61
Quintile 3	0.71	0.44	16,035.10
Quintile 4	0.70	0.47	15,805.54
Quintile 5	0.59	0.30	16,886.56
Joint significance of panel B coefficients (p-value)	0.48	0.59	0.88

Notes: Each column presents estimates from three separate regressions: main effect of fingerprinting in panel A, linear interaction with predicted repayment in panel B, and interactions with quintiles of predicted repayment in panel C. All regressions include stratification cell (location × week of initial club visit) fixed effects. Panel B regressions include the main effect of the level of predicted repayment, and panel C regressions include dummies for quintile of predicted repayment main effects. Standard errors on panel A coefficients are clustered at the club level, while those in panels B and C are bootstrapped with 200 replications and club-level resampling. The p-values in the bottom row are from tests that in panel B, the fingerprinting main effect and the "Predicted repayment × fingerprint" interaction term are jointly statistically significantly different from zero.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

*Significant at the 10 percent level.

While the absence of statistical significance in panel C makes this just a suggestive result, the pattern is in accordance with the theoretical model's prediction that the "worst" borrowers (those whose repayment rates would be lowest in the absence of dynamic incentives) will respond to the imposition of a dynamic incentive by voluntarily reducing their loan sizes.

These results, while only suggestive, are consistent with fingerprinting reducing adverse selection in the credit market, albeit on a different margin than is usually discussed in the credit context. Existing research tends to emphasize that improved enforcement should lead low-quality borrowers to be excluded from borrowing entirely—i.e., the improvement of the borrower pool operates on the *extensive* margin. Our results are suggestive that low-quality borrowers choose smaller loan sizes, which leads the overall loan pool to be less weighted toward low-quality borrowers. The improvement in the borrowing pool operates on the *intensive* margin of borrowing, rather than the extensive margin.

Interpretation of subsequent differences in the repayment rates (discussed below) should keep this result in mind. Improvements in repayment among fingerprinted borrowers (particularly among those in the lowest quintile) may in part result from their decisions to take out smaller loans at the very outset of the lending process and improve their eventual likelihood of repayment.

B. *Loan Repayment*

How did fingerprinting affect ultimate loan repayment? Table 4 presents estimated effects of fingerprinting for the loan recipient sample on three outcomes: outstanding balance (in MK), fraction of loan paid, and an indicator for whether the loan is fully paid, all by September 30, 2008 (the official due date of the loan, after which the loan is officially past due). The next three columns (columns 2–4) are similar, but the three variables refer to "eventual" repayment as of the end of November 2008. The lender makes no attempt to collect past-due loans after November of each agricultural loan cycle, so the eventual repayment variables represent the final repayment status on these loans.

Results for all loan repayment outcomes are similar: fingerprinting improves loan repayment, in particular for borrowers expected ex ante to have poorer repayment performance. Coefficients in panel A indicate that fingerprinted borrowers have lower outstanding balances, higher fractions paid, and are more likely to be fully paid on time as well as eventually (and the coefficients in the regressions for outstanding balance and fraction paid on time are statistically significant at the 10 percent level).

In panel B, the fingerprinting-predicted repayment interaction term is statistically significantly different from zero (at the 5 percent or 1 percent levels) in all regressions. The effect of fingerprinting on repayment is larger the lower is the borrower's ex ante likelihood of repayment. The fingerprint main effect and the "Predicted repayment × fingerprint" interaction term are jointly significantly different from zero at conventional levels (p-values reported in the bottom row of the table).

In panel C, it is evident that the effect of fingerprinting is isolated in the lowest quintile of expected repayment, with coefficients on the fingerprint-quintile 1 interaction all being statistically significantly different from zero at the 5 percent or 1 percent level and indicating beneficial effects of fingerprinting on repayment (lower outstanding balances, higher fraction paid, and higher likelihood of full repayment). Coefficients on other fingerprint-quintile interactions are all smaller in magnitude and not statistically significantly different from zero (with the exception of the positive coefficient on the fingerprint-quintile 5 interaction for outstanding balance and the negative-coefficient corresponding-interaction term for fraction paid, which is odd and may simply be due to sampling variation).

The magnitudes of the repayment effect found for the lowest predicted-repayment quintile are large. The MK7,249.27 effect on eventual outstanding balance amounts to 40 percent of the average loan size for borrowers in the lowest predicted-repayment quintile. While outstanding balance should mechanically be lower due to

Table 4—Impact of Fingerprinting on Repayment

	Loan recipients					
Dependent variable	Balance, Sept. 30 (1)	Fraction paid by Sept. 30 (2)	Fully paid by Sept. 30 (3)	Balance, eventual (4)	Fraction paid, eventual (5)	Fully paid, eventual (6)
Panel A						
Fingerprint	−1,489.945*	0.069*	0.088	−975.181	0.044	0.080
	(836.931)	(0.041)	(0.066)	(762.090)	(0.037)	(0.061)
Panel B						
Fingerprint	−15,173.560***	0.719***	0.847***	−9,800.693**	0.447**	0.614***
	(2,712.601)	(0.108)	(0.180)	(4,150.1)	(0.183)	(0.225)
Predicted repayment × fingerprint	16,987.019***	−0.807***	−0.942***	10,958.377***	−0.500**	−0.663**
	(3,010.827)	(0.120)	(0.199)	(4,452.911)	(0.196)	(0.245)
Panel C						
Fingerprint × quintile 1	−10,844.701***	0.506***	0.549***	−7,249.271**	0.327**	0.408***
	(2,622.283)	(0.125)	(0.144)	(2,918.825)	(0.135)	(0.156)
Fingerprint × quintile 2	−1,007.857	0.056	0.154	−1,006.419	0.057	0.165
	(2,033.207)	(0.105)	(0.165)	(1,870.044)	(0.098)	(0.152)
Fingerprint × quintile 3	−275.604	−0.001	−0.007	−261.204	−0.003	0.002
	(950.669)	(0.048)	(.092)	(878.856)	(0.044)	(0.087)
Fingerprint × quintile 4	812.241	−0.040	−0.064	701.779	−0.032	−0.041
	(915.645)	(0.044)	(0.078)	(863.69)	(0.042)	(0.075)
Fingerprint × quintile 5	1,702.297*	−0.075*	−0.085	1,429.524	−0.060	−0.051
	(968.333)	(0.044)	(0.074)	(906.418)	(0.041)	(0.071)
Observations	1,147	1,147	1,147	1,147	1,147	1,147
Mean of dependent variable	2,912.91	0.84	0.74	2,080.86	0.89	0.79
Quintile 1	6,955.67	0.62	0.52	4,087.04	0.81	0.68
Quintile 2	4,024.05	0.77	0.63	3,331.17	0.81	0.67
Quintile 3	1,571.44	0.92	0.83	1301.79	0.93	0.84
Quintile 4	877.80	0.95	0.85	781.59	0.95	0.87
Quintile 5	1,214.19	0.94	0.85	950.29	0.95	0.88
Joint significance of panel B coefficients (*p*-value)	0.00	0.00	0.00	0.02	0.01	0.00

