Credit Scoring Meets Agricultural Lending: Exogenous Shocks, Recovery, and Access to Formal Credit

Nicolás de Roux†

March 15, 2017

Abstract

Credit scoring has become a widespread tool to assess the creditworthiness of prospective borrowers, and has been found to increase efficiency and welfare in many settings. But this paper identifies a shortcoming in existing credit scoring systems that may lead to a market failure in agricultural lending in developing countries: Farmers’ scores – and their access to credit – decline because of exogenous short-term weather shocks that do not reduce their likelihood of future repayment. I use data on the near universe of formal agricultural loans for coffee production in Colombia to show that excess rainfall shocks cause lower concurrent loan repayment, lower credit scores, and more frequent denial of subsequent loan applications. Drawing on the agronomic literature on coffee production and using survey data, I show that productivity, income and repayment behavior recover faster from these shocks than farmers’ credit histories. The additional loan denials create costs for both farmers and the lender that could be avoided. The results suggest that incorporating verifiable information on individual level shocks into credit scores would increase the efficiency of credit markets.

* A previous version of this paper circulated as my job market paper. I am especially grateful to my sponsor Eric Verhoogen for his support and guidance. I would also like to thank Pierre-André Chiappori, Jonas Hjort and Suresh Naidu for their advice. For useful comments I am grateful to Emily Breza, Zach Brown, Juan Camilo Cárdenas, Felipe Cortés, Xavier Durán, Andrés Escobar, Marcela Eslava, Alejandro Fajardo, Margarita Gáfaro, Chris Hansman, Juan Herreño, Marc Hofstetter, Supreet Kaur, Wojciech Kopczuk, Ajin Lee, Ana María Ibáñez, Luis R. Martínez, Josepa Miquel-Florensa, Andrés Moya, Juan Carlos Muñoz-Mora, José Antonio Ocampo, Kiki Pop-Eleches, Evan Riehl, Yogita Shamdasani, Mica Sviatschi, Diego Ubfal, Miguel Urquiola and to seminar participants at Universidad de los Andes, Universidad Javeriana and Universidad del Rosario. This project could not have been carried out without the hospitality of the Banco Agrario de Colombia and the knowledge of its staff. Many thanks to all members of the Vice-presidency of Risk and the Information Management Unit, and especially to Jairo Esquivel, Moisés Mahecha and Guillermo Otero. Finally, I am grateful to the Federación Nacional de Cafeteros, and in particular to Óscar Bernal, Julián Cantor and Marcela Urueña, and to the the Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia. All errors are my own.

† Department of Economics, Universidad de los Andes, email: nicolas.de.roux@uniandes.edu.co
“If any one owe a debt for a loan, and a storm prostrates the grain, or the harvest fail, or the grain does not grow for lack of water; in that year he need not give his creditor any grain, he washes his debt-table in water and pays no rent for his year.”

— Hammurabi’s Code (c. 1760 B.C)

1 Introduction

Credit scoring has spread quickly across the banking and lending industry in both developed and developing countries (Mester, 1997; Berger, Frame, and Miller, 2005; de Janvry, McIntosh, and Sadoulet, 2010). Theoretically, this tool has the potential to lessen problems of moral hazard and adverse selection (Pagano and Jappelli, 1993; Giné, Goldberg, and Yang, 2012) and many empirical studies have found welfare and efficiency gains from its use (Einav, Jenkins, and Levin, 2012; Einav, Jenkins, and Levin, 2013; Giné, Goldberg, and Yang, 2012).

Existing credit scores are primarily calculated using fixed borrower characteristics and information on prior repayment behavior. These methods are designed for situations in which a borrower’s repayment depends largely on her type or her own actions. But credit scores typically do not take into account information on exogenous shocks that affect the economic environment in which borrowers are operating. Therefore, they may not be as well suited for settings in which large, exogenous shocks that influence repayment are important.

This paper documents a market failure that results from an important shortcoming of credit scores used in agricultural lending in developing countries. In particular, I study a setting in which farmers’ scores and their access to credit decline because of exogenous short-term shocks that do not reduce their likelihood of future repayment. To do this, I use a novel administrative data set with the near-universe of formal loans to small farmers in Colombia. I first show that exogenous weather shocks lead to lower repayment of concurrent loans, lower credit scores, and more frequent denial of subsequent loan applications. I then show that income recovers from these shocks faster than farmers’ access to credit. Furthermore, I present evidence on recovery in loan payments in two different samples, consistent with the recovery in income. These findings imply a market failure because banks using information on weather shocks would want to lend to farmers who are currently denied credit. The implication is that the efficiency of the credit market could be enhanced if credit scores and credit histories were modified to account for these exogenous shocks.

From a theoretical perspective, the market failure arises from the fact that a signal that conveys information on the action and the type of the borrower is not incorporated in the contract with the bank. In my setting, rainfall shocks provide such information (for example, the actions of a borrower who defaults in the absence of a rainfall shock are likely different from the actions of a borrower who defaults when a rainfall shock occurs) and should be incorporated in the credit
score or, more broadly, in the contract between the farmer and the bank.¹

To document the channels described above I proceed in two steps. In the first part of the paper, I estimate the effect of rainfall shocks on concurrent loan repayment, credit scores, and future access to credit, and I establish that this effect is persistent. In the second part, I argue that, on average, farmer income from coffee production and repayment ability recover faster than farmer credit histories after a shock.

For the first step, I use administrative data on loan applications, credit scores, and repayment behavior from the Banco Agrario de Colombia (BAC) [Colombian Agrarian Bank]. The BAC is a publicly owned bank that was created in 1999 to finance rural productive activities. In 2013, it held 97% of formal agricultural loans to small farmers in Colombia (DNP and FINAGRO, 2014). I merge BAC’s individual data from 2005 to 2015 with administrative data from the Federación Nacional de Cafeteros (FNC) [National Federation of Coffee Growers] that allows me to recover geographical coordinates for each farmer and link the loan to the closest rainfall station at the time of loan disbursement.

I estimate the effect of excessive rain shocks during loan tenure on repayment in a sample of loans destined for coffee production. Excessive rainfall is associated with lower productivity of coffee trees.² Therefore, it can lead to a fall in farmer income and to lower loan repayment, as long as insurance markets are incomplete.³ The estimation specifications in these exercises include rainfall station and quarter-year of disbursement times maturity fixed effects. Thus, the identifying assumption is that the occurrence of shocks at different points in time for a given rainfall station is not systematically correlated with other time-varying factors that affect repayment of outstanding loans of farmers close to the station.

I find that excessive rainfall decreases the probability of loan repayment. In particular, loans that experience excess rainfall shocks during their first year have a probability of entering into a period of 30 days of overdues that is 22 percent larger than loans with no shocks. This leads the BAC to downgrade the scores of its clients that are reported to credit bureaus.

I then estimate the effect of rainfall shocks on subsequent loan applications. The loan application process in the BAC consists of two stages. I find that the score reported by the credit bureau is lower and that denial at both stages of the application is more likely for farmers who experienced a shock during the previous loan tenure. For example, in my preferred specification, the probability of loan denial is 13 percent larger in the first stage of the process for applications

¹The Informativeness Principle states that “any additional information about the agent’s action, however imperfect, can be used to improve the welfare of both the principal and the agent” (Holmstrom, 1979). In this regard, in the Appendix I show in a simple model of borrower screening how the the inclusion of observable exogenous shocks in the credit score can reduce the probability of both an inclusion error (lending to a un-profitable borrower) and an exclusion error (denying credit to a un-profitable one).
²I focus on excessive rainfall since coffee production is more sensitive to it than to a lack of rain. See for example Boucher and Moya (2015).
³This is the case for agriculture in Colombia, and in particular for coffee production (Boucher and Moya, 2015).
following loan tenures with rainfall shocks. In the second stage the probability of denial increases by 10 percent. Using a subset of the sample of this exercise, I show that even after two years since the maturity of the first loan, applications following loan tenures where a shock occurred have lower credit scores and are more likely to be denied.

Having documented the effect of rainfall shocks on repayment and future access to credit, I proceed to the second step, which shows that income recovers faster from rainfall shocks than the credit history of the farmer. Based on an extensive agronomic literature, I argue that excessive rainfall in the year before the harvest affects the productivity of coffee trees, but the next harvest is not affected after weather returns to normal. This suggests that the effect of the shocks on the productivity of the coffee tree dies out after one year at most. To test this in the data, I use a representative survey of coffee farmer sales in 2005 to show that income recovers from excessive rainfall shocks on average one year after the shock. The identifying assumption for this exercise is that, in the cross-section, the occurrence of excess rainfall shocks is not correlated with unobservables that might affect coffee sales. I present various robustness checks that support it.

A question that arises at this point is whether the recovery in income translates into a recovery in repayment. One would expect this to be the case since income is the main determinant of the ability of lenders to repay their loans. I present two pieces of evidence of repayment recovery after rainfall shocks. First, in a sample of loans with maturities of five years or more, I show that repayment recovers two years after a shock in the first year of the five year loan tenure. Second, in a sample of individuals with high credit scores, I show that a shock during the tenure of an initial loan causes lower repayment but has no significant effect on acceptance of subsequent loan applications. This suggests that selection driven by shocks is not an important concern in this sample. More importantly, I find no significant effect of shocks during the first loan on repayment of the subsequent loan.

The idea that a transitory shock can affect credit histories, scores, and access to credit long after the recovery of productivity, income and repayment propensities can also apply to contexts other than agricultural lending in developing countries. A failure to take exogenous shocks into account could generate inefficiencies in other credit markets. If transitory shocks can be observed, and the forces driving the recovery are well understood and unrelated to lender characteristics, the costs they generate could be mitigated. As far as I am aware, this is the first paper to empirically document this mechanism. Closely related, Avery, Calem, and Canner (2004) argue that situational circumstances that temporarily affect repayment can generate problems for credit

---

4The transitory nature of rainfall shocks is a common supposition in the development literature. See for example Kaur (2015).

5This idea is consistent with the fact that the BAC’s repayment schedule is organized to match the harvests of its farmers. Empirical evidence on this matter is scant but still suggestive of such relation. For example, Chirwa (1997) finds that crop sales of small farmers are associated with higher loan repayment in Malawi. Acquah and Addo (2011) find that higher fishing income is associated with higher repayment of loans to fishers.
scores in consumer lending markets.\textsuperscript{6}

If exogenous shocks generate costs for the lender, why aren’t they taken into account? This fact is more puzzling for agricultural credit markets in developing countries, since shocks are in principle easy to observe. The quote from Hammurabi’s Code even suggests that using this information is a matter of common sense. But below, I argue that technological constraints only recently allowed for shocks to be measured with the level of precision needed to incorporate them in credit scores.

My paper contributes to four different areas of research in development economics and finance. First, it adds to an empirical literature that estimates the impacts of credit scoring and credit histories (Einav, Jenkins, and Levin, 2012, 2013; Giné, Goldberg, and Yang, 2012; de Janvry, McIntosh, and Sadoulet, 2010). This literature has found almost uniformly that credit scoring leads to better credit market outcomes. My paper points to an inefficiency that arises in contexts where repayment is affected by orthogonal and transitory shocks and they are not recorded in credit histories or included in credit scores. These findings suggest that the optimal information set to estimate credit scores is highly context-specific. In this sense, this paper also contributes to work that investigates how to improve credit scores.\textsuperscript{7}

Second, this paper adds to the research that explains why individuals are credit constrained, in particular in the agricultural sector in developing countries (Conning and Udry, 2007; Giné, Goldberg, and Yang, 2012).\textsuperscript{8} This paper shows that farmers can lose access to formal credit simply from the fact that agricultural production generates volatile income streams that lead to loan default. This loss of access can be persistent due to worse credit histories and credit scores.

Third, this paper contributes to a large literature that studies the effects of information sharing through credit bureaus and the use of credit reports (Sharpe, 1990; Pagano and Jappelli, 1993; Vercammen, 1995; Padilla and Pagano, 2000; Jappelli and Pagano, 2002; de Janvry, McIntosh, and Sadoulet, 2010; Giné, Goldberg, and Yang, 2012; González-Uribe and Osorio-Rodríguez, 2014; Dobbie, Goldsmith-Pinkham, Mahoney, and Song, 2016; Garmaise and Natividad, 2016).\textsuperscript{9}

\textsuperscript{6}Using data from the US, the authors study the effect of county-level unemployment rates and changes in marital status on repayment performance. They find that both variables affect loan repayment after controlling for ex-ante credit scores. Nevertheless, they mention that solving the problem that these circumstances create for credit scoring models is difficult, since information on these episodes is not available within credit reporting agencies. Furthermore, the circumstances they consider are not exogenous, in the sense that their occurrence is likely correlated with borrower characteristics. In my case, rainfall shocks are orthogonal to individual characteristics and they are in principle easy to incorporate in credit histories and scores. Furthermore, it is unclear if individuals will recover from the circumstances the authors consider. For example, a divorce can have long term consequences on borrower’s income and repayment.

\textsuperscript{7}See for example Rocha Sousa, Gama, and Brandao, 2016 and Wei, Yildirim, Van den Bulte, and Dellarocas, 2015. To the best of my knowledge, the credit scoring industry has not yet started to incorporate or search for information on exogenous shocks.

\textsuperscript{8}A large agricultural-economics literature has documented the costs of credit constraints for farmers. See for example Carter and Olinto (2003), Petrick (2004), Guirkinger and Boucher (2008) and Fletschner, Guirkinger, and Boucher (2010).

