CEDE
Centro de Estudios sobre Desarrollo Económico

# Exogenous Shocks, Credit Reports and Access to Credit: Evidence from Colombian Coffee Producers

Documento CEDE

#57

Nicolás de Roux

Noviembre de 2021

Universidad de los Andes
Colombia
Facultad de Economía

Serie Documentos Cede, 2021-57 ISSN 1657-7191 Edición electrónica. Noviembre de 2021



infocede@uniandes.edu.co
http://economia.uniandes.edu.co

Impreso en Colombia – Printed in Colombia



CEDE
Centro de Estudios sobre Desarrollo Económico

## Documento CEDE

**Descripción:** los documentos CEDE son producto de las investigaciones realizadas por al menos un profesor de planta de la Facultad de Economía o sus investigadores formalmente asociados.

Universidad de los Andes Colombia | Facultad de Economía

# Exogenous Shocks, Credit Reports and Access to Credit: Evidence from Colombian Coffee Producers*

Nicolás de Roux†

**Abstract**

Credit reporting systems have become a widespread tool to assess the creditworthiness of prospective borrowers. This paper studies the implications for credit access of using them in contexts where exogenous and transitory shocks affect income and repayment. Using a novel administrative data set with the near universe of formal loans to coffee producers in Colombia together with data from close to 1,200 rainfall stations, I show that transitory weather shocks lead to lower rates of loan repayment, lower credit scores, and more frequent denials of future loan applications. I present evidence that affected producers' incomes and ability to repay recover more quickly from shocks than credit access. This implies that these producers become credit constrained despite their ability to repay a loan. Insurance, contingency-dependent repayment schemes, or the inclusion of information on exogenous shocks in credit scoring models have the potential to alleviate the problem.

**JEL Codes**: G21, O12, O13, Q12, Q14, Q54.
**Keywords**: Shocks, Credit Reports, Access to Credit.

*A previous version of this paper circulated under the title "Weather Variability, Credit Scores, and Access to Credit: Evidence from Colombian Coffee Farmers." I am especially grateful to 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 detailed comments and insights, I am especially grateful to Zach Brown, Margarita Gáfaro, Chris Hansman, Pepita Miquel-Florensa, Evan Riehl, and Mica Sviatschi. For useful comments I would like to thank Francesco Amodio, Pietro Bonaldi, Emily Breza, Juan Camilo Cárdenas, Felipe Cortés, Xavier Durán, Marcela Eslava, Leopoldo Fergusson, Natalia Garbiras-Díaz, Juan Herreño, Marc Hofstetter, Dean Karlan, Supreet Kaur, Wojciech Kopczuk, José Leibovich, Ana María Ibáñez, Diego Jimenez, Rachid Laajaj, Nicola Limodio, Luis R. Martínez, Craig McIntosh, Juan Moreno-Cruz, Moisés Mahecha, Andrés Moya, Juan Carlos Muñoz-Mora, José Antonio Ocampo, Guillermo Otero, Kiki Pop-Eleches, Yogita Shamdasani, Diego Ubfal, Miguel Urquiola, Román Andrés Zárate and seminar participants at NEUDC, Banco de la República de Colombia, Columbia University, Universidad de los Andes, Universidad Javeriana and Universidad del Rosario. Sergio Florez, Lucía Mendoza and Natalia Moreno provided excellent research assistance. Many thanks to all members of the Vice-presidency of Risk of the Banco Agrario. I would especially like to thank Germán Camargo and Ana María Torres for their help with the data. I am also grateful to the Federación Nacional de Cafeteros and the Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia. For financial support, I am grateful to Universidad de los Andes. All errors are my own.

†nicolas.de.roux@uniandes.edu.co, Department of Economics, Universidad de Los Andes - Bogotá.