Notes: Each column presents estimates from three separate regressions: main effect of fingerprinting in panel A, linear interaction with predicted repayment in panel B, and interactions with quintiles of predicted repayment in panel C. All regressions include stratification cell (location × week of initial club visit) fixed effects. Panel B regressions include the main effect of the level of predicted repayment, and panel C regressions include dummies for quintile of predicted repayment main effects. Standard errors on panel A coefficients are clustered at the club level, while those in panels B and C are bootstrapped with 200 replications and club-level resampling. The *p*-values in the bottom row are from tests that in panel B, the fingerprinting main effect and the "Predicted repayment × fingerprint" interaction term are jointly statistically significantly different from zero.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

the lower loan size in the lowest predicted-repayment quintile, the effect is almost three times the size of the reduction in loan size, so by itself lower loan size cannot explain the treatment effect on repayment. The 32.7 percentage point increase in eventual fraction paid and the 40.8 percentage point increase in the likelihood of eventually being fully paid are also large relative to bottom quintile percentages of 81 percent and 68 percent, respectively. Put another way, the fingerprinting-induced increase in repayment for the lowest quintile accounts for nearly the entire gap between repayment absent fingerprinting and full repayment.

C. *Intermediate Outcomes That May Effect Repayment*

In this section we examine decisions that farmers make throughout the planting and harvest season that may contribute to higher repayment among fingerprinted farmers. The dependent variables in the remaining results tables are available from a smaller subset of loan recipients ($N = 520$) who were successfully interviewed in the August 2008 follow-up survey round. To help rule out the possibility that selection into the 520-observation August 2008 follow-up survey sample might bias the regression results for that sample, column 2 of online Appendix Table 3 examines selection of loan recipients into the follow-up survey sample. The regressions are analogous in structure to those in the main results tables (panels A, B, and C), and the dependent variable is a dummy variable for attrition from the baseline (September 2007) to the August 2008 survey. There is no evidence of selective attrition related to treatment status: in no case is fingerprinting or fingerprinting interacted with predicted repayment statistically significantly associated with attrition from the survey.

Online Appendix Table 4 presents regression results for repayment outcomes that are analogous to those in columns 4–9 of main Table 3, but where the sample is restricted to this 520-observation sample. The results confirm that the repayment results in the 520-observation sample are very similar to those in the overall loan recipient sample, in terms of both magnitudes of effects and statistical significance levels.

Land Area Allocated to Various Crops.—One of the first decisions that farmers make in any planting season (which typically starts in November and December) is the proportion of land allocated to different crops. Tables 5A and 5B examine the average and heterogeneous impact of fingerprinting on land allocation; the dependent variables across columns are fraction of land used in maize (Table 5A, column 1), seven cash crops (Table 5A, columns 2–5, and Table 5B, columns 1–3), and all cash crops combined (Table 5B, column 4).[30]

Why might land allocation to different crops respond to fingerprinting? As discussed in the context of the theoretical model (footnote 22), nonproduction of paprika is a form of moral hazard, since the lender can only feasibly seize paprika output (in collaboration with the paprika buyer) and not other crops. By not producing paprika (or producing less), the borrower is better able to avoid repayment. Therefore, by improving the lender's dynamic incentives, fingerprinting may discourage such diversion of inputs and land to other crops.

While none of the effects of fingerprinting in Tables 5A or 5B (either overall in panel A or in interaction with predicted repayment in panels B and C) are statistically significant at conventional levels, the point estimates provide suggestive evidence that there is an impact of fingerprinting on land allocation for borrowers in the first predicted-repayment quintile. In this group, the effect of fingerprinting on land allocated to paprika (Table 5A, column 5, first row of panel C) is marginally significant (with a t-statistic of 1.57) and positive, indicating that fingerprinting leads farmers

[30] For each farmer, the values of the variables across columns 1–8 add up to 1.

Table 5A—Impact of Fingerprinting on Land Use

	Dependent variable: Fraction of land used for:				
	Maize (1)	Soya/Beans (2)	Groundnuts (3)	Tobacco (4)	Paprika (5)
Panel A					
Fingerprint	−0.003	0.015	−0.011	−0.007	0.010
	(0.020)	(0.019)	(0.016)	(0.016)	(0.014)
Panel B					
Fingerprint	−0.047	−0.005	−0.003	−0.029	0.073
	(0.101)	(0.087)	(0.058)	(0.064)	(0.058)
Predicted repayment × fingerprint	0.057	0.024	−0.012	0.027	−0.079
	(0.111)	(0.098)	(0.068)	(0.068)	(0.068)
Panel C					
Fingerprint × quintile 1	−0.087	0.002	0.005	−0.007	0.077
	(0.074)	(0.058)	(0.050)	(0.050)	(0.049)
Fingerprint × quintile 2	0.055	0.019	−0.015	−0.024	−0.023
	(0.055)	(0.041)	(0.039)	(0.0290)	(0.036)
Fingerprint × quintile 3	−0.006	−0.000	−0.010	−0.003	0.014
	(0.041)	(0.043)	(0.032)	(0.021)	(0.036)
Fingerprint × quintile 4	0.005	0.013	−0.022	0.003	0.003
	(0.041)	(0.041)	(0.035)	(0.019)	(0.037)
Fingerprint × quintile 5	0.007	0.036	−0.017	−0.002	−0.009
	(0.041)	(0.037)	(0.036)	(0.022)	(0.033)
Observations	520	520	520	520	520
Mean of dependent variable	0.43	0.15	0.13	0.08	0.19
Quintile 1	0.44	0.07	0.13	0.18	0.17
Quintile 2	0.49	0.10	0.13	0.13	0.15
Quintile 3	0.42	0.21	0.12	0.03	0.20
Quintile 4	0.42	0.19	0.12	0.04	0.21
Quintile 5	0.40	0.17	0.14	0.04	0.23
Joint significance of panel B coefficients (p-value)	0.69	0.62	0.45	0.54	0.21