\textsuperscript{9}Of these, Garmaise and Natividad (2016) is the most related to this paper. The authors show that random credit rating downgrades (that is, credit downgrades not related to default) in consumer credit markets in Peru.
My paper contributes to this literature by providing evidence on the costs of credit reports that do not differentiate among reasons for default. The consequences of this effect can be amplified when information is shared between lenders and the reasons for the default are not specified in the credit history.\textsuperscript{10}

Fourth, this paper adds to a literature that shows that transitory shocks can have long term macroeconomic consequences (Blanchard and Summers, 1986; Ouyang, 2009; Eslava, Galindo, Hofstetter, and Izquierdo, 2010; Ball, 2014; Barrot and Sauvagnat, 2016) and also long term effects on farmer production (Rosenzweig and Wolpin, 1993).\textsuperscript{11} After a weather shock, farmers can lose access to credit, potentially hurting productivity enhancing investments. This mechanism could operate in other credit markets besides the agricultural setting. Firms may lose access to credit due to shocks orthogonal to their characteristics and from which they recover faster than their credit histories. This in turn can lead to costs in terms of investment and total factor productivity. As far as I can tell, my paper is the first to suggest that short term real shocks can have long term consequences through use of credit scoring and credit histories.

The rest of this paper is organized as follows. Section 2 provides some background. In Section 3, I study the effect of rainfall shocks on repayment, credit scores, and future access to credit, and I document the persistence of these effects. In Section 4, I study recovery of coffee tree productivity, recovery of income after rainfall shocks, and repayment recovery. Finally, Section 5 concludes.

2 Agriculture in Colombia and the Banco Agrario

2.1 Agricultural Financing

As is frequent in developing countries, the rural sector in Colombia is under-supplied with capital. According to a national census conducted by the Colombian official statistical agency, DANE, in 2013 only 11\% of agricultural producers demanded credit (either formal or informal). Of these, cause a three-year reduction in financing. These downgrades are generated by credit rating thresholds and not by real shocks affecting repayment. The main difference between the two studies is that in my case exogenous shocks do have real consequences but farmers’ repayment recovers faster than the associated negative credit reports.\textsuperscript{10}

For example, a farmer might refrain himself from applying for a production loan if he anticipates to need a loan (say a consumer loan) in a not-so-distant future. This idea is consistent with the finding of Liberman (2016) that borrowers of a department store in Chile are willing to pay 11\% of their monthly income for a good credit reputation.\textsuperscript{11}

In doing so, a side contribution of my paper relates to a voluminous literature that studies the effects of weather on economic activity. Dell, Jones, and Olken (2014) provide a summary of this research. My paper documents a new margin in which weather matters: subsequent access to formal credit. Estimations of the effect of rainfall on repayment are scant. As far as I can tell, the only other work that estimates at the farmer level this effect is the paper by Pelka, Musshoff, and Weber (2015). My study is the first one to estimate the effect of rainfall shocks on repayment using coordinates of individual farms. See Castro and Garcia (2014) for an estimation of the effect of climate variations in a structural default risk model, using aggregate data from the BAC.
89.6% got a loan. The ELCA rural panel,\(^{12}\) representative of four rural regions in the country, shows better results. In 2013, 52% of households had no credit, 25% had one or more formals loans (and no informal loans), 16% had one or more informal loans (and no formal loans) and 7% had a formal and an informal loan (or more) (Cadena and Quintero, 2015). Therefore, at least according to the ELCA, formal loans are the main source of capital for rural households.

Despite these low rates of credit access in Colombia, the BAC is the main player in this market. The BAC is a publicly owned bank, created in 1999, with the mandate from the government to finance agricultural productive activities in the country.\(^{13}\) In 2013 it lent 97% of formal loans to small farmers\(^ {14}\) (DNP and FINAGRO, 2014) and it is the only bank in the country to have presence in all of its 1123 municipalities. 89% of the banks’ branch offices are in rural areas. For small farmers, collateral is guaranteed with resources from the Fondo Agropecuario de Garantías, (FAG) [Agricultural Guarantees Fund], a public fund created for this purpose.\(^ {15}\)

The fact that the government provides the collateral for loans might suggests that the BAC has no incentives to screen borrowers. This is not the case though as indicated by the following three facts. First, the process to recover the guarantee is cumbersome and costly. Second, the regulations that apply to the BAC are the same as those of other banks and the main indicators of the bank’s performance are tied to borrower default. If these indicators turn out to be unfavorable, financial authorities can intervene. Third, and as discussed in more detail below, the BAC has developed its own models to screen borrowers.

\[ \begin{align*}
\text{2.2 The BAC’s Loan Application Process} \\
\text{The typical farmer who applies for a loan from the BAC does so in the branch office closest to his farm. In the office, a loan officer does a consult with CIFIN, the credit bureau in business with the BAC. From this consult, a report is issued. It contains information on the credit history of the borrower, the CIFIN Credit Score (which results from CIFIN’s own credit scoring model), and whether or not the application process can continue. At this point the application process can end with the farmer being denied credit. The immediate denial at the CIFIN stage depends on the CIFIN Credit Score, on other variables in the farmer’s credit history and on the BAC’s}
\end{align*} \]

\(^{12}\)The ELCA panel is a longitudinal survey conducted by the Universidad de los Andes. It carries out surveys for both rural and urban areas. Two waves have been conducted thus far, in 2010 and 2013. For rural areas, around 4500 households where interviewed in 17 municipalities, representative of four rural regions of Colombia.

\(^{13}\)According to the bank’s statutes, 70% of the loan balance has to correspond to agriculture related activities.

\(^{14}\)Government institutions define a small farmer as one with an amount of total assets in Colombian Pesos smaller than 120 millions (around $US 39000, at the average daily exchange rate of the first semester of 2016 equal to 3098 Colombia Pesos per US$). This includes all household assets (for example a car or a television), and not only those used for farm production.

\(^{15}\)The FAG is administered by the Fondo para el Financiamiento del Sector Agropecuario (FINAGRO) [Fund for the Financing of the Agricultural Sector]. FINAGRO is a second tier bank also created with the objective of financing rural productive activities. FINAGRO lends “rediscount” funds to first tier banks. It lends at an interest rate \(i\) and first tier banks lend at a rate \(i + x\) to agricultural producers. In 2013, 85% of FINAGRO’s “rediscount” resources where allocated by the BAC (DNP and FINAGRO, 2014). This is another indicator of the BAC’s importance for financing small farmers in the country.
policies. For example, there is a threshold set by the BAC for the CIFIN score. If the farmer’s score is below this threshold, the application is denied. There are other variables that determine if the application is denied at this stage. An example are periods with overdues during past loan tenures. The BAC’s policies regarding the CIFIN consult are constantly being changed and tuned up. I refer to this stage of the application process as the CIFIN Consult Stage.

Once the loan application makes it through the CIFIN stage, it arrives to the main offices of the BAC in Bogotá, the capital of Colombia. There, each loan application is individually revised by a credit analyst. Different inputs are taken into account by the credit analyst when he reviews the application. An important one is a score obtained from the BAC’s own credit scoring model (which is different from CIFIN’s model). Other examples are the projected income and cost flows of the project the farmer wants to finance. The credit analyst makes the final decision on loan approval. I refer to this stage of the application process as the Analysis Stage.\textsuperscript{16}

\section{The Effect of Rainfall Shocks on Repayment, Credit Scores and Future Access to Credit}

In this section of the paper I study the effects of rainfall shocks on repayment behavior, credit scores and future access to credit and the persistence of this effect. I divide the presentation of the results in two parts. First, I study the effect of rainfall shocks on repayment behavior. Next I study their effect on subsequent access to credit and the persistence of this effect.

\subsection{Effect on Repayment}

\subsubsection{Data and Construction of the Estimation Sample}

The following are the three main data sources that I use to estimate the effects on rainfall shocks on repayment behavior. First, I use administrative records from the BAC. Second, I use data from the Sistema de Información Cafetera (SICA) [System of Coffee Information] from the FNC. The SICA contains yearly information on plot characteristics from coffee farmers in Colombia that interact with the FNC. Finally, I use rainfall data from the Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia, (IDEAM) [Institute of Hydrology, Meteorology and Environmental Studies].

To construct the estimation sample, I first take all the farmers that are ever observed in the SICA data from 2006 to 2014. The SICA is updated every time a farmer interacts with the FNC. This interaction is frequent since coffee farmers receive social services and technical support from this institution. Also, many smaller cooperatives of coffee growers are affiliated to the FNC, and

\textsuperscript{16}The application process has seen a series of changes in time. For example, the BAC’s model used to score small farmers entered in operation in 2012.
through them, farmers interact with the FNC (Muñoz-Mora, 2016).

For the list of farmers that ever appeared in the SICA data in 2006-2014, I obtain the loans that were disbursed by the BAC in the period of 2005-2011. For this set of loans, I select those of farmers with at least one farm observed in the SICA in the year of loan disbursement. The largest farm is then used to link the loan to the rainfall station. The farms in the SICA and in the IDEAM rainfall station are geo-referenced. For each farmer, I use the Euclidean distance to find the closest rainfall station to the largest farm at the time of loan disbursement. In panel A of Figure 1, I depict the distribution of coffee farms in the SICA across Colombia, for the years 2010-2013. Panel B shows the distribution of the rainfall station that are close to at least one farm in the SICA. Figure 1 shows that the distribution of the IDEAM rainfall stations is dense in coffee growing areas of Colombia. Therefore, the closest rainfall station provides a good measure of the extreme weather events that the farmers faces. Finally, from the set of loans linked to a rainfall station, I take only loans whose destination is related to coffee production. I consider all maturities of coffee loans.

Since the focus of the paper is on the effect of rainfall shocks on loan applications that follow an initial loan tenure, I construct a sample that allows me to look at this effect. That is, I select a sample of farmers with an initial loan who then applied for a subsequent loan. I refer to this sample as the sample of Loans With Posterior Applications. Note that this sample is different from the sample of all disbursed loans. To construct it, I start with the most recent loan originated in 2008-2011. This results in a sample of one loan per farmer. For each farmer, I then find the first application after the maturity of the initial loan in both the CIFIN Consult data and the Analysis Stage data.

---

17 As indicated below, the mean distance between the farm and the closest rainfall station in my main estimation sample is 6.5 kilometers.
18 The BAC data has detailed descriptions of the destination of the loans. The following are two examples:
1) “Traditional coffee - sustainment of agricultural production”: refers to loans for sustaining traditional coffee production; which include for example fertilizer purchases. ‘Traditional’, as opposed to ‘technical’ refers to coffee production with parameters such as type of coffee seed that are not as good as the ones used in ‘technical’ production.
2) “Coffee renovation by plantation - plantation and maintenance”: broadly refers to loans destined to the plantation and maintenance of new coffee trees.
19 There are close to 822,000 coffee farmers ever observed in the SICA data in 2006-2014. For them there are a total of 498,000 loans in the BAC data in the period 2005-2011. Around 299,000 (60%) correspond to coffee production and close to 274,000 can be linked to a rainfall station at the quarter of loan disbursement. By selecting the sample in this manner, I ensure that the farm used to link the rainfall station to the loan is observed in the SICA data in the same year of loan disbursement and is associated to the same farmer.
20 The reason to use as starting date 2008 is that in that year the Law 1266/ 2008 was introduced. This law required credit bureaus to erase negative information (past overdue periods for example) that was sufficiently old by 2008. For details on the Law see González-Uribe and Osorio (2016). To avoid any confounding effects of the law I start with loan that where disbursed in 2008 or afterwards. Also, I study the effect on the CIFIN Consult Stage of an application following an initial loan. Data of the CIFIN consults is only available starting in 2010. This restriction on the sample ensures that the maturity of the initial loan is close enough to 2010, so I can find in the CIFIN data the subsequent loan application.
21 For reasons explained bellow (see in particular footnote 27) I work with quarterly dates. For a few cases, there are multiple loans for the same farmer originated in the same quarter. When this is the case, I randomly select one of the loans.
22 Note that for a given initial loan, the first application in the CIFIN consult does not need to correspond to the
3.1.2 Definition of Rainfall Shocks

I now describe the definition of rainfall shocks for which I use historical data from the IDEAM. I use monthly data on precipitation available from 1982 to 2012. I construct rainfall shocks at the rainfall station level. For each station I add observed monthly precipitations to obtain a quarterly measure of precipitation. Then, I define shocks at the calendar-quarter level in the following manner: for each quarter I obtain the 80th percentile of the observed rainfall distribution from 1982 to 2012. I say that in a given quarter there was an excessive rainfall quarter-shock if observed rainfall was above the 80th percentile of the corresponding distribution. Note that under this definition, for each calendar year I can observe up to four quarter-shocks in each station. For a given calendar year, I will say that a rainfall shock occurred (and refer to it as Rainfall Shock) if at least two of these quarter-shocks occurred. This definition takes into account seasonality both at the station and at the calendar-quarter level. According to this definition, a shock occurs if in a given year, rainfall is particularly high compared to the historical raininess of the rainfall station during that year.

3.1.3 Empirical Approach

I estimate by OLS the following linear-probability model of repayment of the initial loan in the sample of Loans With Posterior Applications:

\[ y_{ijm\tau} = \beta s_{j\tau} + \mu_{m\tau} + \delta_j + \epsilon_{ijm\tau} \]  

where \( y_{ijm\tau} \) is a dummy equal to 1 if loan \( i \), close to rainfall station \( j \), of maturity \( m \) and originated in quarter \( \tau \), was ever overdue by 30 days or more. \( s_{j\tau} \) is a dummy equal to 1 if a rainfall shock occurred in the first year after loan disbursement. My results are robust to considering different definitions of shocks. For example if I consider the number of quarter-shocks in a given calendar year, I find similar effects on repayment and subsequent access to credit.