# Choques Exógenos, Reportes Crediticios y Acceso al Crédito: Evidencia de Productores de Café Colombianos*

Nicolás de Roux†

**Resumen**

Los sistemas de reportes de crédito se han convertido en una herramienta de uso generalizado para determinar la calidad crediticia de futuros prestatarios. Este artículo estudia las implicaciones para el acceso al crédito de usar dichos sistemas en contextos donde choques exógenos y transitorios afectan el ingreso y la capacidad de pago. Usando datos administrativos novedosos con casi la totalidad de préstamos formales de productores de café en Colombia junto con datos de cerca de 1,200 estaciones de lluvia, muestro que los choques climáticos transitorios llevan a un menor repago de los préstamos, a menores puntajes crediticios y a rechazos más frecuentes de solicitudes de créditos futuras. Adicionalmente, presento evidencia de que el ingreso de los productores afectados y su capacidad de pago se recuperan más rápido de los choques que su acceso al crédito. Esto implica que estos productores pierden la posibilidad de adquirir préstamos a pesar de que tienen la capacidad para pagarlos. Los seguros, los esquemas de pago contingentes o la inclusión de información sobre choques exógenos en los modelos de puntajes crediticios tienen el potencial de aliviar el problema.

**Códigos JEL**: G21, O12, O13, Q12, Q14, Q54.
**Palabras clave**: Choques, reportes de crédito, acceso a crédito.

---

*Una versión anterior de este artículo circuló bajo el título "Variabilidad Climática, Puntajes Crediticios y Acceso al Crédito: Evidencia desde la Caficultura Colombiana." Estoy especialmente en deuda con Eric Verhoogen por su guía y apoyo. También quisiera agradecer a Pierre-André Chiappori, Jonas Hjort y Suresh Naidu por sus consejos. Por sus comentarios detallados, agradezco especialmente a Zach Brown, Margarita Gáfaro, Chris Hansman, Pepita Miquel-Florensa, Evan Riehl y Mica Sviatschi. Agradezco los comentarios de Francesco Amodio, Emily Breza, Juan Camilo Cárdenas, Felipe Cortés, Xavier Durán, Marcela Eslava, Leopoldo Fergusson, Natalia Garbiras-Díaz, Juan Herreño, Marc Hofstetter, Dean Karlan, Supreet Kaur, Wojciech Kopczuk, José Leibovich, Ana María Ibáñez, Diego Jimenez, Rachid Laajaj, Luis R. Martínez, Craig McIntosh, Juan Moreno-Cruz, Moisés Mahecha, Andrés Moya, Juan Carlos Muñoz-Mora, José Antonio Ocampo, Guillermo Otero, Kiki Pop-Eleches, Yogita Shamdasani, Diego Ubfal, Miguel Urquiola, Román Andrés Zárate y de las personas que participaron en seminarios en NEUDC, el Banco de la República de Colombia, Columbia University, la Universidad de los Andes, la Universidad Javeriana y la Universidad del Rosario. Sergio Florez, Lucía Mendoza y Natalia Moreno hicieron un trabajo excelente como asistentes de investigación. Muchas gracias a todas las personas de la Vicepresidencia de Riesgos del Banco Agrario. Por su ayuda con los datos estoy especialmente agradecido con Germán Camargo y Ana María Torres. Agradezco también a la Federación Nacional de Cafeteros y al Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia. Por su apoyo financiero, agradezco a la Universidad de los Andes. Todos los errores son de mi entera responsabilidad.

†Facultad de Economía, Universidad de los Andes, Bogotá, E-mail: nicolas.de.roux@uniandes.edu.co.

# 1 Introduction

Credit reporting systems have spread quickly in the banking and lending industries facilitating access to credit of previously underserved populations across the globe (Mester, 1997; Berger et al., 2005). Their wide adoption stems from their potential to lessen problems of moral hazard and adverse selection (Pagano and Jappelli, 1993). Indeed, growing empirical work has documented welfare and efficiency gains from their use (Einav et al., 2012). However, individuals' credit histories may not fully reflect their future ability to repay a loan if they were affected by a transitory shock out of their control. While a loan officer may be able to differentiate between borrower luck and borrower type, credit scores generally do not make this distinction. This paper seeks to understand the implications of using credit reporting systems in contexts where transitory and exogenous shocks are common.