Notes: Each column presents estimates from three separate regressions: main effect of fingerprinting in panel A, linear interaction with predicted repayment in panel B, and interactions with quintiles of predicted repayment in panel C. All regressions include stratification cell (location × week of initial club visit) fixed effects. Panel B regressions include the main effect of the level of predicted repayment, and panel C regressions include dummies for quintile of predicted repayment main effects. Standard errors on panel A coefficients are clustered at the club level, while those in panels B and C are bootstrapped with 200 replications and club-level resampling. The p-values in the bottom row are from tests that in Panel B, the fingerprinting main effect and the "Predicted repayment × fingerprint" interaction term are jointly statistically significantly different from zero. The sample is limited to individuals who took out loans in 2008 and who were included in a follow-up survey in 2009.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

to allocate 7.7 percentage points more land to paprika. This effect is roughly half the size of the paprika land allocation in the lowest quintile of predicted repayment.

It is worth considering that the effect on land allocated to paprika may be smaller than it might be otherwise because farmers began preparing and allocating land earlier in the agricultural season than our treatment. If land is less easily reallocated than other inputs from one crop to another, then we would anticipate smaller short-run effects on land allocation than on the use of inputs such as fertilizer and chemicals (to which we now turn). In the long run, when farmers incorporate the additional cost of default due to fingerprinting into their agricultural planning earlier in the season, we might find larger impacts on land allocation.

TABLE 5B—IMPACT OF FINGERPRINTING ON LAND USE

	Dependent variable: Fraction of land used for:			
	Tomatoes (1)	Leafy vegetables (2)	Cabbage (3)	All cash crops (4)
Panel A				
Fingerprint	−0.001	−0.002	−0.000	0.003
	(0.003)	(0.003)	(0.001)	(0.020)
Panel B				
Fingerprint	0.013	0.003	−0.005	0.047
	(0.009)	(0.014)	(0.004)	(0.101)
Predicted repayment × fingerprint	−0.017	−0.006	0.006	−0.057
	(0.012)	(0.015)	(0.005)	(0.111)
Panel C				
Fingerprint × quintile 1	0.008	0.004	−0.003	0.087
	(0.007)	(0.013)	(0.003)	(0.074)
Fingerprint × quintile 2	−0.003	−0.008	−0.001	−0.055
	(0.007)	(0.008)	(0.002)	(0.055)
Fingerprint × quintile 3	0.007	−0.002	−0.000	0.006
	(0.007)	(0.006)	(0.002)	(0.041)
Fingerprint × quintile 4	−0.003	−0.001	0.003	−0.005
	(0.009)	(0.007)	(0.003)	(0.041)
Fingerprint × quintile 5	−0.011	−0.003	−0.001	−0.007
	(0.008)	(0.006)	(0.002)	(0.041)
Observations	520	520	520	520
Mean of dependent variable	0.01	0.00	0.00	0.57
Quintile 1	0.01	0.01	0.00	0.56
Quintile 2	0.00	0.00	0.00	0.51
Quintile 3	0.01	0.00	0.00	0.58
Quintile 4	0.01	0.01	0.00	0.58
Quintile 5	0.01	0.01	0.00	0.60
Joint significance of panel B coefficients (p-value)	0.18	0.58	0.82	0.39

Notes: Each column presents estimates from three separate regressions: main effect of fingerprinting in panel A, linear interaction with predicted repayment in panel B, and interactions with quintiles of predicted repayment in panel C. All regressions include stratification cell (location × week of initial club visit) fixed effects. Panel B regressions include the main effect of the level of predicted repayment, and panel C regressions include dummies for quintile of predicted repayment main effects. Standard errors on panel A coefficients are clustered at the club level, while those in panels B and C are bootstrapped with 200 replications and club-level resampling. The p-values in the bottom row are from tests that in panel B, the fingerprinting main effect and the "Predicted repayment × fingerprint" interaction term are jointly statistically significantly different from zero. The sample is limited to individuals who took out loans in 2008 and who were included in a follow-up survey in 2009.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Inputs Used on Paprika.—After allocating land to different crops, the other major farming decision made by farmers is input application. Nonapplication of inputs on the paprika crop facilitates default on the loan and is therefore another form of moral hazard, again since only paprika output can feasibly be seized by the lender.

It is worth keeping in mind that input application takes place later in the agricultural cycle than land allocation, and agricultural inputs are more fungible than land. Also, inputs are added multiple times throughout the season, so farmers can incorporate new information about the cost of default into their use of inputs but cannot change land allocation after planting. Thus, we may expect use of inputs to respond more quickly to the introduction of fingerprinting than allocation of land.

Table 6 examines the effect of fingerprinting on the use of inputs on the paprika crop. The dependent variables in the first five columns (all denominated in MK) are applications of seeds, fertilizer, chemicals, man-days (hired labor), and all inputs together. Columns 6 and 7 look at, respectively, manure application (denominated in kilograms because this input is typically produced at home and not purchased) and the number of times farmers weeded the paprika plot. We view the manure- and weeding-dependent variables as more purely capturing labor effort exerted on the paprika crop, while the other dependent variables capture both labor effort and financial resources expended.

The results for paid inputs (columns 1–5) indicate that, particularly for farmers with lower likelihood of repayment, fingerprinting leads to higher application of inputs on the paprika crop. In panel B, the coefficients on the fingerprint-predicted repayment interaction are all negative in sign, and the effects on the use of fertilizer and paid inputs in aggregate are statistically significantly different from zero.[31] In panel C, the coefficient on the fingerprint-quintile 1 interaction is positive and significantly different from zero at the 1 percent confidence level for spending on seeds and is marginally significant for spending on fertilizer (*t*-statistic 1.44) and for all paid inputs (*t*-statistic 1.54). The negative and significant impact on use of paid labor in the fourth quintile is puzzling and may be attributable to sampling variation.

Results for inputs not purchased in the market are either nonexistent or ambiguous. No coefficient is statistically significantly different from zero in the regressions for manure (column 6) or times weeding (column 7).