I focus on excessive rainfall shocks since coffee growing in Colombia is more sensible to periods of excessive rain according to conversations with personnel from the FNC. In estimations not reported in this paper I find no effect of shocks of low rainfall. See also Gast, Benavides, Sanz, Herrera, Ramírez, Cristancho, and Marín (2013), Boucher and Moya (2015), and Turbay, Nates, Jaramillo, Vélez, and Ocampo (2014).

In Appendix A, I present robustness exercises where I estimate the effect in a cross-section of loans disbursed in 2005-2011. In that sample, I do not select a single loan for each farmer and instead consider all loans disbursed. The effects are very similar to the ones estimated in the main text.

Since I define rainfall shocks using calendar quarters and also for simplicity, I handle other dates at the quarter
the first year of loan maturity. The reason is that my sample consists of a continuum of maturities starting with maturities of less than a year. \( \mu_{m\tau} \) are quarter-of-disbursement times maturity fixed effects. \( \delta_j \) denote rainfall station fixed effects and \( \epsilon_{ij\tau} \) is a mean-zero error term. The coefficient of interest is \( \beta \) and it is expected to be greater than 0, indicating that rainfall shocks increase the probability of entering in a period of 30 days with overdues.

The inclusion of \( \mu_{m\tau} \) and \( \delta_j \) implies that the coefficient \( \beta \) is identified from variation in \( s_{j\tau} \) across time within rainfall station, for loans of similar maturity. Therefore the coefficient \( \beta \) is identified and has a causal interpretation as long as the occurrence of shocks for a given rainfall station is not systematically correlated with other time-varying factors that affect repayment of outstanding loans linked to the station. Formally, I assume that:

\[
E[\epsilon_{ijm\tau}|s_{j\tau}, \mu_{m\tau}, \delta_j] = 0
\]

This assumption is reasonable given my definition of rainfall shocks. \( s_{ij\tau} \) captures atypical rainfall realizations for a given year in a given station. This variation is likely to be uncorrelated with other time varying factors that affect repayment of loans linked to the rainfall station.

### 3.1.4 Results

Table 1 shows summary statistics of loan characteristics in my estimation sample. The average loan has a maturity close to two years and an annual interest rate of 11%. The mean distance between the farm and the closest rainfall station is 6.54 kilometers. The mean value of \( s_{j\tau} \) is 0.42 and the 75th percentile is 1. This captures the fact that the period I consider was a particularly rainy one because of the occurrence of the “La Niña” climatic phenomenon in the second half of 2010 and in 2011. On average, 14% of loans entered in a period of 30 days past due and 10% level too. For example, if a loan was disbursed in March 2007, I say that the loan was disbursed in the first quarter of 2007 (denoted by 2007-1). For that particular loan, \( s_{j\tau} \) will take a value of 1 if at least two of the quarters in \{2007-1, 2007-2, 2007-3, 2007-4\} where quarter-shocks.

Although I have a continuum of maturities, its distribution is bimodal as shown in Figure 2. The first mode corresponds to loans between one or two year maturities which are usually related to the sustainment of agricultural production (e.g. buy fertilizer). The second mode corresponds to loans of five years or more which correspond to long term investments like planting new coffee trees or investing in production infrastructure. Therefore I define my set of maturity fixed effects as consisting of a single dummy equal to one for loans of maturities of three years or more.

My main outcome of interest is whether or not the loan entered in a period of 30 days past due. The reason is that this is the main indicator of default used by the BAC. According to its statistics, 80% of loans that enter in a 30 days overdue period end up in longer periods with overdues. The 30 days overdue period is also the main indicator that financial regulations in Colombia require banks to use. According to BAC officers, this is the main indicator of default used in other countries as well. Furthermore, in the BAC data there is not a variable for default, since this is a concept that can be defined in different ways.

La Niña [The Girl] is an ocean-atmospheric phenomenon. It is generated by lower than average temperatures of the Pacific Ocean, in its Eastern-Central part. La Niña is associated with higher levels of rain and cloudiness in some countries of South America like Colombia, Ecuador and Peru (Gast, Benavides, Sanz, Herrera, Ramírez, Cristancho, and Marín, 2013). Its counterpart, El Niño [The Boy] phenomenon, is associated with lower temperatures than average and has the opposite effects on rainfall than La Niña.
entered in a period of 60 days past due.

Table 2 presents the results from the estimation of equation 1 (in the sample of Loans With Posterior Applications), and different variants of it. Each column corresponds to a different regression. All the estimations include rainfall station fixed effects. The set of quarter-of-disbursement times maturity fixed effects varies with the specification as indicated in the bottom of the table. I cluster the error terms at the rainfall station level. The first three columns of the table correspond to the baseline specification. Column (1) reports the coefficient of estimating exactly equation 1. It implies that a rainfall shock during the first year of loan tenure increases the probability of entering into a period with overdues in 22% compared to loans with no shock. In Column (2), I report results of an alternative definition of shocks where I use as dependent variable the number of quarter-shocks in the first year after loan disbursement. As explained before, this variable can take a value of 1, 2, 3 or 4. The coefficient implies that each additional quarter-shock in the first year increases the probability that the loan enters into a period of 30 days past due in 11.3%, compared to mean of the outcome in the entire sample.

The next four columns present robustness exercises of the previous results. Column (3) shows that a rainfall shock increases the probability of entering into a period of 60 days past due in 19%. This implies that the effect documented before is also present for longer overdue periods, consistent with the fact that most loans that enter into a period of 30 also enter longer overdue periods (see footnote 29). Column (4) shows the results for a subsample where the distance between the farm and the rainfall station is lower than the mean of the full sample (equal to 6.54 km). The estimated coefficient is again positive and significant at the 5% level. Column (5) shows results for the subsample of short terms loans. These are loans with maturities of up to three years. The coefficient is almost identical to that of the baseline scenario. Finally, column (5) reports results for long term loans. I define long term loans as loans of maturities of three years or more. Again, a rainfall shock in the first year of loan tenure increases the probability that the loan enters into a period of 30 day or more past due, relative to long term loans with no shocks in the initial year. The economic magnitude of these coefficients is sizable. For the baseline case (Column (1)) it corresponds to a 22% increase relative to loans with no shocks. This is saying that because of exogenous reasons outside of the control of the farmer, default increases by 22%.

In sum, the results presented in Table 1, imply a robust effect of rainfall shocks on repayment. Loans with shocks are more likely to enter into long past due periods. Now, I turn to the first piece of evidence of the effects of shocks on credit scores. In particular, I look at the effect of shocks on the scores that the BAC reports to credit bureaus and financial authorities.

### 3.2 Effect on Reported Scores

Following regulations from financial authorities, the BAC has to report each month to credit bureaus a score for each client with an outstanding loan with the BAC (hereafter *BAC Reported*...
This score ranges from A (highest) to E (lowest). Every time a loan is generated, the loan is assigned a score of A. By rule from the financial authorities, a single score has to be reported for each client. When different loans have different scores for the same client, the lowest score is assigned to the client and is the one reported. For small farmers, the BAC reported score depends mainly on days past due and the BAC’s policies (some predetermined thresholds for example).

Table 3 presents results of the effect of rainfall shocks on the BAC Reported Score. Here, the sample is identical to the one I use in the baseline exercises of Table 2 and the estimated equation is identical to equation 1 but I consider two different outcomes. The first one is a dummy for whether or not the BAC Reported Score fell (to any score different than A) at any point during loan tenure. The second one is a dummy for whether or not the score fell to the worst possible score, E. The results imply that rainfall shocks lead to a worse BAC reported score. In particular, a rainfall shock during the first year of loan tenure causes the probability of the BAC reported score to fall or to fall to E in both cases by 20%, compared to loans with no rainfall shock. Therefore, rainfall shocks generate negative information for borrowers that is then reported to credit bureaus. Again, the size of these coefficients is important.

The reports from the BAC are one of the channels through which the farmer’s credit history feeds into credit bureaus and the financial system in general. A negative credit history will have effects on the credit bureau’s credit scores and future access to credit as I document in the next section.

3.3 Effect on Credit Bureau Scores and Subsequent Access to Credit

In this section, I study the effect of rainfall shocks in credit bureaus’ credit scores and subsequent access to credit. The sample used is the same as the one used in the previous section, that is the sample of Loans With Subsequent Applications.

3.3.1 Data and Empirical Approach

I have data for both the CIFIN Consult Stage and the Analysis Stage. Regarding the CIFIN Consult Stage, data is only available for the period 2010-2015. For this period, I observe all the CIFIN consults corresponding to all loan applications received by the BAC. For the Analysis Stage, data is available from 2005 to 2015 for all loan applications that made it to this stage (and therefore, passed the CIFIN Consult Stage).

To analyze the effect of rainfall shocks on subsequent access to credit, I compare loan applications following loan tenures where a rainfall shock occurred during the first year with applications

---

31 In the financial literature terminology, the information a lender shares or reports to a credit bureau is divided in “negative” and “positive” information. Negative information consists mainly of defaults, while positive information includes aspects like outstanding debt. In the case of the reports to the BAC (and in general of all banks in Colombia) both negative and positive information is reported.
after loan tenures with no shock. More precisely, I estimate by OLS the following model:

\[ x_{ijmτ} = αs_{jτ} + θ_{mτ} + γ_j + u_{ijmτ} \] (3)

where \( x_{ijmτ} \) is one of three outcomes for loan application \( i \), after the maturity of a loan of maturity \( m \), originated in quarter \( τ \), close to rainfall station \( j \). The three outcomes I consider are the CIFIN Score at the CIFIN Consult Stage, a dummy for denial at this stage and a dummy for denial at the Analysis Stage. As before, \( s_{jτ} \) is a dummy equal to 1 if a rainfall shocks occurred in the first year of loan tenure (of the initial loan). \( θ_{mτ} \) are quarter-of-disbursement times maturity fixed effects of the initial loan. \( γ_j \) are rainfall station fixed effects and \( u_{ijmτ} \) is a mean-zero error term.

Similar to the case where I studied the effect of rainfall shocks on repayment, the identifying assumption relies on \( s_{jτ} \) being uncorrelated with the error term. Again, this assumption seems reasonable given my definition of rainfall shocks.

I estimate equation 3 for two different samples. Recall that in the construction of the sample of Loans With a Subsequent Application I start with an initial loan and then find the most recent loan application. The selection of this initial loan is irrespective of loan maturity. The two samples I consider differ on the conditions imposed on this initial loan. The first one starts with loans irrespective of maturity, and is the one I have used thus far. The second one starts with initial loans of one year maturity. The reason to consider this second sample is that I can subset it in a way that keeps constant the amount of time elapsed between the rainfall shocks and the next application. I will provide below more details to justify the need of considering this second sample, where I argue that repayment recovers faster than credit access.

3.3.2 Results

Table 4 presents the results of the effect of rainfall shocks on subsequent access to credit. Panel A shows the results for the sample of applications that follow initial loans of all maturities while panel B shows the results for the sample that starts with one year maturity loans.

In column (1), I report results of estimating the effect of a rainfall shock during loan tenure on the probability of applying for a new loan. That is, I estimate equation 3 where the outcome is a dummy for whether or not the farmer applied for a loan after maturity of the original loan. In both panels A and B, I find no significant effects of rainfall shocks on the probability of applying for a new loan. In other words, there is no selection in application caused by rainfall shocks. For exposition purposes, Column (2) presents the effect of rainfall shocks on the first loan. In the case of Panel A, this coefficient is that same as the one reported in Column (1) of Table 2.

Now I turn to the results of the effect of shocks during the initial loan tenure on the next loan application. For the sample in panel A (the one with initial loans of all maturities), column (3)
shows that the occurrence of a rainfall shock in the first year of loan tenure decreases the CIFIN score by 5.7 points. This corresponds to a decrease of around 0.05 standard deviations of the CIFIN score in the sample. The coefficient in column (4) is positive and statistically significant, indicating that a rainfall shock causes an increase in the probability of denial at the CIFIN stage for the next loan application. The magnitude is large. The size of the coefficient implies that this increase in the probability equals 12.6% of the average frequency of denial of applications following loans with no rainfall shock. Regarding the probability of denial in the Analysis Stage, the effect is also important. Applications that arrive to this stage are 9.8% more likely to be denied if a rainfall shock occurred compared to the denial rate of applications with no shocks during the previous tenure.

Regarding applications that followed loans of one year maturity, the results are very similar (and somewhat larger) to the previous ones, as can be seen in Panel B of Table 4. In particular, a rainfall shock during the first year of loan tenure of the initial loan leads to a decrease in the CIFIN score of 7.5 points and an increase in the probability of denial of 17.4% and 11.8% in the CIFIN Consult and the Analysis Stage respectively, compared to the average denial rate of applications following loans with no rainfall shocks.

The previous results show that rainfall shocks have a large effect on credit bureaus’ scores and on subsequent access to credit. On average, the occurrence of rainfall shocks during loan tenure is associated with lower credit scores, and more frequent denial of subsequent loan applications. In the following section, I show that this result remains as the time window between the initial loan maturity and the next application grows, which indicates that the effect persists over time.

3.3.3 Persistence

In this section, I document the persistence of the effect of rainfall shocks on credit bureaus’ scores and subsequent access to credit. It is important to note at this point that in 2008 a law was introduced in Colombia that requires that negative information (past defaults) cannot remain on credit bureau registries for more than four years. Therefore, information older than this cannot be reported back to banks or used in the estimation of the credit bureaus’ scoring models. This puts a time limit to the effect of negative information on future access to credit.32

To study the persistence of the effect, I consider a subset of the samples of the previous section. More precisely, I take the set of first applications after maturity of the original loan but add an additional requirement: that at least a given amount of time elapsed between maturity of the original loan and the first application. I consider a one and a two year time window. For these samples I estimate the same regressions of the previous section.