This question is relevant for three main reasons. First, in many settings, producers are exposed to transitory and exogenous shocks, like exchange rate fluctuations (see e.g. Berman et al., 2012), commodity price shocks (see e.g. Fafchamps, 1992), or weather shocks (see e.g. Dell et al., 2014). These shocks may temporarily affect repayment ability even if borrowers are otherwise low-risk. Second, the number of institutions using credit scores is increasing rapidly, especially in developing countries where policymakers have encouraged their adoption (World Bank, 2007). For example, micro-credit institutions are seeking information from credit bureaus or sharing information about their clients as a discipline device (de Janvry et al., 2010; Giné et al., 2012). Public development banks, an important source of financing in low and middle-income countries, are also likely to use credit scores to rate potential borrowers (de Luna-Martínez and Vicente, 2012; de Olloqui, 2013). Third, there may be scope to improve credit rating systems or complement them with other financial tools to alleviate the inefficiencies that result from their interaction with exogenous shocks to borrowers' ability to repay. Addressing these issues may be important for improving credit access.

In order to examine how the interaction of exogenous shocks and credit rating systems can hinder access to credit, I study the case of coffee farmers in Colombia who use loans from a large agricultural bank to finance their production. I analyze the effect of weather shocks, which are unrelated to coffee producers' characteristics (including their risk type) and only have a transitory effect on the productivity of coffee trees (Gast et al., 2013). Therefore, they constitute an ideal shock to examine these issues. Specifically, I estimate the effect of these exogenous shocks on multiple credit market outcomes. To do so, I combine novel administrative data on the near universe of formal agricultural loans to small farmers in the country with information from nearly 1,200 rainfall stations. I show a negative effect on the rates of loan repayment and the credit scores reported to credit bureaus. Most importantly, I show that affected coffee producers have lower credit scores in the future and are more likely to have subsequent loan applications rejected; these effects are highly persistent and can last up to five years. I then present additional evidence that this exclusion from credit markets not only affects producers:[1] it also represents missed income for the lender. Specifically, I show that affected producers' income and repayment recover faster than credit scores and access to credit. This is inefficient since many producers who are rejected for future loans would have been able to repay them. As I discuss in more detail below, the policy implications are immediate: this problem can be alleviated with insurance, the use of contingency-dependent repayment schemes, or even with the inclusion of information on exogenous shocks in credit scores.

To document these channels, I use administrative data on loan applications, credit scores, and repayment from the Banco Agrario de Colombia (BAC) [Colombian Agrarian Bank], which in 2013 held 97% of all formal agricultural loans to small farmers in Colombia (DNP and FINAGRO, 2014). I merge the bank's individual-level data from 2005 to 2015 with administrative data from the Federación Nacional de Cafeteros (FNC) [National Federation of Coffee Growers], which provides geographical coordinates for each farmer. I link each loan to the closest rainfall station at the time of the loan disbursement. This allows me to measure with precision rainfall variability at the loan level, using information from nearly 1,200 rainfall stations going back to 1982. I estimate the effect of excessive rainfall on repayment using the universe of BAC loans to small and medium farmers for coffee production from 2005–2011.[2] I focus on excessive rainfall as my measure of weather variability because it reduces the productivity of coffee trees.[3] My empirical specification includes fixed effects for rainfall stations and for quarter-year of disbursement interacted with loan maturity. Thus, the identifying assumption is that the occurrence of weather shocks in different geographic areas and time periods is not systematically correlated with other time-varying factors that affect loan repayment. I present balance tests showing that loan and farm characteristics are not systematically correlated with rainfall shocks.

I find that excessive rainfall decreases producers' ability to repay their outstanding loans, which in turn has a persistent negative effect on their future access to credit. For each quarter of excessive rainfall during the first year of the loan, the probability that it enters into a

[1] 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 et al. (2010).

[2] I use the FNC's definition of small and medium size farmers – those who cultivate 10 hectares or less.

[3] See Boucher and Moya (2015), Gast et al. (2013), and Turbay et al. (2014).

period of 30 days with overdues increases by 0.004 points, equal to 3% of the sample mean. I show that affected farmers have lower credit scores on subsequent applications and are more likely to be denied future loans. In my preferred specification, the probability of loan denial is 0.008 points higher (equal to 7.5% of the sample mean) for each quarter of excessive rainfall in the first year of the previous loan. Using individual-level panel data, I show that the effect on credit scores and loan denial persists for at least five years after the initial excessive rainfall.