It is worth asking whether the impact of fingerprinting seen in Table 6 means that farmers are less likely to divert input to use on other crops, or, alternatively, less likely to sell or barter the inputs for their market value. To address this, we examined the impact of fingerprinting on use of inputs on all crops combined. Results were very similar to Table 6's results for input use on the paprika crop only. (Results are available from the authors upon request.) This suggests that in the absence of fingerprinting, inputs were not used on other nonpaprika crops. (If fingerprinting simply led inputs to be substituted away from nonpaprika crops to paprika, the estimated impact of fingerprinting on input use on all crops would be zero.) It therefore seems most likely that fingerprinting made farmers less likely to dispose of the inputs via sale or barter.

In sum: for borrowers with a lower likelihood of repayment, fingerprinting leads to increased use of marketable inputs in growing paprika. While this effect is at best only marginally significant for borrowers in the lowest predicted repayment quintile, the magnitudes in that quintile are substantial. For the lowest predicted-repayment subgroup, fingerprinted farmers used MK6,566 more paid inputs in total, which is substantial compared to the mean in the lowest predicted-repayment subgroup of MK7,440.

Farm Profits.—Given these effects of fingerprinting on intermediate farming decisions such as land allocation and input use, what is the effect on agricultural revenues

[31] Joint tests, at the bottom of the table, indicate that the panel B coefficients are jointly marginally significant for the fertilizer (column 2) and all paid inputs (column 5) regressions, and jointly significant (10 percent level) for man-days (column 4).

TABLE 6—IMPACT OF FINGERPRINTING ON AGRICULTURAL INPUTS

Dependent variable	Seeds (MK) (1)	Fertilizer (MK) (2)	Chemicals (MK) (3)	Man-days (MK) (4)	All paid inputs (MK) (5)	KG manure (6)	Times weeding (7)
Panel A							
Fingerprint	84.536	1037.378	357.103	–408.599**	1,070.419	44.863	0.048
	(54.312)	(1297.753)	(219.533)	(188.581)	(1,523.582)	(37.258)	(0.141)
Panel B							
Fingerprint	279.401**	10953.768*	424.276	425.804	12,083.249*	34.902	0.176
	(138.828)	(5856.494)	(510.020)	(484.087)	(6270.836)	(162.308)	(0.450)
Predicted repayment × fingerprint	−243.229	−12236.414*	−77.737	−1043.976	−1,3601.356*	12.306	−0.162
	(185.108)	(6503.934)	(673.749)	(663.119)	(7,108.123)	(183.395)	(0.570)
Panel C							
Fingerprint × quintile 1	214.555***	5852.606	384.382	114.901	6566.444	56.139	0.406
	(82.610)	(4058.444)	(339.435)	(207.522)	(4,262.700)	(124.425)	(0.329)
Fingerprint × quintile 2	91.985	4241.768	260.137	−206.938	4,386.952	53.378	−0.379
	(96.385)	(3043.436)	(400.035)	(443.448)	(3,383.939)	(75.779)	(0.314)
Fingerprint × quintile 3	121.291	−316.432	484.330	−427.907	−138.718	97.375	−0.118
	(107.279)	(2332.776)	(449.485)	(485.596)	(2,741.751)	(95.154)	(0.320)
Fingerprint × quintile 4	−18.632	−1315.729	201.476	−973.256*	−2,106.140	−8.974	−0.196
	(121.874)	(2501.755)	(457.487)	(532.087)	(3,066.267)	(73.012)	(0.328)
Fingerprint × quintile 5	47.757	−1874.942	431.041	−417.203	−1,813.347	37.594	0.548
	(121.474)	(2343.55)	(438.313)	(561.762)	(2,853.287)	(91.110)	(0.362)
Observations	520	520	520	520	520	520	520
Mean of dependent variable	247.06	7499.85	671.31	665.98	9,084.19	90.84	1.94
Quintile 1	174.13	6721.24	401.30	143.48	7,440.15	97.39	1.47
Quintile 2	140.00	6080.46	620.67	238.94	7,080.08	39.25	1.55
Quintile 3	269.90	8927.65	674.48	836.98	10,709.00	105.73	2.05
Quintile 4	292.07	7649.51	715.08	936.29	9,592.95	93.23	2.24
Quintile 5	340.18	8078.58	892.05	1065.18	10,375.99	118.13	2.28
Mean of dependent variable (US $)	1.70	51.72	0.63	4.59	62.65	NA	NA
Joint significance of panel B coefficients (p-value)	0.19	0.14	0.44	0.07	0.13	0.47	0.35

Notes: Each column presents estimates from three separate regressions: main effect of fingerprinting in panel A, linear interaction with predicted repayment in panel B, and interactions with quintiles of predicted repayment in panel C. All regressions include stratification cell (location × week of initial club visit) fixed effects. Panel B regressions include the main effect of the level of predicted repayment, and panel C regressions include dummies for quintile of predicted repayment main effects. Standard errors on panel A coefficients are clustered at the club level, while those in panels B and C are bootstrapped with 200 replications and club-level resampling. The p-values in the bottom row are from tests that in panel B, the fingerprinting main effect and the "Predicted repayment × fingerprint" interaction term are jointly statistically significantly different from zero. The sample is limited to individuals who took out loans in 2008 and who were included in a follow-up survey in 2009.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

and profits? Table 7 presents regression results where the dependent variables are market crop sales, the value of unsold crops, and profits (market sales plus value of unsold crops minus value of inputs used), all denominated in MK. The magnitudes of the overall impacts of fingerprinting on value of sales, unsold harvest, and total profits (panel A), and in the bottom two quintiles (panel C), are large and positive, but the effects are imprecisely estimated and none are statistically significantly different from zero. To help deal with the problem of outliers in the profit figures, column 4 presents regression results where the dependent variable is the natural log of agricultural profits.[32] The effect of fingerprinting in the bottom quintile of predicted repayment is positive but not statistically significant (*t*-statistic 1.21). Joint tests (reported at the bottom of the table) indicate that the panel B coefficients are jointly significant at the 10 percent level for market crop sales and log agricultural profits.

In sum, then, it remains possible that increased use of paid inputs led ultimately to higher revenue and profits among fingerprinted farmers in our sample, but the imprecision of the estimates prevents us from making strong statements about the impact of fingerprinting on farm profits.

V. Discussion and Additional Analyses

In sum, the results indicate that for the lowest predicted-repayment quintile, fingerprinting leads to substantially higher loan repayment. In seeking explanations for this result, we have provided evidence that for this subgroup, fingerprinting leads farmers to take out smaller loans, devote more land to paprika, and apply more inputs on paprika.