Table 5 presents the results. Columns (1) to (3) report results for the sample that starts with loans irrespective of maturity. Columns (4) to (6) correspond to the sample that starts with

32 This is the same law discussed previously in Footnote 20.
loans of one year. Panel A presents the results for a time window of one year. Similar to the results presented in Table 4, for both samples, a rainfall shock during the first year of loan tenure decreases the CIFIN score, and increases the probability of denial at both the CIFIN Consult and Analysis Stages. In all cases the coefficients are statistically significant and their size are very close to the ones reported in Table 4. These results imply that even after a year has elapsed between the time of the shock and the loan application, loans are still more likely to be denied. In Section 4, I will argue that this is enough time for the income stream of the farmer to recover and that under some scenarios it is also enough for repayment to recover. Therefore the credit scoring system is leading to denial of loans that on average would be profitable to the lender.

Panel B of Table 5 shows the results for a time window of 2 years. The sign of the coefficients is the same as in the case of a time window of one year. For both samples of maturities, a rainfall shock during loan tenure decreases the CIFIN score. Applications after one year maturity loans are more likely to be denied at the CIFIN Consult Stage if a rainfall shock occurred. The rest of the coefficients (columns (2), (3) and (11)) are not significant at conventional statistical levels. Despite this fact, the size of the coefficients is similar, and in some cases larger, than the coefficients in Panel A. Given that the sample size drops by almost half in the regressions of panel B relative to those in Panel A, this is not surprising. In sum, I conclude that even after two years have passed since the maturity of the original loan, subsequent applications are still more likely to be denied. Again, this time window is longer than the one required for the income stream of the farmer to recover from a shock and is long enough, in some plausible scenarios, for repayment to recover.

A small caveat needs to be mentioned at this point. Rainfall shocks might affect the timing of the next loan application. So the comparison between farmers with the same time window, some of which had a shock during the initial loan tenure and some of which did not, might not be the ideal one. Despite this, the fact that the sign and size of the coefficients remains practically unchanged for different time windows provides reassurance in that the effect of shocks is persistent.

The effect of rainfall shocks on credit scores reported by the BAC, credit bureaus’ scores, and future access to credit implies a cost to the farmer who is credit constrained. This cost is larger, the longer the time his credit history keeps record of his defaults. In the next section, I show that the farmer’s income recovers faster than his access to credit. This implies that the lender is also incurring a cost, as long as future income is the main determinant of future loan repayment. I also present evidence on repayment recovery and argue that under some scenarios, repayment recovers faster than access to credit. The lender could have lent to farmers who experienced a shock, since on average their income stream recovers fast enough to allow for repayment of the next loan.
4 Recovery

In this section I argue that the income stream of the farmer recovers faster from rainfall shocks than the credit history and present empirical evidence on repayment recovery. To show that income recovers I proceed in two steps. First, I draw from a large agronomic literature to argue that rainfall shocks affect the productivity of the coffee tree if they occur at most one year before harvest. Shocks occurring before that have no effect. Second, I use data from a representative survey of coffee farmers in Colombia conducted in 2006 to show that the income stream recovers from rainfall shocks faster than credit access, consistent with the agronomic literature. Regarding repayment, I show that it recovers in a sample of long term loans and in a sample of high credit score farmers. I discuss how the results on repayment on long term loans together with the results on persistence of the effect on credit access imply that under plausible scenarios, the bank is not lending to farmers whose repayment recovers in time to repay the next loan.

4.1 The Productivity of the Coffee Tree and Rainfall Shocks

Coffee grows in trees and the seed of the cherry is the coffee bean. Usually, farmers harvest the cherry when it is ripe, remove the pulp and dry it to obtain “parchment” coffee. Then they sell it to cooperatives, the FNC or other buyers. The price paid to farmers basically depends on the international price (after conversion to Colombian Pesos at the ongoing exchange rate), plus a premium that recognizes the quality of Colombian coffee. For most coffee farmers, the price is fairly similar across the country.\(^{33}\)

On average, the coffee tree starts producing coffee one year after its plantation. From year one to five, the productivity of the coffee tree is increasing. Then, from years 5 to 8 it is decreasing. After that age, the coffee tree can still produce coffee but renovation is highly recommended. Different factors affect this pattern including, for example, the quality of the soil (Arcila, Farfán, Moreno, Salazar, and Hincapié, 2007). When the productivity of the tree is decreasing, farmer frequently cut the tree’s branches to increase its productivity in subsequent harvests.\(^{34}\)

During the life time of the coffee tree, and depending on the zone of the country, there are one or two harvest each year.\(^{35}\) There are two main phases that determine that amount and quality of the coffee cherries of the harvest. The first one is the flowering stage, when coffee flowers blossom and then transform into the coffee cherries. The second one is the coffee fruit development stage, when the coffee berry grows. The flowering phase lasts between 3 to 5 months and the fruit development stage lasts between six and seven months (Gast, Benavides, Sanz, Herrera, Ramírez, Cristancho, and Marín, 2013). Therefore the critical period that determines the quality of the coffee cherries is the flowering stage and the coffee fruit development stage.

\(^{33}\)Some high quality coffees are paid higher prices.

\(^{34}\)This is known as renovation by “zoca” or “zoqueo” in Colombian coffee jargon.

\(^{35}\)The coffee tree is permanently producing coffee. But depending on the region, there are some times in the year where the level of production is concentrated. These are the harvest periods.
harvest starts 12 to 9 months before.

For a given harvest, the productivity of the coffee tree is highly sensible to the weather events during this period. The agronomic literature has documented that high levels of rain during the flowering phase hinder the development of the coffee flower and are associated with lower levels of cherry production of the coffee tree. Furthermore, lack of solar radiation during the fruit development phase is also associated with lower production (Gast, Benavides, Sanz, Herrera, Ramírez, Cristancho, and Marín, 2013; Arcila, Farfán, Moreno, Salazar, and Hincapié, 2007; Boucher and Moya, 2015). Therefore, periods of excessive rainfall affect production directly during the flowering phase, and indirectly during the fruit development phase since clouds diminish the amount of sun light the tree receives.

In sum, periods of excessive rainfall affects the productivity of the coffee tree if they occur up to one year before the harvest. The next harvest will not be affected by these events since its flowering and fruit development phases will be under no distress after rainfall has returned to normal.

Although the effect of rainfall shocks on the productivity of the coffee tree is transitory and can last up to one year, the duration of the effect on income can be different. The income stream generated by coffee growing is an economic quantity that depends on farmer’s decisions and not only on the productivity of the coffee tree. In the next section, I show empirically that income recovers from excessive rainfall shocks relatively quick.

4.2 Income Recovery

4.2.1 Data and Empirical Strategy

To study income recovery I use a representative survey of small coffee farmers in Colombia, conducted by the FNC in 2006 and named “[encuesta de] Mercado Laboral Cafetero y Acceso al Crédito para Productores de Café en Colombia” (MLYCC) [Analysis of the labor market and access to credit for small Colombian coffee growers]. The survey contains information on coffee production from coffee farmers with less than 5 hectares cultivated with coffee. Crucially it contains the following question: “How many ‘arrobas’ of parchment coffee did you sell the previous year.”

Coffee farmers in Colombia usually sell all the coffee they produce, so the answer to this

---

36 Although not as important as the reasons mentioned in the main text, excessive rainfall can also lead to excessive soil humidity which can affect coffee production, wash away fertilizers, and is associated to higher incidence of diseases in the coffee tree.

37 Although I have no measures of solar radiation to confirm its correlation with rainfall levels, it has been documented that periods of high rainfall come in hand with lower sunshine in many coffee growing regions of Colombia (Gast, Benavides, Sanz, Herrera, Ramírez, Cristancho, and Marín, 2013; Turbay, Nates, Jaramillo, Vélez, and Ocampo, 2014).

38 This recovery argument excludes catastrophic events. While the coffee tree is very resistant to weather events there are some that can destroy it; a landslide for example.

39 The ‘Arroba’ is a traditional Spanish weight measure. The exact weight of an “arroba” varies across countries and regions. In Colombia it is equivalent to 12.5 kilos, or 27.6 pounds.
question provides a measure both of production and income.

The survey was conducted between March and April of 2006 and asks about the decisions of the farmer in the previous year (Muñoz-Mora, 2016; Lozano, 2009). Although the question is not specific, I assume from the way it is phrased that it refers to coffee sales in 2005. Depending on the region of Colombia, there are between one and two harvests in a given calendar year. Therefore, I expect shocks occurring in 2004 and 2005 to potentially affect the coffee sales of 2005. Figure 3 depicts a time-line where I further clarify this. From the previous section, I expect shocks to have an effect if they occur up to one year before the harvest. For example, and as shown in the figure, harvests in early 2005 are affected by weather events of early 2004. But shocks occurring before January 2004 should not affect 2005’s harvest.

I then proceed to link the farms of the survey to the SICA data, allowing me to obtain its geographical coordinates and the closest rainfall station. With these data in hand I estimate by OLS the following regression:

\[ r_{t,ijk} = \alpha_0 + \alpha_1 s_{t-1,j} + \alpha_2 s_{t-2,j} + \alpha_3 s_{t-3,j} + Z_i'\rho + \phi_k + u_{ijk} \]  

where \( r_{t,ijk} \) is the amount of coffee sold in 2005 per-hectare cultivated with coffee, by farmer \( i \), close to rainfall station \( j \), in coffee growing region \( k \). \( s_{t-1,j} \) is a dummy that takes a value of 1 if a rainfall shock occurred in 2005 or in 2004. Similarly, \( s_{t-2,j} \) and \( s_{t-3,j} \) are dummies for the occurrence of a rainfall shock in 2003-2002 and 2001-2000, respectively. \( Z_i \) is a vector of controls that varies depending on the specification. \( u_{ijk} \) is a mean zero error term. \( \phi_k \) are coffee-region fixed effects. I expect \( \alpha_1 \) to be negative since excessive rainfall shocks occurring in 2005 and 2004 affect the productivity of the coffee tree for the 2005 harvest. Consistent with the hypothesis of recovery from the shocks, I expect the \( \alpha_2 \) and \( \alpha_3 \) to be non-negative. When estimating equation 4, I cluster errors at the rainfall station level and use sampling weights provided in the survey’s data.

The identification assumption here is that \( s_{t-1,j}, s_{t-2,j} \) and \( s_{t-3,j} \) are uncorrelated with \( u_{ijk} \), or in other words, that they are as good as randomly assigned. This assumption might seem strong at first. It might be possible that farmers close to rainfall stations with shocks are different in unobservable characteristics from farmers close to rainfall stations with no shocks. I cannot

\[ \text{See Footnote 35. Again, the coffee tree produces coffee beans permanently. In some regions of the country, production concentrates in two different times of the year leading to two different harvests.} \]

\[ \text{To define rainfall shocks in this section I use the 90th percentile of the historical rainfall distribution instead of the 80th percentile. As discussed before, the period under analysis in the previous sections includes La Niña phenomenon, so shocks defined with the 80th percentile are strong enough to alter repayment. In the current exercise, I need to consider the 90th percentile to capture an effect on coffee production. Arguably, this is due to the fact that no extreme weather events like la Niña occurred in the period under consideration.} \]

\[ \text{I follow Muñoz-Mora (2016) and define four coffee regions that differ in natural conditions, like altitude, that affect coffee production.} \]

\[ \text{The ideal data to identify the recovery from rainfall shocks would be a long enough panel at the farm level with data on production or sales. Unfortunately such a panel does not exist for coffee production in Colombia.} \]
control for this possibility, since the variation I use here is at the rainfall station level, which excludes rainfall station fixed effects. Nevertheless, I am using variation across space in the timing of the shocks (different rainfall stations where shocked at different points in time). Again, given my definition of shocks, this variation is likely to be as good as random even in the cross section.

Figure 4 shows the distribution across Colombia of rainfall stations for which at least two quarter-shocks occurred in 2005 or 2004 (that is $s_{t-1,j} = 1$). Dots in green depict rainfall stations with $s_{t-1,j} = 0$ and that where close to at least one farm in the MLYCC survey (where close is at a distance of 6.1 km or less). Dots in purple depict rainfall stations with $s_{t-1,j} = 1$. The takeaway from the figure is that rainfall stations with $s_{t-1,j} = 1$ are distributed across the country and not concentrated in a particular region. This is consistent with the idea that shocks are as good as random in the cross section, and affect farmers in different regions of the country. In the next section, where I present the results, I show additional exercises that provide support for this claim.\footnote{The estimations presented here use farms at a distance of 6.15 km or less to the closest rainfall station. My estimations for the effects on repayment are irrespective of distance. Nevertheless, Boucher and Moya (2015) only find effects of rainfall on forecasts of coffee tree productivity for distances smaller than to 3 km, although they do not consider rainfall shocks but rainfall levels instead. To be conservative, I use a 6.15 km distance (which is the median distance in my sample) in the baseline specification. The results presented here are robust to considering different distances and total sales (instead of total sales per-hectare) as outcome.}

4.2.2 Results

Table 6 reports the results from the estimation of equation 4 under different specifications. The first column shows the results without the inclusion of controls. The estimated value of $\alpha_1$ is presented in row “Rainfall Shock 04-05”. It is negative and statistically significant. Its size implies that the occurrence of rainfall shock in 2004-2005 is associated with a reduction in the amount of coffee sold (per-hectare cultivated with coffee) equivalent to 32% of the sample mean or to 0.31 standard deviations. The median farm in the estimation sample has 67% of its area cultivated with coffee so coffee sales are likely to be the main income source of most farmers in the survey. Therefore, the reduction in sales caused by excessive rainfall shocks have an important impact on farmers’ income. Regarding excessive rainfall shocks in the periods of 2002-2003 and 2002-2001 the estimated values of $\alpha_2$ and $\alpha_3$ are positive and non-significant at conventional statistical levels. At the bottom of the table, I report the p-value of tests where the null hypothesis is $\alpha_1 = \alpha_2$ and $\alpha_1 = \alpha_3$. I reject equality for the first test at a level of significance of 10%. For the second test, I cannot reject equality at conventional levels of confidence. Nevertheless, the p-value is low (equal to 0.12).