Next, and to provide additional evidence that this is inefficient, I show that farmers' incomes and repayment ability recover faster than credit histories. Agricultural research shows that the effects of excessive rainfall on coffee trees' productivity last for only one harvest season (Arcila et al., 2007; Gast et al., 2013), thereby the recovery in productivity of the next harvests should lead to a recovery in income. To test this prediction empirically, I use data from the 2014 Colombian Agricultural Census, which contains information on the 2013 harvests of all farms in the country. I find that excessive rain has a large negative effect on farmers' income in the following year and a half but no effect in any subsequent years, suggesting that their income quickly recovers from this type of shock.[4] Assuming that the results from these exercises also apply to the farmers in the loan data, these findings imply that income recovers faster than access to credit.

But does income recovery necessarily translate into repayment recovery? If the ability to repay depends on factors such as household assets, it may not. Using a sample of long-term loans to measure repayment over a long period of time, I show that excessive rainfall lowers rates of repayment in years 2 and 3 after the shock, but that the effect is smaller in year 4 and disappears in year 5. In other words, repayment starts to rebound in year 3 after the shock and has fully recovered by year 5. These results imply that some coffee producers are denied loans that they would be able to repay since in year 5 credit access is still affected by the shock, which represents lost benefits to the bank. *Ex ante*, it is possible that shocks reveal information to the bank about producers' types that would help improve its lending decisions. My results show that a substantial fraction of producers is able to resume loan repayments a few years after the shock. If this were not the case, I would not observe an average pattern of recovery in the sample of long-term loans.

This paper contributes to several strands of the literature. First, it is related to a finance literature on credit reporting systems. For example, Jappelli and Pagano (2002) and Djankov et al. (2007) document cross-country positive correlations between bank lending and

[4]These findings rely on a stronger identification assumption than in the first part of the paper since the census is only a cross-section. Below I present results from various robustness exercises that support it.

information sharing. Einav et al. (2013), and de Janvry et al. (2010) estimate positive welfare impacts of credit scores. Furthermore, different papers have shown that negative credit reports reduce future access to credit (Musto, 2004; Bos and Nakamura, 2014; Dobbie et al., 2020).[5] Importantly, Garmaise and Natividad (2017) use data from all financial institutions of Peru to document that purely random negative reports decrease credit access. In their setting, customers can borrow in local currency or in U.S dollars and the interaction between exchange rate fluctuations and credit rating thresholds generates exogenous variation in the occurrence of negative credit reports. Interestingly, although banks have all the information to realize that these shocks are purely bad luck and unrelated to the repayment ability of borrowers, they do nothing to unravel the resulting problem.[6] I contribute to this literature by documenting how shocks with an actual effect on income and repayment can lead to an inefficiency. Here the problem arises because shocks are exogenous and with only a transitory effect on productivity. Therefore, they are unrelated to producers' characteristics and they affect access to credit longer than income and repayment.

Second, this paper is related to work in the intersection of finance and development that studies how the features of developing countries interact with financial systems. For example, Conning and Udry (2007) discuss the information frictions that pervade rural agricultural settings with negative consequences for credit access and Limodio (2021) provides causal evidence that the higher volatility of deposits in developing countries lowers the supply of long-term funds. Credit scores were introduced in the 1950s in the United States and were designed for the consumers and the small and medium-sized enterprises of developed countries (Mester, 1997). Acemoglu and Zilibotti (2001) argue that many technologies used in less developed countries were invented with the needs and conditions of rich countries in mind, which explains part of the productivity gap between these two groups. My paper suggests that traditional credit rating systems should be adapted to settings where income is more volatile or complemented with other financial products.

Third, this paper contributes to a large development literature on the consequences of income volatility. For example, extreme weather events can lead to a lower accumulation of productive assets and investment (see e.g. Rosenzweig and Wolpin, 1993), lower consumption (see e.g. Kazianga and Udry, 2006), and lower investment in human capital (see e.g. Jensen, 2000). My results also relate to research on why farmers are credit constrained (see e.g.