In the context of our theoretical model, we interpret these results as indicating that, for the farmers with the lowest ex ante likelihood of repayment, fingerprinting reduces adverse selection and ex ante moral hazard. The reduction in adverse selection (a reduction in the riskiness of the loan pool) comes about not via the extensive margin of loan approval and take-up, but through farmers' decisions to take out smaller loans if they are fingerprinted (the intensive margin of loan take-up).

In this section we summarize the results of additional robustness checks that are presented in greater detail in the online Appendix (online Appendix F provides further detail on all analyses discussed below). We then provide additional evidence that our results are not likely to reflect reductions in ex post moral hazard. Finally, we report results of a test of the positive correlation property that reveals the presence of asymmetric information.

Impact of Fingerprinting in Full Sample.—Most results presented so far are for the subsample of farmers who took out a loanWe have argued that when restricting ourselves to this subsample, estimated treatment effects are not confounded by selection concerns because treatment has no statistically significant effect on selection into borrowing, either on average or in interaction with predicted repayment (Table 3, column 2). That said, one may raise a concern about statistical power: 95 percent confidence intervals around the point estimates in Table 3, column 2 admit nonnegligible effects of treatment on selection into borrowing. The concern would be that there was in fact selection into borrowing in response to fingerprinting, which would cloud the interpretation of our results. For example, one might worry that that fingerprinting led borrowers in quintile 1 of predicted repayment to be on

[32] For seven observations, profits are zero or negative, and in these cases ln(profits) is replaced by 0. These observations do not drive the results; results are essentially identical when these observations are excluded.

Table 7—Impact of Fingerprinting on Agricultural Profits

Dependent variable	Market sales (self report, MK) (1)	Value of unsold harvest (regional prices, MK) (2)	Profits (market sales + value of unsold harvest − cost of inputs, MK) (3)	Ln(profits) (4)
Panel A				
Fingerprint	5,808.270	3,571.446	11,457.127	0.043
	(9,376.512)	(10,525.289)	(14,071.809)	(0.094)
Panel B				
Fingerprint	70,072.212	−33,288.163	23,141.194	0.682
	(55,485.33)	(59,698.16)	(78,424.26)	(0.426)
Predicted repayment × fingerprint	−79,154.870	42,761.646	−16,899.862	−0.793
	(57,954.66)	(72,680.72)	(91,455.98)	(0.495)
Panel C				
Fingerprint × quintile 1	32,123.244	168.559	25,730.854	0.434
	(39,966.77)	(33,675.88)	(53,903.61)	(0.359)
Fingerprint × quintile 2	44,570.113	11,320.950	53,763.203	0.249
	(36,480.13)	(62,585.49)	(71,637.71)	(0.260)
Fingerprint × quintile 3	−22,828.231	−21,841.768	−35,149.036	−0.242
	(17,754.30)	(59,742.76)	(63,280.91)	(0.218)
Fingerprint × quintile 4	−14,121.491	19,577.449	12,923.138	−0.036
	(14,551.19)	(47,941.73)	(50,773.18)	(0.218)
Fingerprint × quintile 5	−2,035.855	6,086.165	4,342.320	−0.078
	(14,838.68)	(58,414.17)	(61,584.75)	(0.230)
Observations	520	520	520	520
Mean of dependent variable	65,004.30	80,296.97	117,779.16	11.44
Quintile 1	60,662.57	82,739.24	121,222.50	11.36
Quintile 2	89,028.25	29,995.27	91,652.71	11.55
Quintile 3	57,683.74	96,247.91	123,242.30	11.44
Quintile 4	61,088.27	104,927.50	136,467.50	11.45
Quintile 5	56,593.43	85,817.08	115,172.50	11.39
Mean of dependent variable (US $)	448.31	553.77	812.27	NA
Joint significance of panel B coefficients (*p*-value)	0.08	0.56	0.20	0.08

Notes: Each column presents estimates from three separate regressions: main effect of fingerprinting in panel A, linear interaction with predicted repayment in panel B, and interactions with quintiles of predicted repayment in panel C. All regressions include stratification cell (location × week of initial club visit) fixed effects. Panel B regressions include the main effect of the level of predicted repayment, and panel C regressions include dummies for quintile of predicted repayment main effects. Standard errors on panel A coefficients are clustered at the club level, while those in panels B and C are bootstrapped with 200 replications and club-level resampling. The *p*-values in the bottom row are from tests that in panel B, the fingerprinting main effect and the "Predicted repayment × fingerprint" interaction term are jointly statistically significantly different from zero. The sample is limited to individuals who took out loans in 2008 and who were included in a follow-up survey in 2009.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

average different from control group borrowers in quintile 1 (along various observed and unobserved dimensions) in ways that make them more likely to repay, to devote land to paprika, and to use fertilizer on paprika.

Analyses of the full sample of farmers, without restricting the sample only to borrowers, can help address such concerns about selection bias. Estimated effects of treatment (and interactions with predicted repayment) would then represent effects of being fingerprinted on average across treated individuals, whether or not the individual took out a loan. While such an analysis makes little sense for outcomes specific to loans such as repayment (as in the outcomes in Table 4), we carry out this analysis for the other examined variables from the August 2008 follow-up survey, namely land use, input use, and profits (the outcomes in Tables 5A, 5B, 6, and 7).

As it turns out, full-sample regression results are very similar to those from the borrower-only regressions. The general pattern is for coefficients that were significant before to remain statistically significant, but to be only around half the magnitude of the coefficients in the borrowing sample regressions. This reduction in coefficient magnitude is consistent with effect sizes in the full sample representing a weighted average of no effects for nonborrowers and nonzero effects for borrowers (slightly less than half of individuals in the full sample are borrowers). We conclude that selection into borrowing is not driving the treatment effect estimates of Tables 5A, 5B, 6, and 7.

Results with "Simple" Predicted Repayment Regression.—Results discussed so far on treatment effect heterogeneity construct the predicted repayment variable from the regression in column 3 of Table 2. The right-hand side of this regression has farmer-level characteristics, as well as stratification cell (locality × week of initial club visit) fixed effects.