These results are consistent with recovery of farmer’s income: excessive rainfall shocks during 2005-2004 significantly reduce the total amount (or weight to be more precise) of coffee sold, while shocks during the periods of 2003-2002 and 2001-2000 have no negative effect.
the timing of this effect matches with what is expected given the timing of the effect of rainfall shocks on the coffee tree’s productivity. Recall that only shocks occurring up to one year before the harvest affect the productivity. So, as I discussed before, shocks up to 2004 should have an effect on 2005’s harvest. The estimated effects come from different rainfall stations distributed across the country. For example, there were a total of 128 farms that had a rainfall shock in 2004-2005, corresponding to 26 different rainfall stations out of 246 that are close to farmers in the survey.

The second column of Table 6 includes a series of controls that are arguably predetermined at the start of the period under consideration (i.e. 2000-2005): the size of the farm, the size of the household and the level of education of the household head. The inclusion of this set of controls does not considerably change the estimate of $\alpha_1$. This is reassuring in the sense that if rainfall shocks are as good as randomly assigned, the estimated coefficient should not be affected by the inclusion of predetermined controls. The same is true for the point estimates of $\alpha_2$, and $\alpha_3$, which do not change considerably with the inclusion of these controls.

Finally, in the last column I present the results of estimating equation 4 but with controls that are not predetermined at the time of rainfall shocks in the periods 2000-2001, and 2002-2003. In particular, I include the value reported in the survey of the density of the coffee crop (number of trees per hectare cultivated with coffee), the average age of the coffee crop, dummies for the variety of seed used, and a dummy for whether or not the crop is completely exposed to sunlight.45 All these variables have been documented to affect the productivity of coffee crops (Muñoz-Mora, 2010; Gast, Benavides, Sanz, Herrera, Ramírez, Cristancho, and Marín, 2013). Since these variables correspond to coffee farmer’s decisions, they are likely to depend on rainfall shocks during the periods of 2000-2001 and 2002-2003 and therefore might bias the coefficients of interest (Angrist and Pischke, 2008). The results reported under column (3) show that even with the inclusion of “bad controls” the estimate of $\alpha_1$ does not change much. This is not true though for the values of $\alpha_2$ and $\alpha_3$ which change considerably compared to the values under columns (1) and (2), consistent with the idea that the additional controls depend on these variables. Still, the estimated coefficient is positive and not significant suggesting that these shocks do not affect the total amount of coffee sold per-hectare in 2005.

The fact that the estimate of $\alpha_1$ does not vary much across specifications is consistent with the idea that rainfall shocks are as good as random in the cross-section. As discussed previously, Figure 4 shows that treaded rainfall stations are distributed across the country and not concentrated in a particular region. In Table 7, I present a balance test where treated farms are those with a rainfall shock in the period 2004-2005. Large differences in observables across treated and control groups are not observed. The results of Figure 4 and Table 7 are consistent with the idea that rainfall shocks are as good as randomly assigned in the cross section.

---

45It is frequent for coffee crops to be intertwined with other trees that shadow coffee trees.
To conclude, the agronomic coffee literature has documented that excessive rainfall shocks affect the productivity of the coffee if they occur within one year of the harvest. Once weather returns to normal, so does productivity. The results from the exercises with the survey data show that income follows a pattern that is consistent with the evolution of productivity. It is affected by shocks close to the current harvest but not by older shocks.

4.3 Repayment Recovery

A question that arises at this point is whether income recovery translates into a recovery in repayment. This is likely to be the case since one of the main determinants (if not the main determinant) of repayment is the income stream of the farmer.\footnote{Empirical evidence on this matter is scant but still suggestive of such relation. For example, Chirwa (1997) finds in a cross section of loans to small farmers that crop sales are associated with higher loan repayment in Malawi. Acquah and Addo (2011) find that higher fishing income is associated with higher repayment of loans to fishers.} Consistent with this idea is the fact that the BAC’s repayment schedule is organized to match the harvests of its farmers. Nevertheless, it is still possible that after income recovers from a rainfall shock, the farmer does not repay his loan.

I face a fundamental problem in answering the question of whether or not a farmer would repay a loan after defaulting in a previous one, as a result of a rainfall shock. As documented previously, rainfall shocks cause on average higher rates of denial of posterior loan applications so the sample of farmers who get their next loan approved is selected. Furthermore, I do not observe repayment of farmer’s who are denied a loan. To be more precise, this is a quantity that does not exist. So, without making strong assumptions, answering this question in the sample I have used thus far is hard.\footnote{One could postulate a Tobit selection model and estimate a first stage of loan acceptance. Nevertheless, this approach needs strong distributional assumptions. Angrist and Pischke (2008) warn against estimating Tobit models where the latent variable has no empirical counterpart.} Nonetheless, in this section I show that repayment recovers for two different subsets of loans: loans with long maturities and loans from farmers with high ex-ante credit scores. I consider that these results can be extrapolated to other samples, given the results presented previously on income recovery. Furthermore, I explain in detail how the timing of the recovery in long term loans and the persistent effect on access to credit documented in Section 3.3.3, imply that the bank is not lending to borrowers who could repay a second loan.
4.3.1 Recovery of Long Term Loans

In this section, I present evidence on repayment recovery in a sample of loans of maturities of five years or more. More precisely, I estimate by OLS the following equation, for \( k \) in \{1,2,3,4,5\}:

\[
y_{k,ijk \tau} = \beta_k s_{ij \tau} + \psi_{\tau} + \delta_j + \nu_{kij \tau} \tag{5}
\]

This equation is identical to equation 1 but the outcome is a dummy equal to 1 if loan \( i \) ever entered in a period of 30 days past due at age \( k \) (where \( k \) is in given in years). As before, \( i \) indexes loans, \( \tau \) indexes quarter of origination, and \( j \) rainfall stations. \( s_{ij \tau} \) is a dummy for rainfall shocks in the first year after the loan was disbursed. \( \psi_{\tau} \) are quarter-of-disbursement fixed effects, \( \delta_j \) are rainfall station fixed effects and \( \nu_{kij \tau} \) is a mean-zero error term. I cluster errors at the rainfall station level.

In Figure 5, I plot the estimate of \( \beta_k \) for \( k \) in \{1,2,3,4,5\} along with bars indicating 10% confidence intervals. I do this exercise for two different samples. Recall that I have monthly data on repayment for each loan. It is possible that I don’t observe the loans until their maturity. This can either be because the loan was repaid in full before maturity or because the loan was restructured. In the left panel of Figure 5, I don’t take into account the fact that loans can be restructured and assign a 0 to loans even if I don’t observe them in a given year. The right panel removes from the sample loans that were restructured.

The left panel of Figure 5 shows that the occurrence of a rainfall shock in the first year of tenure has a statistically and positive effect on the probability of entering in a period with past dues in years 2 and 3. The effect then fades out and gets close to zero for year 5. The small effect in year one can be explained by the fact that most long term maturity loans have a grace period (of usually one year) where farmers are not required make any capital payments. The pattern depicted in the figure suggests that repayment recovers from the shock in the first year of loan tenure. The right panel of Figure 5, where I remove from the estimating sample loans that were restructured, is similar to that of the left panel. Although it is less pronounced, it also shows that the effect of the shock is monotonically decreasing with the time elapsed since the shock.

---

48Loans of such a long maturity are used in coffee production to finance investments such as planting new coffee trees, or build warehouses to dry coffee, known as “beneficiaderos” in the Colombian coffee jargon.

49Loan restructuring occurs when the bank agrees with the client to start a new obligation that gathers previous obligations. The new obligation normally has different conditions than the previous one (for example a different payment schedule). This is usually done to offer alternatives of repayment to the client when he has entered overdue periods or manifested that he cannot pay.

50For example, consider loan A, that had zero overdues in year 1, zero overdues in year 2, a period of 30 days past due in year 3, but that was repaid in full in year 4. Then I code \( y_{1,A} = 0, y_{2,A} = 0, y_{3,A} = 1, y_{4,A} = 0, y_{5,A} = 0 \).

51I use a list provided by the BAC of restructured loans. This list is not complete though. I constructed a list of restructured loans by taking loans that end before maturity and establishing if there was a loan issued one month before, the same month or one month after for the same farmer. This list is very similar to the list provided by the BAC.
importantly, in both panels there is no significant effect of the shock in year 4 or in year 5. In sum, both figures suggest that in the sample of loans of five years or more, the effect of the shock on repayment dies out in time.

4.3.2 The Timing of Repayment Recovery and the Persistence in Exclusion from Credit Access

Does the timing of the recovery described in the previous section imply that the bank should lend to borrowers not currently receiving loans? In short, the answer is yes. Recall from Section 3.3.3 that the effect of shocks on future access to credit is persistent. In particular, in that section I documented that the effect of a shock on subsequent access can last two years. That amount of time, combined with the repayment recovery presented in Figure 5 implies that the bank is not lending to at least some important fraction of farmers whose repayment will have recovered at the time of the first payment of a potential second loan. In Figure 6, I present a time schematic where I consider four plausible scenarios to illustrate this point. The idea is that the time it takes for the farmer to apply for a new loan together with the time of the bank’s grace periods are enough for the farmer to recover.\footnote{Grace periods are usually one of the components of the loans granted by the bank. For example, for one year maturity loans, the farmer can repay the full amount he owes one year after the loan is disbursed.}

In this section, I work with the results on persistence obtained with the sample that starts with one year maturity loans (columns (4) to (6) of Table 5). The advantage of using this sample is that for a fixed time window between loan maturity and the next application, the time elapsed between a shock and the next application remains constant. This is not the case for the sample that starts with loans of all maturities. In that sample, the time between the first year of the initial loan and the next application varies with maturity (even if the time window between loan maturity and the next application is constant).

I consider two time windows between maturity of the first loan and the subsequent application, corresponding to those studied in Section 3.3.3 (Table 5, and in particular Column (5)). For the sample starting with one year maturity loans a shock during tenure increases the probability of denial of the subsequent loan application (at the CIFIN stage). This increase is of 22\% for a time window of two years between the maturity of the first loan and the next application and of 20\% for a time window of one year.

I consider four different scenarios depending on the time window and on whether or not I allow for a grace period on the second loan. To be conservative, I consider only one year grace periods. This yields four scenarios to study. Scenario 1: a two years time window and no grace period. Scenario 2: a two years time window with a grace period. Scenario 3: a one year time window and a grace period. Scenario 4: a one year time window and no grace period.

Consider Scenario 1 depicted in Figure 6 (top-left panel). In this Scenario, a shock occurred
during tenure of the initial loan, two years elapsed until the next application and there was no grace period. In this case, the probability of denial of the next loan is 22% higher as a consequence of the occurrence of the shock. If the recovery in repayment is the same as the one depicted in Figure 5 (Panel B), then repayment of a second loan will start at a point in time where repayment will have already recovered from the shock. As it can be seen in the schematic, repayment recovers in years 4 and 5 and repayment of the second loan would start in year 4. Nevertheless, the applications following tenures with a shock are 22% more likely to be denied. This implies that the bank is denying loans that in principle would be repaid.

In Scenario 2 (top-right panel of Figure 6) I show a situation where two years elapse between maturity of the first loan an there is a grace period for the second loan. In this case, repayment of a potential second loan would start in year 5, a point in time when repayment has recovered for more that a year. The increase in the denial rate is the same as before: 22% higher for applications after a loan with a shock, even though repayment would have recovered at the time of the first payment.

In Scenario 3 and 4 (bottom-left and bottom-right panels of Figure 6, respectively), I consider a one year time window between maturity of the first loan and the subsequent application. Again, for Scenario 3, the grace period guarantees that repayment will have recovered at the time of the first payment of the second loan (year 4). Despite this fact, loan denial is 20% higher for applications with a shock during the previous loan tenure. Scenario 4 is the only one from those considered in this analysis where repayment of a second loans starts at year 3, a point in time where repayment has not recovered yet.

The three scenarios where repayment recovers in time for the first payment of the second loan are plausible scenarios. For example, the entire sample that starts with one year maturity loans in Table 5 consists of 20,549 applications. Of these, 22% (4,558) correspond to a time window of two years between maturity and the next application. As the two top panels of Figure 6 show, even with no grace period, for all these loans repayment will have recovered at the time of the first payment of a potential second loan. Nevertheless, denial is 22% more likely for applications following a loan with a shock. Scenario 3 is also frequent, given that all one year maturity loans have a one year grace period.