[5]Bos et al. (2018) find that negative credit reports reduce labor market opportunities in Sweden while Dobbie et al. (2020) find no such effects in the US.

[6]Relatedly, Avery et al. (2004) argue that circumstances that affect repayment temporarily (like a divorce or economic conditions) can generate problems for credit reporting systems in consumer lending markets.

Giné et al., 2012). The lender I study cannot differentiate between high-risk borrowers and those who experienced a temporary shock, which leads to credit constraints for farmers who would be able to repay future loans. Therefore, farmers can lose access to credit because agricultural production generates volatile income streams that lead to loan default. The longer-term impact on their credit histories and credit scores can lead to a persistent lack of access to future loans.[7]

Agricultural insurance can help overcome the consequences of volatile weather and in particular the problems outlined by this paper. Numerous papers have studied the effects of insurance (see e.g. Cole et al. (2017) and Karlan et al. (2014)) and the determinants of its low demand (see e.g. Cole et al. (2013) and Casaburi and Willis (2018)). My results point to an additional benefit of insurance since it can alleviate the effect of weather variability on income, repayment, credit scores, and future access to credit. But there are other potential alternatives to mitigate the negative channels I document, like weather-dependent repayment schemes, the deletion of past negative reports, or the inclusion of information on shocks in credit scores.[8] I discuss these policy implications in detail in the conclusion section.

The rest of the paper is organized as follows. Section 2 provides background information on the BAC. Section 3 documents the effects of shocks on current loan outcomes while Section 4 shows that shocks have a negative and persistent effect on credit scores and future credit access. Section 5 presents evidence on the recovery of farmers' income and repayment after rainfall shocks. Section 6 concludes.

## 2 The Banco Agrario

The BAC is a publicly owned bank created in 1999 to finance agriculture. It is the main player in Colombia's rural credit market. In 2013 it supplied 97% of all formal loans to small farmers, and it is the only bank present in all of the country's 1,123 municipalities (DNP and FINAGRO, 2014). The vast majority (89%) of the bank's branch offices are in rural areas. The Fondo Agropecuacio de Garantías, (FAG) [Agricultural Guarantees Fund]

[7] More generally, the paper relates to a literature on the effects of weather on economic activity (see e.g Dell et al., 2014). To my knowledge, the only other paper that estimates the effect of rainfall on loan repayment is Pelka et al. (2015), although their rainfall measures are at the bank branch level. See Castro and Garcia (2014) for an estimation of the effect of climate variation in a structural default risk model, using aggregate data from the BAC.

[8] In the Appendix I show in a simple model of borrower screening how using information on exogenous shocks in a credit score can reduce the probability of an inclusion error (lending to an un-profitable borrower) and the probability of an exclusion error (denying credit to a profitable one). The model also allows me to define mathematically what I mean by transitory and exogenous shock.

provides collateral for small farmers.[9] Although the FAG provides most of the collateral, the BAC has strong incentives to screen borrowers since, by regulation, it still has to try to reclaim some collateral from borrowers who default which is expensive. Furthermore, the bank the BAC has to follow similar regulations as private banks and its performance is monitored with indicators of borrower default. Poor performance can force the financial authorities to intervene.

The typical farmer applies for a BAC loan at the branch office closest to his farm. In the office, a bank employee enquires about his standing with the credit bureau. This query produces a credit history report for the borrower, which includes his credit score (which I refer to as the *Bureau Score*) and indicates whether the application process can continue. Denial at this stage of the application depends on the Bureau Score, other variables such as credit history, and BAC policies. For example, applications from borrowers with a Bureau Score below a certain threshold are denied. This paper focuses on the effect of shocks at this stage of the application process. Applications that make it past this stage are sent to the BAC's head office in Bogotá, where a credit analyst assesses each one and decides whether to approve the loan.[10]

# 3 Shocks and Current Loan Outcomes

This section investigates the effect of exogenous shocks on current loan outcomes. I find that shocks increase the probability of default and lower the loan score reported by the BAC to credit bureaus and financial authorities. I start by presenting the data sources and the construction of the variables. I then present the empirical strategy and discuss the results.