Because the baseline farmer-level characteristics listed in Table 2 are the most readily interpretable, we check the robustness of the results to constructing predicted repayment using only baseline farmer-level characteristics. The alternative predicted repayment regression is that of column 3 of Table 2, except that stratification cell fixed effects are dropped. This regression is then used to predict repayment for the full sample, and the predicted repayment variable is interacted with treatment to examine heterogeneity in the treatment effect.

Regression results are very similar when using this simpler index of predicted repayment. Overall, the general conclusion stands: fingerprinting has more substantial effects on repayment and activities on the farm for individuals with lower predicted repayment, even when repayment is predicted using only a restricted set of baseline farmer-level variables.

Results Where Predicted Repayment Coefficients Obtained from Partition of Control Group.—In heterogeneous treatment effect results presented so far, there may be a concern that, for idiosyncratic reasons, control farmers in some geographic areas could have unusually low repayment rates compared to treatment farmers in the same areas. If this were the case, then the main analyses we have conducted so far might mechanically find a positive effect of treatment in cohorts where control group farmers had idiosyncratically low repayment rates.

We address this type of concern in two ways. First, we point to the robustness check described above, where we find that results are very similar when the predicted repayment index is estimated without stratification cell fixed effects. These results reveal that the patterns of treatment effect heterogeneity we emphasize are not simply an artifact of inclusion of these (locality × week of initial club visit) fixed effects in the predicted repayment regression.

Second, we gauge the extent to which our main results diverge from those of an alternative approach that involves partitioning the control group into two parts: one part used to generate coefficients in the predicted repayment regression, and the other part used as a counterfactual for the treatment group in the main regressions. Because observations used to generate coefficients in the auxiliary predicted repayment regression are not then used as counterfactuals for the treatment observations, this approach avoids the possibility that our results arise mechanically from overfitting the repayment model.

Due to sampling variation, different randomly determined partitions of the control group will yield different results, so we conduct this exercise 1,000 times and then examine the distribution of the regression coefficients generated. We focus our attention on coefficients on the interaction between the treatment indicator and the indicator for quintile 1 of predicted repayment (in panel C) for the dependent variables of Tables 3 to 7.

We find that in all cases the quintile 1 interaction term coefficient falls within the 95 percent confidence interval of the coefficients generated in the partitioning exercise. Furthermore, whenever the interaction term coefficient is statistically significantly different from zero in Tables 3 to 7, the 95 percent confidence interval of the coefficients generated in the partitioning exercise does not include zero or coefficients of the opposite sign.

We therefore conclude that our main results are not mechanically driven by idiosyncratically low repayment among some control farmers in certain localities.

Evidence for a Reduction in Ex Post Moral Hazard.—Reductions in ex ante moral hazard may help encourage higher loan repayment by improving farm output so that farmers have higher incomes with which to make loan repayments. Reductions in adverse selection—reduced loan sizes for the "worst" borrowers—also help increase repayment performance. But a question that remains is whether any of the increase in repayment is due to reductions in ex post moral hazard. In other words, are there reductions in strategic or opportunistic default by borrowers, holding constant loan size and farm profits?

We investigate this by running regressions for repayment outcomes, but including controls for profits and loan size. Results are reported in online Appendix Table 12.[33] The profits and total borrowed variables are flexibly specified as indicators for the borrower being in the first through tenth decile of the distribution of the variable (one indicator is excluded in each resulting group of ten indicators.)

When controlling for loan size and profits, the effect of fingerprinting on on-time repayment for the worst borrowers declines in magnitude. For all key coefficients in columns 1–3 (those on the panel B interaction term and the panel C interaction with quintile 1), magnitudes fall substantially vis-à-vis corresponding estimates in Appendix Table 4. The tests of differences in these coefficients vis-à-vis those in Appendix Table 4, reported in the bottom of the table, indicate that the key coefficients are statistically significantly different when the controls for loan size and profits are included in the regression. That said, in the regression for "Balance, September 30," the linear interaction term and the interaction term with quintile 1 of predicted repayment remain statistically significant at the 5 percent and 10 percent

[33] We restrict to the $N = 520$ sample because of the need to control for profits, which was only observed in the August 2008 survey. These results should be compared with online Appendix Table 4, which is for the same sample.

levels, respectively. The interaction term with quintile 1 is also significant at the 10 percent level in the "Fraction Paid by September 30" regression.

Results for eventual repayment are less conclusive. We cannot reject the hypothesis that fingerprinting has no effect on eventual repayment (columns 4–6) once we control for agricultural profits and original loan size. Coefficient estimates that were previously statistically significant (in online Appendix Table 4) are now uniformly smaller in magnitude and not statistically significantly different from zero. But significance tests at the bottom of the table indicate that for five out of six key coefficients in columns 4–6 (those on the panel B interaction term and the panel C interaction with quintile 1), we cannot reject the null that corresponding coefficients in Appendix Table 4 are the same.

These results suggest that when the dependent variable is on-time repayment, reductions in both ex ante and ex post moral hazard may be driving the increase in repayment: effects of fingerprinting for the worst borrowers remain statistically significant (or nearly so) in columns 1–3 even when controls for loan size and profits are included in the regression (suggesting a role for ex post moral hazard), but are statistically significantly smaller in magnitude than when such controls are not included (consistent with the presence of ex ante moral hazard). For eventual repayment (columns 4–6), this test is inconclusive because coefficients decline but are not statistically significantly different when loan size and profits controls are added to the regression.

Test of the Positive Correlation Property.—Following several recent articles that use data from insurance markets to test for the presence of asymmetric information (Chiappori and Salanié 2000, 2003; Chiappori et al. 2006), the predictions of the theoretical model of Section III can be used to perform a similar test. In the insurance market context, many models of adverse selection and possibly moral hazard that assume competitive insurance markets predict a positive correlation between coverage and the probability of the event insured, conditioning on the information available to the insurer. In our context, the test involves a positive correlation between loan size and default.

In order to test this prediction, multiple loan contracts must coexist in equilibrium, but according to the model (see Appendix Figure 3), all agents should borrow the high amount b_H when dynamic incentives cannot be used, and so there should be no correlation. With dynamic incentives, however, both high and low loan sizes (b_L and b_H) will be taken, and so the correlation can be tested. Using data on the loan size and default at maturity date, we find, as expected, no correlation for borrowers in the control group (t-stat = 1.13), but find a strong positive correlation in the treatment group (t-stat = 3.30). In the treatment group, a MK1,000 increase in the loan amount is associated with an increase in the probability of default (not being fully paid at the loan due date) of roughly 3 percentage points.