This analysis assumes though that the repayment recovery of Figure 6 applies to all loans (and not only long term maturity loans). But given the results on income recovery obtained with the production data (Section 4.2) and according to which only shocks not older than one year affect production and sales, this seems a reasonable assumption. Furthermore and as a general rule, a farmer has many plots with different characteristics within the same farm. Plots in a single coffee farm can differ in many ways, for example in the age of the coffee trees planted, the coffee variety, the density (trees per hectare) and the pattern in which they are planted. All these variables have implications in terms of the productivity of the coffee tree. Usually, farmers have plots with
different characteristic to smooth production (a practice that is also recommended by the FNC). Long terms loans are most frequently used for the plantation of new coffee trees, most probably in a single plot out of the many in the farm.\footnote{For example, in the SICA data in 2011 there are about 705,000 farms observed. The average number of plots across these farms is 4.74. Averaging across farms the mean age of their plots yields 10.4 years and averaging the standard deviation of the age of their plots yields 4.5 years. This implies that within farms there is considerable variation in the age of the coffee trees across plots. For the smoothing reasons mentioned above, it is very unlikely that a farmer will renew all the trees of his farm at a single point in time.} Although the plots with the new coffee trees are not producing any coffee, the farmer can pay the loan with the income generated by the other plots. Therefore, the repayment of the long term loan reflects the repayment ability generated from the production of the whole farm (and all of its plots) and not only the plot where new trees were planted using the long term loan.

The loses for the bank of not lending to the set of borrowers who are denied a loan because of rainfall shocks are large. Under some assumptions, they amount in my sample to 7,146 millions of Colombian pesos in revenue (about 3.4 millions of US$), and about 6,066 millions of Colombian pesos in utilities (about 2.9 millions of US$). These are approximately 2.2\% and 5.6\% of the bank’s revenues and utilities in the first quarter of 2013 respectively.\footnote{In my full sample, I observe around 274,000 loans disbursed in the period of 2005-2011 and with an average maturity of 3.4 years (these are the loans in Table A.1 in the Appendix). Suppose that the proportion of farmers who apply for a second loan is the same as the one in the estimation sample (64\%). This implies 174,300 new applications under a re-application rate of 64\%. For this sample, 39\% of the loans had a shock in the first year. So there were around 70,000 applications for which a shock in the previous loan occurred. Now, the average time window between the end of the first loan and the second loan is 1.3 years (taken from the sample in Panel A of Table 4). From column (2) of Panel A in Table 5, the CIFIN denial rate is 0.017 points larger for applications - occurring more than a year after the maturity of the first loan - that had a shock in first the loan tenure. Therefore in the CIFIN stage, out of the 70,000 applications, 1,190 applications (= 0.017 \times 70,000) are denied because of the shock. Of the 70,000 applications (and assuming that the rate of loan denial in the CIFIN stage is that of the control group in the same column, and equal to 0.144), around 58,200 applications following a loan tenure with a shock make it to the analysis stage. The rate of denial in this case is 0.027 points higher for applications following a tenure with a shock. Therefore, in the analysis stage, about 1,570 (= 0.027 \times 58,200) additional applications are denied. This gives a total of 2,760 loans denied because of the rainfall shocks. Now, during this period the average yearly interest rate for loans to small farmers was about 14.4 percent, and the average size of the loan in my sample of coffee farmers was about 10,415 thousands of Colombian pesos. If I assume that all loans denied because of the shock would have been repaid, a simple amortization schedule for a maturity of 3.4 years with an annual interest rate of 14.4 percent, implies that the bank is losing 7,146 millions of Colombian pesos in interest revenues (about 3.4 millions of US$), and -subtracting the costs of operation for this amount of loans – about 6,066 millions of Colombian pesos in utilities (about 2.9 millions of US$).} The losses of the BAC are not restricted to the ones calculated in this sample. In August of 2014, the bank had 860,298 loans for small farmers outstanding. If the mechanisms documented in this paper apply for agricultural activities beyond coffee, it is likely that the losses of not incorporating information on exogenous shocks in credit histories and credit scores are larger. Furthermore, these calculations do not include the welfare losses coming the revenues lost by the farmers who are denied a loan they could repay.
4.3.3 Recovery of High Credit Score Farmers

In this section I focus on loans from farmers with high credit scores at the time of loan origination. The advantage of considering this sample is that even if these farmers get a rainfall shock during loan tenure, it is less likely that they will be denied a subsequent loan. Therefore, selection is less of a concern in this scenario and it is more likely that I can observe repayment of a second loan. I study in this sample whether a shock during the first loan tenure has an effect on repayment of the next loan.

Since I only have data on the CIFIN Stage starting in 2010 and data on rainfall shocks up to 2012, I focus on the sample of loans originated in the period 2010-2011. For these set of loans I observe the CIFIN credit score at the time of loan origination. I restrict the sample to loans of farmers with a CIFIN score in the top quartile of the original sample. As in section 3.3.1, I construct a sample that starts with an initial loan and then look for the subsequent loan application.

I estimate the effect of a rainfall shock on different outcomes. Table 8 present the results. Column (1) presents results where the outcome is dummy for whether or not the initial loan was overdue for 30 days or more.\textsuperscript{55} In both panels, I find a significant effect of a rainfall shock on repayment of the initial loan. This result implies that rainfall shocks affect repayment behavior even for ex-ante high credit score farmers. Columns (2) and (3) present the results of the effect of a rainfall shock on a dummy for denial at the CIFIN Stage and the Analysis Stage respectively, for loan applications following the initial loan tenure.\textsuperscript{56} I find no significant effect on the probability of denial, which confirms that selection is less important in the sample of high credit score loans. Finally, Column (4) presents results where the outcome is repayment of a second loan. In this case, the sample consists of subsequent loans (i.e. that were approved) after tenure of the initial loan. Again, I find no significant effect of shocks in the first loan on repayment of the second loan.

In sum, for high credit score farmers, I find that shocks in the first loan do not have a significant effect on repayment of a subsequent loan. This implies that repayment recovers from rainfall shocks for this sample of high credit risks, consistent with the findings on income recovery.

4.3.4 Frequency of Default, Borrowers Quality and Exogenous Shocks

It is interesting to note in Table 8 that the average rate of default of initial loans with no rainfall shock is 1.3%. A rainfall shock increases this default rate to 108%. These magnitudes are different than my main estimates from Table 2. There, the average default rate is 15.5% and the rainfall shock causes an increase in the default probability of 22%.

Although these differences seem large at first they correspond to what is expected. First, high ex-ante credit score borrowers (those in the sample of Table 8) have a much lower rate of

\textsuperscript{55}The estimated equation is identical to equation 1.
\textsuperscript{56}In this case, the estimated equation is identical to equation 3.
default that the rest of the population (for example, borrowers in the sample of Table 2). More interestingly, exogenous shocks increase by much more the rate of default (relative to the mean) in the case of high quality borrowers. The average borrower defaults for many different reasons, one of which are exogenous shocks. For high quality borrowers, there are very few causes of default, therefore exogenous shocks carry much more importance in causing a default.

5 Discussion and Concluding Comments

In this paper, I documented a market failure that results from the use of traditional credit scoring and credit reports in agricultural lending. In particular, I have shown that excessive rainfall shocks affect farmers’ repayment, credit scores, and future access to credit. Furthermore, I showed that income and repayment recover faster than credit access. These results imply costs both to the farmer and to the lender. Incorporating information on rainfall shocks in credit scores would likely increase efficiency and welfare in this credit market. A large recent literature has documented the benefits of using weather information to create index based insurance. In a similar fashion, this paper shows that precise information on weather events can be used to obtain better credit scoring systems and to allocate credit more appropriately in developing countries.

A concern worth discussing is that my study considers a single bank in a single developing country. However, at least five countries in Latin America have banks or micro-finance institutions lending to farmers and consulting their credit histories in credit bureaus. There is additional evidence of micro-credit institutions seeking information from credit bureaus or sharing information of their clients as a discipline device. Finally, since the late ’90s there has been a return of Public Development Banks. A survey of 90 of these banks in 61 countries conducted in 2011 by the World Bank revealed that 92 percent target small and medium enterprises, and 83 percent target agricultural business (de Luna-Martínez and Vicente, 2012). Additionally, there is a consensus among policy makers that Public Development Banks should be regulated as private banks (de Olloqui, 2013). This usually implies maintaining and reporting credit histories of bank clients to credit bureaus and financial authorities. Therefore, it is likely that the mechanisms documented in this paper are relevant for other Public Development Banks. Also, as mentioned before, the mechanism outlined here are likely to apply to credit scores in other credit markets, although in those scenarios, orthogonal shocks might be less important determinants of repayment or information on them might be hard to obtain, in different contexts.

If the mechanisms that I document in this paper exist, why haven’t lending institutions taken them into account? The most plausible hypothesis is that until recently it was very costly or simply not feasible to use precise measures of weather in the computation of credit scores. It is no mystery

57 See for example Karlan, Osei, Osei-Akoto, and Udry (2014).
58 See for example de Janvry, McIntosh, and Sadoulet (2010) and Giné, Goldberg, and Yang (2012)
to institutions lending to small farmers that weather affects lender’s repayment and profitability. Nevertheless the exercises performed in this paper require levels of measurement and precision that historically could not be implemented because of cost or technological constraints.\textsuperscript{59} For example, geo-referencing technologies became available at low costs only recently. In Colombia, most farmers are still not geo-referenced. Also, dense networks of weather stations with long time series of weather information remain uncommon.\textsuperscript{60} A second possibility is that the fact that credit scoring systems are well suited for consumer lending in developed countries might lead one to conclude that they can be applied with no strings attached to agricultural lending in developing countries. In other words, there might be inertia in banking practices across banks. Furthermore, the fact that lending institutions are ignoring information on the sources of credit downgrades is not unique to my setting.\textsuperscript{61}

In the absence of complete insurance markets, a direct policy implication emerges from the results of this paper. Precise weather information should be used when computing credit scores. This entails geo-referencing farmers and establishing systems that accurately measure weather events near them. Once these systems are in place, detailed records should be kept. These can be included in credit reports and used when computing credit scores. Furthermore, the findings that I presented constitute an example of a situation where a policy, practice or institution used in developed countries can and should be adjusted for the particular setting of developing countries.\textsuperscript{62}

The results of this paper open different avenues for future research. First, the welfare implications of omitting exogenous shocks from credit scores can be estimated in a structural model of lender and borrower decisions. Second, there may be other credit markets where differentiating between causes of default and credit downgrades could improve credit allocation. Third, it opens the question of how to better measure and incorporate information on exogenous shocks for credit histories and credit scores.

\textsuperscript{59}The BAC is aware of a theoretical relationship between weather and repayment and it invest resources in monitoring weather at an aggregate level, for example, the municipality level. But according to BAC officers, previous exercises at this level of aggregation show no effect on repayment behavior. The study by Castro and Garcia (2014) use BAC data to estimate in a structural risk model the effect of weather at an aggregate level. These levels of aggregation though cannot inform individual credit histories and credit scores.

\textsuperscript{60}One of the advances of this paper consists in finding a scenario where such a network is available. In particular geo-referenced coffee farms combined with data from the IDEAM. It is important to note that at the time of this study the IDEAM information was not available for the bank to be used at the level of disaggregation of this paper.

\textsuperscript{61}Garmaise and Natividad (2016), discussed in the introduction, document a similar situation in consumer lending in Peru. In their case, lending institutions have all the information needed to understand the source of an exogenous score downgrade but still do not use it. This echoes the fact that the BAC does not use such information on weather shocks.

\textsuperscript{62}There is a recent literature in the intersection of public finance and development that studies how the policies that are optimal in developed countries might not be optimal for developing countries. See for example Gordon and Li (2009) and Best, Brockmeyer, Kleven, Spinnewijn, and Waseem (2015). My scenario is different in that the inefficiency this paper documents from the use of traditional credit scoring likely exists in developed countries as well. The point is that for agricultural lending in developing countries we can observe the exogenous shocks that generate the inefficiency.
References


ESLAVA, M., A. J. GALINDO, M. HOFSTETTER, AND A. IZQUIERDO (2010): “Scarring Recessions and


Investigaciones Geográficas, Boletín del Instituto de Geografía, 2014(85), 95–112.


Figure 1: Coffee Farmers and Rainfall Stations Distribution Across Space

A. SICA farms 2010-2013

B. Rainfall Stations (close to a coffee farm)
Figure 2: Distribution of Maturities in Estimation Sample
Figure 3: Harvest Time Line

- 2003: Weather events do not affect 2005’s coffee tree productivity
- 2004: Weather events affect 2005’s coffee tree productivity
- 2005: Harvest
- Survey

Figure 4: Treated Rainfall Stations
Figure 5: Repayment Recovery in Loans of Five Year or More Maturities

Notes: Each panel plots coefficients for the estimated effect of a shock in the first year of loan tenure on a dummy equal to one if the loan entered in a period of 30 days past due, at any given age of the loan. The age of the loan is represented in years in the horizontal axis. Vertical bars correspond to 10% confidence intervals. The sample consists of loans of maturities of five or more years disbursed in the period 2008-2011.
Figure 6: Recovery Schematic

**Scenario 1:**
Application After Two Years, No Grace Period

**Scenario 2:**
Application After Two Years, One Year Grace Period

**Scenario 3:**
Application After One Year, One Year Grace Period

**Scenario 4:**
Application After One Year, No Grace Period
## Table 1: Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>St. Dev.</th>
<th>Min</th>
<th>Pctl(25)</th>
<th>Median</th>
<th>Pctl(75)</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maturity (years)</td>
<td>1.65</td>
<td>1.40</td>
<td>0.17</td>
<td>1.00</td>
<td>1.00</td>
<td>2.00</td>
<td>6.58</td>
</tr>
<tr>
<td>Interest Rate (annual)</td>
<td>10.67</td>
<td>2.72</td>
<td>2.47</td>
<td>9.50</td>
<td>10.02</td>
<td>12.54</td>
<td>43.98</td>
</tr>
<tr>
<td>Distance to Rainfall S.</td>
<td>6.54</td>
<td>3.98</td>
<td>0.04</td>
<td>3.86</td>
<td>5.96</td>
<td>8.37</td>
<td>38.13</td>
</tr>
<tr>
<td>Rainfall Shock, year 1</td>
<td>0.42</td>
<td>0.49</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td># quarter-shocks, year 1</td>
<td>1.35</td>
<td>1.06</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>30 Days Overdue</td>
<td>0.14</td>
<td>0.35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>60 Days Overdue</td>
<td>0.10</td>
<td>0.30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BAC Score Fell</td>
<td>0.14</td>
<td>0.35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>BAC Score Fell to E</td>
<td>0.07</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

*Notes:* The data source is the BAC administrative data. Included loans are for coffee production, originated in the period of 2008-2011 and for which there is a subsequent application observed in the CIFIN Stage.
Table 2: Effect of Rainfall Shocks on Repayment

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th></th>
<th>Heterogeneous Effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30 Days</td>
<td>60 Days</td>
<td>Dist. to Stat.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Overdue</td>
<td>Overdue</td>
<td>&lt; Median</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maturity: Short</td>
<td>Maturity: Long</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td></td>
</tr>
<tr>
<td>Rainfall Shock, year 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.034***</td>
<td>0.022***</td>
<td>0.017*</td>
<td>0.032***</td>
</tr>
<tr>
<td># quarter-shocks, year 1</td>
<td>0.016***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.009)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>Mean (control group)</td>
<td>0.155</td>
<td>0.117</td>
<td>0.11</td>
<td>0.12</td>
</tr>
<tr>
<td>Mean (all obs.)</td>
<td>0.142</td>
<td>0.114</td>
<td>0.12</td>
<td>0.138</td>
</tr>
<tr>
<td>Origin Date * Matu. FE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Origin Date FE</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>32,512</td>
<td>32,512</td>
<td>16,590</td>
<td>28,047</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4465</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.125</td>
<td>0.124</td>
<td>0.114</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.138</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.061</td>
</tr>
</tbody>
</table>

Notes: The data source is the BAC administrative data. Included loans are for coffee production, originated in the period of 2008-2011 and for which there is a subsequent application observed in the CIFIN Stage. In a given calendar year, a rainfall shock is defined by rainfall realizations in at least two quarters above the 80th percentile of the quarter-year and station specific rainfall distribution of 1982-2012. Each loan is linked to the closest rainfall station at the time of loan disbursement using the coordinate of the farmer’s largest farm and the rainfall station coordinate, using the Euclidean distance. Origination date fixed effects are effects for the quarter of origination (for example 2005-2). The maturity fixed effect distinguishes between long and short term loans (equal to 1 for loans with maturities of three years or more). The outcome of columns (1), and (3) – (6) is a dummy for loans that ever entered into a period with overdues of 30 days or more. The outcome of column (2) is a dummy for loans that ever entered into a period with overdues of 60 days or more. Standard errors clustered at the rainfall station level are reported in parentheses.

*p<0.1; **p<0.05; ***p<0.01
Table 3: Effect of Rainfall Shocks on Reported BAC Scores

<table>
<thead>
<tr>
<th></th>
<th>Score Fell from A</th>
<th>Score Fell to E</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(1)</strong></td>
<td><strong>(2)</strong></td>
<td></td>
</tr>
<tr>
<td>Rainfall Shock, year 1</td>
<td>0.029***</td>
<td>0.017***</td>
</tr>
<tr>
<td></td>
<td>(0.007)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Mean (control group)</td>
<td>0.150</td>
<td>0.085</td>
</tr>
<tr>
<td>Origin Date * Matu. FE</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. FE</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst.</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>32,512</td>
<td>32,512</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.16</td>
<td>0.075</td>
</tr>
</tbody>
</table>

**Notes:** Included loans are for coffee production, originated in the period of 2008-2011 and for which there is a subsequent application observed in the CIFIN Stage. The outcome in column (1) is a dummy for loans for which the BAC Reported Score fell from A to any other score at any point during loan tenure. The outcome in column (2) is a dummy for loans for which the BAC Reported Score fell from A to the lowest possible score, E. Standard errors clustered at the rainfall station level are reported in parentheses.

*p<0.1; **p<0.05; ***p<0.01
Table 4: Effect of Rainfall Shocks on Future Credit Access

<table>
<thead>
<tr>
<th>Applied New Loan</th>
<th>Initial Loan Overdue</th>
<th>CIFIN Score</th>
<th>CIFIN Denial</th>
<th>Analysis Denial</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>A. Initial Loan Maturity: All</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall Shock, (0.005)</td>
<td>0.034*** (0.007)</td>
<td>-5.747*** (1.96)</td>
<td>0.015*** (0.005)</td>
<td>0.017*** (0.007)</td>
</tr>
<tr>
<td>Mean (control group) 0.816</td>
<td>0.21</td>
<td>925</td>
<td>0.119</td>
<td>0.173</td>
</tr>
<tr>
<td>Origin Date FE N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Origin Date * Matu. FE Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. FE Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst. Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations 51,102</td>
<td>32,512</td>
<td>31,939</td>
<td>32,512</td>
<td>24,083</td>
</tr>
<tr>
<td>Adjusted R² 0.21</td>
<td>0.13</td>
<td>0.074</td>
<td>0.048</td>
<td>0.019</td>
</tr>
<tr>
<td>B. Initial Loan Maturity: 1 Year</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall Shock, (0.009)</td>
<td>0.024*** (0.006)</td>
<td>-7.148*** (2.155)</td>
<td>0.019*** (0.007)</td>
<td>0.019*** (0.009)</td>
</tr>
<tr>
<td>Mean (control group) 0.835</td>
<td>0.149</td>
<td>941</td>
<td>0.109</td>
<td>0.16</td>
</tr>
<tr>
<td>Origin Date FE Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Origin Date * Matu. FE N</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Rainfall St. FE Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst. Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations 28,177</td>
<td>20,549</td>
<td>20,161</td>
<td>20,549</td>
<td>16,368</td>
</tr>
<tr>
<td>Adjusted R² 0.28</td>
<td>0.152</td>
<td>0.049</td>
<td>0.059</td>
<td>0.019</td>
</tr>
</tbody>
</table>

Notes: The data source is the BAC administrative data. One observation corresponds to one loan application of one individual, following loans irrespective of maturity (panel A) and applications following one year loans (panel B). The outcome of Column (1) is a dummy for individuals who applied for a new loan after the initial loan. The outcome of Column (2) is a dummy for loans that entered into a period of 30 days past due in the sample of initial loans. The outcome of Column (3) is the score reported by CIFIN when the farmer applies for a new loan and the outcome of Column (4) is a dummy for applications denied at the CIFIN Stage. The outcome of Column (5) is a dummy for denial at the Analysis Stage.

*p<0.1; **p<0.05; ***p<0.01
Table 5: Effect of Rainfall Shocks on Future Credit Access: Persistence

<table>
<thead>
<tr>
<th></th>
<th>First Loan: All Maturities</th>
<th>First Loan: 1 Year Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CIFIN Score</td>
<td>CIFIN Denial</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>A. Time Lapse Maturity to Application: 1 Year</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall Shock, year 1</td>
<td>-5.7***</td>
<td>0.017*</td>
</tr>
<tr>
<td></td>
<td>(3.3)</td>
<td>(0.009)</td>
</tr>
<tr>
<td>Mean (control group)</td>
<td>914</td>
<td>0.144</td>
</tr>
<tr>
<td>Observations</td>
<td>12,797</td>
<td>13,108</td>
</tr>
<tr>
<td>B. Time Lapse Maturity to Application: 2 Years</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rainfall Shock, year 1</td>
<td>-9.7*</td>
<td>0.2</td>
</tr>
<tr>
<td></td>
<td>(5.168)</td>
<td>(0.014)</td>
</tr>
<tr>
<td>Mean (control group)</td>
<td>913</td>
<td>0.148</td>
</tr>
<tr>
<td>Observations</td>
<td>6,380</td>
<td>6,457</td>
</tr>
<tr>
<td>Origin Date FE</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Origin Date * Matu. FE</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. FE</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst.</td>
<td>Y</td>
<td>Y</td>
</tr>
</tbody>
</table>

Notes: The data source is the BAC administrative data. One observation corresponds to one loan application of one individual. Columns (1) to (3) correspond to applications after loans irrespective of maturity whereas columns (4) to (6) correspond to applications following one year loans. Each panel corresponds to a different sample depending on the duration of the time window between maturity of the initial loan and the next loan application. Each reported coefficient corresponds to a different regression. The outcomes in Columns (1)-(2) and (4)-(5) correspond to the first application observed after the maturity of the corresponding loan, in the CIFIN Consult Stage. Columns (3) and (6) correspond to the first application that makes it to the Analysis Stage. The CIFIN Score (i.e. the outcome of Columns (1) and (4)) corresponds to the score reported by the CIFIN Credit Bureau (at the time of loan application). The outcome of Columns (2) and (5) is a dummy for denial at the CIFIN Consult Stage. The outcome of Columns (3) and (6) is a dummy for denial at the Analysis Stage. Standard errors clustered at the rainfall station level are reported in parentheses.

*p<0.1; **p<0.05; ***p<0.01
## Table 6: Income Recovery from Rainfall Shocks

<table>
<thead>
<tr>
<th></th>
<th>No Controls</th>
<th>Controls (pred.)</th>
<th>Controls (all)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
</tr>
<tr>
<td>Rainfall Shock 04-05</td>
<td>-7.67**</td>
<td>-7.46**</td>
<td>-6.43**</td>
</tr>
<tr>
<td></td>
<td>(3.52)</td>
<td>(3.49)</td>
<td>(2.92)</td>
</tr>
<tr>
<td>Rainfall Shock 02-03</td>
<td>3.75</td>
<td>3.64</td>
<td>6.15</td>
</tr>
<tr>
<td></td>
<td>(4.70)</td>
<td>(4.61)</td>
<td>(5.46)</td>
</tr>
<tr>
<td>Rainfall Shock 00-01</td>
<td>3.82</td>
<td>4.30</td>
<td>2.96</td>
</tr>
<tr>
<td></td>
<td>(5.22)</td>
<td>(5.19)</td>
<td>(5.73)</td>
</tr>
<tr>
<td>Farm Area</td>
<td>0.018</td>
<td>-0.042</td>
<td>(0.20)</td>
</tr>
<tr>
<td></td>
<td>(0.20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Household Size</td>
<td>0.64</td>
<td>0.51</td>
<td>(0.40)</td>
</tr>
<tr>
<td></td>
<td>(0.40)</td>
<td></td>
<td>(0.38)</td>
</tr>
<tr>
<td>Education</td>
<td>1.86*</td>
<td>1.79</td>
<td>(1.13)</td>
</tr>
<tr>
<td></td>
<td>(1.13)</td>
<td></td>
<td>(1.09)</td>
</tr>
<tr>
<td>Gender</td>
<td>-5.39***</td>
<td>-4.70**</td>
<td>(1.56)</td>
</tr>
<tr>
<td></td>
<td>(1.56)</td>
<td></td>
<td>(1.63)</td>
</tr>
<tr>
<td>Density</td>
<td></td>
<td></td>
<td>2.6***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(5)</td>
</tr>
<tr>
<td>Average Age</td>
<td>-0.16*</td>
<td></td>
<td>(0.088)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.88)</td>
</tr>
<tr>
<td>Sun Exposed</td>
<td></td>
<td></td>
<td>1.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.94)</td>
</tr>
<tr>
<td>Constant</td>
<td>24.4***</td>
<td>26.6***</td>
<td>15.3***</td>
</tr>
<tr>
<td></td>
<td>(1.61)</td>
<td>(3.49)</td>
<td>(3.70)</td>
</tr>
</tbody>
</table>

### Notes:
The data comes from the MLYCC survey. Each column corresponds to a different regression. Only farms in a distance smaller than 6.15 km to the rainfall station are included in the sample. The outcome in all regressions is the number of arrobas (one arroba = 12.5 kg) sold per-hectare cultivated with coffee. Rainfall Shock 04-05 is a dummy for the occurrence of a rainfall shock in year 2004 or 2005. Rainfall Shock 02-03 and Shock 00-01 are defined analogously. Farm Area is the size of the farm in hectares. Household Size is the number of household members. Education is an ordered variable that increases with the level of education of the household head. Gender is a dummy for female household head. Density is trees per-hectare of the coffee plot. Average Age is average age (in years) of coffee plots. Sun exposed is dummy for farms with coffee plots with no shade. 128 farms had a rainfall shock in 2004-2005. There are 246 rainfall station clusters. Of these, 26 had a rainfall shock in 2004-2005, 7 in 2002-2003, and 13 in 2000-2001. Standard errors clustered at the rainfall station level are reported in parentheses.