VI. Benefit-Cost Analysis

The analysis so far has estimated the gains to the financial institution (MRFC) from using fingerprinting to identify new borrowers as part of the process of loan screening. These gains need to be weighed against the costs of fingerprinting. We conduct a benefit-cost analysis of biometric fingerprinting of borrowers. The analysis is most valid for institutions similar in characteristics to those of our partner institution, MRFC, but we have made the elements of the calculation very transparent so that they can easily be modified for other institutions with different characteristics.

Under reasonable assumptions, total benefit per individual fingerprinted is MK490.50 (US$3.38). We consider three types of costs: equipment costs (which need to be amortized across all farmers fingerprinted), loan officer time costs, and transaction costs per fingerprint checked. Summing these costs, total cost per individual fingerprinted is MK209.20. The net benefit per individual fingerprinted is therefore MK266.30 (US$1.84), and the benefit-cost ratio is an attractive 2.34. (Details of this calculation are in online Appendix G.)

For several reasons, this benefit-cost calculation is likely to be quite conservative. First of all, under reasonable circumstances some of the individual costs could be brought down considerably. The cost for equipment units could fall substantially if a fingerprinting function were integrated into equipment packages that had multiple functionalities, such as the handheld computers that MRFC is considering providing for all of its loan officers. Transaction costs for fingerprint checking could fall due to volume discounts if the lending institution banded together with other lenders to channel all their fingerprint identification through a single service provider (in the context of a credit bureau, for example).

In addition, there are other benefits to the lending institution that this benefit-cost calculation is not capturing. The impact of fingerprinting on loan repayment may become larger in magnitude over time as the lender's threat of enforcement becomes more credible. We have also assumed that all the benefits come from fingerprinting new loan customers (the subject of this experiment), but there may also be increases in repayment among existing customers who are fingerprinted (on which this experiment does not shed light). Finally, there may be broader benefits that are not captured by the lending institution, such as increased income due to more intensive input application by fingerprinted farmers.[34]

VII. Conclusion

We conducted a field experiment where we randomly selected a subset of potential loan applicants to be fingerprinted, which improved the effectiveness of dynamic repayment incentives for these individuals. For all the recent empirical work on microcredit markets in developing countries, to our knowledge this is the first randomized field experiment of its kind, and the first to shed light (thanks to a detailed follow-up survey of borrowers) on the specific behaviors germane to the presence of asymmetric information problems.

We find heterogeneous effects of being fingerprinted, with the strongest effects among borrowers expected (ex ante) to have the worst repayment performance. Fingerprinting leads these "worst" borrowers to raise their repayment rates dramatically, partly as a result of voluntarily choosing lower loan sizes as well as devoting more agricultural inputs to the cash crop that the loan was intended to finance. In the

[34] Unfortunately, our estimates of the impact of fingerprinting on profits are too imprecise to say whether profits definitely increased due to this intervention.

context of a simple model of asymmetric information in a credit market, we interpret the treatment-induced reduction in loan size as a reduction in adverse selection, and the increase in agricultural input use as a decline in moral hazard.[35]

The short-term improvements in repayment estimated in this paper may indeed be smaller than the effects that would be found over a longer horizon. First of all, borrowers' assessments of the effectiveness of the technology and the credibility of the threat to withhold credit would likely rise over time as they gained further exposure to the system, observed that their past credit performance was being correctly retrieved by the lender, and saw that credit history information was indeed being shared with other lenders. In addition, the lender should be able to selectively allocate credit to the pool of good-performing borrowers over time, further improving overall repayment performance of the borrowing pool. Finally, because there is less risk involved for the lender, the credit contract terms could be made more attractive to borrowers, which may further improve repayment.[36]

By revealing the presence of specific asymmetric information problems and the behaviors that result from them, this paper's findings can help guide future theoretical work on rural credit markets. To be specific, models of credit markets in contexts similar to rural Malawi should allow for adverse selection on the intensive margin of loan take-up (i.e., the choice of loan size), ex ante moral hazard (actions during the production season that may affect farm profits), and ex post moral hazard (strategic or opportunistic default).[37]

Our results also have implications for microlending practitioners, by quantifying the benefits from exploiting a commercially available technology to raise repayment rates. Beyond improving the profitability and financial sustainability of microlenders, increased adoption of fingerprinting (or other identification technologies) can bring additional benefits if lenders are thereby encouraged to expand the supply of credit, and if this expansion of credit supply has positive effects on household well-being.[38] Credit expansions enabled by improved identification technology may be particularly large in previously underserved areas, such as the rural sub-Saharan context of our experiment, where problems with personal identification are particularly severe.

Another potential implication of this research is that in the absence of an alternative national identification system, fingerprints could serve as the unique identifier

[35] In practice, adverse selection and moral hazard may be more intertwined than is typically formulated in theoretical models (see Karlan and Zinman 2009). In the context of this experiment, one could, in principle, isolate the various asymmetric information problems by fingerprinting borrowers at different points in time along the loan cycle. For example, a subset of borrowers (group 1) could be fingerprinted before loan decisions are made, then another group (group 2) could be fingerprinted immediately after loans are granted but before funds are invested into production, and yet another group (group 3) could be fingerprinted once production has taken place but before repayment. A final group of borrowers would not be fingerprinted (group 0). With full compliance—that is, when all subjects agree to be fingerprinted—one could then measure adverse selection by comparing group 1 and 2, and could measure ex ante moral hazard by comparing 2 and 3, and strategic default by comparing 3 and 0. Given the number of farmers in our study, it was infeasible to implement this design because power calculations suggested we could have at best two groups. Our study therefore consists of groups 0 and 1.

[36] After learning about the benefits of biometric technology, MRFC applied for a grant from a donor agency to finance the purchase of handheld devices and software to mainstream the collection of biometric information from all its clients. Opportunity International Bank of Malawi, a competitor that operates in mostly urban areas, collects an electronic fingerprint from every borrower.

[37] But keep in mind that our results on ex post moral hazard must be taken as merely suggestive.

[38] To be sure, this research sheds no light on the impact of microcredit availability on household well-being.

that allows individual credit histories to be stored and accessed in a cross-lender credit bureau. It has been noted that a key obstacle to establishment of credit bureaus is the lack of a unique identification system (Conning and Udry 2007; Fafchamps 2004; Mylenko 2008). Our results indicate that borrowers (particularly the worst borrowers) do perceive fingerprinting as an improvement in the lender's dynamic enforcement technology, and so support the use of fingerprints as an identifier in a national credit bureau.