* p<0.1; ** p<0.05; *** p<0.01
Table 7: Covariate Balance

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household Size</td>
<td>3.93</td>
<td>4.10</td>
<td>0.300</td>
</tr>
<tr>
<td>Education</td>
<td>0.99</td>
<td>0.89</td>
<td>0.060 *</td>
</tr>
<tr>
<td>Gender</td>
<td>1.18</td>
<td>1.16</td>
<td>0.597</td>
</tr>
<tr>
<td>Coffee Area</td>
<td>3.1</td>
<td>2.8</td>
<td>0.204</td>
</tr>
<tr>
<td>Density</td>
<td>4323</td>
<td>4030</td>
<td>0.059 *</td>
</tr>
<tr>
<td>Average Age</td>
<td>8.3</td>
<td>8.7</td>
<td>0.540</td>
</tr>
<tr>
<td>Farm Area</td>
<td>5.8</td>
<td>5.7</td>
<td>0.840</td>
</tr>
<tr>
<td>Sun Exposed</td>
<td>0.19</td>
<td>0.24</td>
<td>0.130</td>
</tr>
</tbody>
</table>

Notes: The data comes from the MLYCC survey. Farms in the treatment group correspond to farms for which the closest rainfall station had a rainfall shock in 2005 or 2004. Control farms are the rest. The reported p-value corresponds to a test where the null hypothesis is equality in means across treated and control groups.

*p<0.1; **p<0.05; ***p<0.01
Table 8: Repayment Recovery of High Credit Score Farmers

<table>
<thead>
<tr>
<th></th>
<th>1st Loan</th>
<th>CIFIN</th>
<th>Analysis</th>
<th>2nd Loan</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Overdue (1)</td>
<td>Denial (2)</td>
<td>Denial (3)</td>
<td>Overdue (4)</td>
</tr>
<tr>
<td>Rainfall Shock</td>
<td>0.014** (0.008)</td>
<td>0.011 (0.012)</td>
<td>0.012 (0.022)</td>
<td>0.033 (0.021)</td>
</tr>
<tr>
<td>Mean (control group)</td>
<td>0.013</td>
<td>0.055</td>
<td>0.141</td>
<td>0.063</td>
</tr>
<tr>
<td>Origin Date FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Origin Date * Matu. FE</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Rainfall St. FE</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst.</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>2,681</td>
<td>3,785</td>
<td>3,141</td>
<td>2,550</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.028</td>
<td>0.078</td>
<td>0.025</td>
<td>0.043</td>
</tr>
</tbody>
</table>

Notes: The data source is the BAC administrative data. Each observation corresponds to one individual. The sample of Column (1) consists of loans originated in 2010-2011, from farmers in the top quartile of the CIFIN credit score distribution and that applied for a subsequent loan. The samples of Columns (2) and (3) correspond to one loan application following loans in the sample of Column (1). The sample in Column (4) consists of loans that were approved after the initial loan. The outcomes are: Columns (1) and (4), a dummy equal to one if the corresponding loan ever entered into a period of 30 days past due, Column (2), a dummy for denial of the subsequent application at the CIFIN Stage, and Column (3), a dummy for denial of the subsequent application at the Analysis Stage. Standard errors clustered at the rainfall station level are reported in parentheses.

*p<0.1; **p<0.05; ***p<0.01
Appendix A: Robustness Exercises

This appendix presents robustness results on the effect of rainfall shocks on repayment (Table 2) and on scores reported by the BAC (Table 3). They differ from the results presented in the main text in that this sample includes all loans disbursed between 2005 and 2011.

Table A.1: Effect of Rainfall Shocks on Repayment

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Heterogeneous Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>30 Days</td>
<td>60 Days</td>
</tr>
<tr>
<td></td>
<td>Overdue</td>
<td>Overdue</td>
</tr>
<tr>
<td>Rainfall Shock, year 1</td>
<td>0.008***</td>
<td>0.006**</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td># quarter-shocks, year 1</td>
<td>0.004***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>Mean (control group)</td>
<td>0.157</td>
<td>0.118</td>
</tr>
<tr>
<td>Mean (all obs.)</td>
<td>0.157</td>
<td></td>
</tr>
<tr>
<td>Origin Date * Matu. FE</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. FE</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst.</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>274,198</td>
<td>274,198</td>
</tr>
<tr>
<td>Adjusted R²</td>
<td>0.079</td>
<td>0.079</td>
</tr>
</tbody>
</table>

Notes: The data source is the BAC administrative data. Included loans are for coffee production and originated in the period of 2005-2011. In a given calendar year, a rainfall shock is defined by rainfall realizations in at least two quarters above the 80th percentile of the quarter-year and station specific rainfall distribution of 1982-2012. Each loan is linked to the closest rainfall station at the time of loan disbursement using the coordinate of the farmer’s largest farm and the rainfall station coordinate, using the Euclidean distance. Origination date fixed effects are effects for the quarter of origination (for example 2005-2). The maturity fixed effect distinguishes between long and short term loans (equal to 1 for loans with maturities of three years or more). The outcome of columns (1), and (3) – (6) is a dummy for loans that ever entered into a period with overdues of 30 days or more. The outcome of column (2) is a dummy for loans that ever entered into a period with overdues of 60 days or more. Standard errors clustered at the rainfall station level are reported in parentheses.

*p<0.1; **p<0.05; ***p<0.01
Table A.2: Effect of Rainfall Shocks on Reported BAC Scores

<table>
<thead>
<tr>
<th>Score Fell from A</th>
<th>Score Fell to E</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>Rainfall Shock, year 1</td>
<td>0.008**</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
</tr>
<tr>
<td>Mean (control group)</td>
<td>0.150</td>
</tr>
<tr>
<td>Origin Date * Matu. FE</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. FE</td>
<td>Y</td>
</tr>
<tr>
<td>Rainfall St. Clst.</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>274,198</td>
</tr>
<tr>
<td>Adjusted R$^2$</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Notes: The data source is the BAC administrative data. Included loans are for coffee production and originated in the period of 2005-2011. The outcome in column (1) is a dummy for loans for which the BAC Reported Score fell from A to any other score at any point of loan tenure. The outcome in column (2) is a dummy for loans for which the BAC Reported Score fell from A to the lowest possible score, E. Standard errors clustered at the rainfall station level are reported in parentheses.

*p<0.1; **p<0.05; ***p<0.01
Appendix B: Theory

In this Appendix, I show with a simple model of lending and credit scoring that the exclusion of exogenous shocks that affect repayment leads the lender to make mistakes more frequently. A mistake is defined as lending to a non-profitable borrower or denying credit to a profitable one. This model is related to work by de Janvry, McIntosh, and Sadoulet (2010). However, unlike previous work, I show how exogenous shocks can lead to a market failure.

Consider a scenario with two periods where a borrower and a lender interact. Suppose that the borrower was granted a loan in period $t-1$ and denote the repayment of this loan by $\pi_{t-1}$. The borrower always applies for a new loan for period $t$ and the lender must decide if he grants this subsequent loan. The borrower is characterized by a level of profitability $\pi_0$ which is unobserved by the lender.

Repayment of the borrower in $t-1$ depends on $\pi_0$ and two random components. More precisely, I assume that:

$$\pi_{t-1} = \pi_0 + z + \epsilon$$  \hspace{1cm} (A1)

where $z$ is an exogenous shock (in the sense that it is independent of $\pi_0$) and that is potentially observable by the lender. I refer to $z$ as a “rainfall shock”. $\epsilon$ is another exogenous component (independent of both $\pi_0$ and $z$) but unobservable to the lender. I assume that $z \sim N(0, \sigma_z^2)$ and $\epsilon \sim N(0, \sigma_\epsilon^2)$. Furthermore, I assume that the lender knows the process that generates $\pi_{t-1}$ (that is, it knows equation A1) but does not observe all of its components. That is, he cannot observe $\pi_0$, $z$, and $\epsilon$ separately.

The repayment of the subsequent loan (in the case the lender is given one) is denoted by $\pi_t$. I assume away any uncertainty in the repayment of this second loan (once the lender has made his decision) so that $\pi_t = \pi_0$. Therefore, the lender makes a positive profit in the second loan if $\pi_0 > 0$ and negative one if $\pi_0 < 0$. To make the decision of whether or not to grant a second loan, the lender makes a prediction of $\pi_t$ based on past borrower behavior. In particular, he forms a “credit score” based on $\pi_{t-1}$. In the case where the rainfall shock is not observed, it is given by:

$$E[\pi_t | \pi_{t-1}] = \pi_{t-1}$$  \hspace{1cm} (A2)

The lender grants the loan if $E[\pi_t | \pi_{t-1}] \geq 0$ and does not otherwise.

Note that this setup has various parallels with my empirical setting. First $z$ affects repayment of the first loan $\pi_{t-1}$ but does not affect repayment of the second loan $\pi_t$. This is consistent with the effect of rainfall shocks. Since there is recovery, rainfall shocks affect repayment of the first loan but don’t affect repayment of the second one. Second, the main determinant of the credit
score is past repayment behavior. Individual characteristics (which are sometimes included in
credit scores) could be included in this setup, but they would not modify the results of the model.

Under the stated assumptions I can compute the probability that the lender makes a mistake.
In particular, I compute the probability that lender grants a loan to an un-profitable borrower
and the probability that it does not grant a loan to a profitable one. I do this for two scenarios:
one in which the rainfall shock \( z \) is not observed by the lender and one in which it is.

**Scenario 1: \( z \) is Unobservable to the Lender**

The lender grants a second if \( E[\pi_t | \pi_{t-1}] \geq 0 \) which is equivalent to \( \pi_{t-1} = \pi_0 + z + \epsilon \geq 0 \) from
equation (A2) and (A1). I denote by \( P_u \) the probability that a loan is granted to an unprofitable
borrower. Therefore, \( P_u = P(\pi_{t-1} \geq 0) = P(\pi_0 + z + \epsilon \geq 0) \) given \( \pi_0 < 0 \). Note that \( z + \epsilon \) is
distributed \( N(0, \sigma_z^2 + \sigma_\epsilon^2) \) since \( z \) and \( \epsilon \) are independent so that \( P(\pi_0 + z + \epsilon \geq 0) = P(z + \epsilon \geq -\pi_0) = P(z + \epsilon \leq \pi_0) \) and

\[ P_u = \Phi\left(\frac{\pi_0}{\sqrt{\sigma_z^2 + \sigma_\epsilon^2}}\right) \] (A3)

Note that this last expression is increasing in \( \sqrt{\sigma_z^2 + \sigma_\epsilon^2} \) given that \( \pi_0 < 0 \). Therefore, the lender
is more likely to make the mistake of lending to an unprofitable borrower the larger the variance
of \( z \) or \( \epsilon \). The intuition is that a larger variance implies that the signal \( \pi_{t-1} \) is less informative on
the profitability of the second loan, \( \pi_0 \).

Now I consider the probability that the lender denies a loan to a profitable borrower and
denote it by \( P_d \). \( P_d \) is given by \( P(\pi_0 + z + \epsilon \leq 0) \) given \( \pi_0 > 0 \). Therefore, \( P_d = P(z + \epsilon \leq -\pi_0) \)
so that, under the distributional assumptions of \( z \) and \( \epsilon \), it can be written as:

\[ P_d = \Phi\left(\frac{-\pi_0}{\sqrt{\sigma_z^2 + \sigma_\epsilon^2}}\right) \] (A4)

Given that \( \pi_0 > 0 \), this expression is also increasing in \( \sqrt{\sigma_z^2 + \sigma_\epsilon^2} \). Again, the probability that
lender makes the mistake (in this case of not lending to a profitable borrower) is increasing in the
variance of the terms \( z \) and \( \epsilon \).

**Scenario 2: \( z \) is Observed by the Lender**

If \( z \) is observed by the lender, his prediction of \( \pi_t \) changes. In particular, since he knows the
process generating \( \pi_{t-1} \) (given by equation A1) he discounts \( \pi_{t-1} \) with the observed value of \( z \).
In other words, the credit score is now \( E[\pi_t | \pi_{t-1}, z] = \pi_{t-1} - z \). Substituting equation A1 yields
\( E[\pi_t|\pi_{t-1}, z] = \pi_0 + \epsilon \). As before, the lender grants the second loan if \( E[\pi_t|\pi_{t-1}, z] \geq 0 \) and does not otherwise.

In this case, the probability of granting a loan to an unprofitable borrower is given by 
\[
P_u = P(E[\pi_t|\pi_{t-1}, z] \geq 0) = P(\pi_0 + \epsilon \geq 0) \text{ with } \pi_0 < 0.
\]
Therefore,
\[
P_u = \Phi \left( \frac{\pi_0}{\sigma_\epsilon} \right)
\]
(A5)

with \( \pi_0 < 0 \). The probability of denying a loan to a profitable borrower, \( P_d \) in this scenario is given by 
\[
P_d = P(\pi_0 + \epsilon \leq 0) \text{ with } \pi_0 > 0,
\]
that is:
\[
P_d = \Phi \left( \frac{-\pi_0}{\sigma_\epsilon} \right)
\]
(A6)

with \( \pi_0 > 0 \).

The probabilities of lender mistakes of both scenarios can be easily compared from equations A3 to A6. Clearly, the probability of lending a loan to an un-profitable borrower, \( P_u \), is larger in Scenario 1 (equation A3) than in Scenario 2 (equation A5), given that both expressions are increasing in the term in the denominator and \( \sigma_z^2 > 0 \). Similarly, the probability of denying a loan to a profitable borrower, \( P_d \), is larger in Scenario 1 (equation A4) than in Scenario 2 (equation A6). The difference in the mistakes probabilities is larger the larger is the variance of the “rainfall shock”, \( \sigma_z^2 \). The intuition is that if the variance of \( z \) is larger, the precision of the signal \( \pi_{t-1} \) increases more when the \( z \) is included in the credit score.

These results imply that the inclusion of exogenous shocks in the credit score increase its precision and reduce the probability of lender mistakes. The gains from this inclusion are larger if the variance of the excluded shocks is large. This is particularly true in the context of agricultural lending in developing countries where exogenous shocks can have large effects on production and income, as shown in the main text.