As is the case with all empirical analysis, it is important to replicate this study in other contexts to gauge the external validity of the results. Our experiment was conducted in a context where there is currently no unique identification system and the credit market is still undeveloped. So while our findings might approximate impacts in other parts of rural sub-Saharan Africa with similar levels of economic and financial development, effects in other environments could be quite different. It would be important to gauge the extent to which impacts are different in populations that are, for example, more urban, more accessible to microcredit, and for which personal identification technologies (e.g., government-issued photo ID) have been implemented more widely. As mentioned above, the effects of fingerprinting on repayment could very well rise over time, and so future studies should monitor effects beyond a single loan cycle. Future work should also make sure to examine responses by the lender, such as changes in the credit contract, approval rates, or in loan officer monitoring. While in our case loan officers did not behave differently toward treated borrowers, in other contexts, perhaps under different loan officer incentives, this may not be the case. We view these and related questions as promising areas for future research.

REFERENCES

Ausubel, L. M. 1999. "Adverse Selection in the Credit Card Market." Unpublished.

Bencivenga, Valerie R., and Bruce D. Smith. 1991. "Financial Intermediation and Endogenous Growth." *Review of Economic Studies* 58 (2): 195–209.

Chiappori, Pierre-André, Bruno Jullien, Bernard Salanié, and François Salanié. 2006. "Asymmetric Information in Insurance: General Testable Implications." *Rand Journal of Economics* 37 (4): 783–98.

Chiappori, Pierre-Andre, and Bernard Salanié. 2000. "Testing Asymmetric Information in Insurance Markets." *Journal of Political Economy* 108 (1): 56–78.

Chiappori, Pierre-Andre, and Bernard Salanié. 2003. "Testing Contract Theory: A Survey of Some Recent Work." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, edited by Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, 115–49. Cambridge: Cambridge University Press.

Conning, Jonathan, and Christopher Udry. 2007. "Rural Financial Markets in Developing Countries." In *The Handbook of Agricultural Economics, Vol. 3: Farmers, Farm Production, and Farm Markets*, edited by Robert E. Evenson and Prabhu Pingali, 2857–908. Amsterdam: Elsevier Science.

de Janvry, A., C. McIntosh, and E. Sadoulet. 2010. "The Supply and Demand Side Impacts of Credit Market Information." *Journal of Development Economics* 93 (2): 173–88.

Edelberg, Wendy. 2004. "Testing for Adverse Selection and Moral Hazard in Consumer Loan Markets." Board of Governors of the Federal Reserve System Finance and Economics Discussion Paper Series 2004-9.

Efron, Bradley, and Robert Tibshirani. 1993. "An Introduction to the Bootstrap." Monograph on Statistics and Applied Probability 57.

Fafchamps, Marcel. 2004. *Market Institutions in Sub-Saharan Africa: Theory and Evidence*. Cambridge, MA: MIT Press.

Ghosh, Parikshit, Dilip Mookherjee, and Debraj Ray. 2001. "Credit Rationing in Developing Countries: An Overview of the Theory." In *Readings in the Theory of Economic Development*, edited by Dilip Mookherjee and Debraj Ray, 283–301. Malden, MA: Blackwell.

Giné, Xavier, Jessica Goldberg, and Dean Yang. 2012. "Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi: Dataset." *American Economic Review*. http://dx.doi.org/10.1257/aer.102.6.2923.

Giné, Xavier, Pamela Jakiela, Dean Karlan, and Jonathan Morduch. 2010. "Microfinance Games" *American Economic Journal: Applied Economics* 2 (3): 60–95.

Giné, Xavier, and Stefan Klonner. 2005. "Financing a New Technology in Small-Scale Fishing: the Dynamics of a Linked Product and Credit Contract." Unpublished.

Giné, Xavier, and Dean Yang. 2009. "Insurance, Credit, and Technology Adoption: Field Experimental Evidence from Malawi." *Journal of Development Economics* 89 (1): 1–11.

Government of India. 2005. *Performance Evaluation of Targeted Public Distribution System* (*TPDS*). New Delhi: Government of India.

Karlan, Dean, and Jonathan Zinman. 2009. "Observing Unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment." *Econometrica* 77 (6): 1993–2008.

King, Robert G., and R. Levine. 1993. "Finance and Growth: Schumpeter Might Be Right." *Quarterly Journal of Economics* 108 (3): 717–37.

Klonner, S., and A. S. Rai. 2009. "Adverse Selection in Credit Markets: Evidence from Bidding ROSCAs." Unpublished.

Ligon, Ethan. 1998. "Risk-Sharing and Information in Village Economies." *Review of Economic Studies* 65 (4): 847–64.

Livshits, Igor, James MacGee, and Michele Tertilt. 2010. "Accounting for the Rise in Consumer Bankruptcies." *American Economic Journal: Macroeconomics* 2 (2): 165–93.

Mylenko, Nataliya. 2008. "Developing Credit Reporting in Africa: Opportunities and Challenges." Paper presented at African Finance for the 21st Century. Seminar organized by the IMF Institute and the Joint Africa Institute. Tunis, Tunisia.

Padilla, J., and M. Pagano. 2000. "Sharing Default Information as a Borrower Discipline Device." *European Economic Review* 44 (10): 1951–80.

Pagano, Marco, and Tullio Jappelli. 1993. "Information Sharing in Credit Markets." *Journal of Finance* 48 (5): 1693–1718.

Paulson, Anna L., Robert M. Townsend, and Alexander Karaivanov. 2006. "Distinguishing Limited Liability from Moral Hazard in a Model of Entrepreneurship." *Journal of Political Economy* 114 (1): 100–44.

Polgreen, Lydia. 2011. "Scanning 2.4 Billion Eyes, India Tries to Connect Poor to Growth." *New York Times*, September 2, 2011, p. 1.

Stiglitz, Joseph E., and Andrew Weiss. 1983. "Incentive Effects of Terminations: Applications to the Credit and Labor Markets." *American Economic Review* 73 (5): 912–27.

Visaria, S. 2009. "Legal Reform and Loan Repayment: The Microeconomic Impact of Debt Recovery Tribunals in India." *American Economic Journal: Applied Economics* 1 (3): 59–81.