## 3.1 Data and Variables

One of the main objectives of this paper is to document the effect of transitory and exogenous shocks on repayment, credit scores, and access to credit. This requires precise measures of the shocks experienced by producers and linked to individual loan data. The combination of the data sets that I use in this paper is unique and allows me to fulfill this strong requirement

[9]The FAG is administered by the Fondo para el Financiamiento del Sector Agropecuario (FINAGRO) [Fund for the Financing of the Agricultural Sector] – a public institution created to promote agriculture in Colombia. FINAGRO lends "rediscount" funds to other banks like the BAC, which use them to lend to farmers. It is thus an example of a second-tier bank that lends to first-tier banks. In 2013, the BAC allocated 85% of FINAGRO's "rediscount" resources (DNP and FINAGRO, 2014).

[10]Farmers with no previous record in the financial system start directly with the analysis stage.

for the following reasons. First, I have access to information from a large number of rainfall stations with time series spanning more than 30 years. The long time series allow me to construct precise measures of exogenous shocks with important variation across space and time. Second, the data on coffee plots provides geographical coordinates of farmers that, together with the weather data, allow me to measure with precision the shocks experienced by the producer at different points in time. Third, the granular data from the BAC allows me to measure loan demand, default, and credit scores. The combination of these three data sets gives me a unique opportunity to document how shocks interact with credit scores and credit access.

More in detail, the analysis of this section draws on three data sources:

(a) Administrative records 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. Importantly, it contains the geographical coordinates of the farms. I have data for 2006–2014.
(b) Precipitation data from the Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia (IDEAM) [Institute of Hydrology, Meteorology and Environmental Studies]. This data set contains monthly rainfall data from more than 1,200 rainfall stations from 1982 to 2016.
(c) Administrative records from the BAC, which contain information on individual loans, their characteristics like interest rate, maturity, monthly overdues, and a loan score.

I use these data sets to obtain a sample at the loan level, which I refer to as the *Loan Sample*. To do so, I find the loans in the Banco Agrario data of the SICA coffee farmers. I keep only loans for coffee production that I can link to a farm of 10 hectares or less, observed in the SICA in the year the loan was disbursed.[11] I then use the coordinates of the corresponding farm to link the loan to the closest rainfall station according to the Euclidean distance.[12] The resulting Loan Sample consists of about 242,000 loans disbursed to 129,000 farmers in 2005–2011.[13] For each loan in the Loan Sample there is associated a farm and the closest rainfall station. The sample contains four sets of variables with information on a)

[11] The mechanisms I describe in this paper might not apply to large producers, who can smooth shocks more easily. Therefore, I keep only small and medium size coffee farmers who, according to the FNC, have farm extensions of up to 10 hectares.

[12] In the Data Appendix I describe the variables and the construction of the sample in more detail.

[13] Although I have information on loans disbursed in 2005–2015, to construct loan default outcomes I need to observe repayment behavior in subsequent years. Since an important part of the analysis below uses long-term loans (five years or more) I restrict the sample to loans disbursed no later than 2011 in order to observe repayment at least five years after the loan was disbursed.

the characteristics of the farm, b) the shocks the coffee producer experienced while the loan was outstanding, c) loan characteristics, and d) loan outcomes. In the following paragraphs, I explain the variables in each set.

*Farm Characteristics*: I have information on the distance to the closest rainfall station and, from the SICA data, on the size of the farm (in hectares). For the coffee plot of the farm, the SICA also contains information on its size (in hectares), its age, and on three important determinants of coffee productivity (Gast et al., 2013): density (number of coffee trees in thousands per hectare), percent shaded area (fraction of the farm with bigger trees that provide shade to the coffee trees, which is beneficial for coffee production), and tree variety (percentage cultivated with the Caturra variety, which is the most common coffee variety in the country and is less productive than other varieties).

*Weather Shocks*: I obtain a measure of the weather shocks that the coffee producer experienced while the loan was outstanding, using the information of the closest rainfall station. I start with the monthly precipitation data at the rainfall station level and calculate total rainfall for each quarter-year from 1982 to 2016. I obtain a separate distribution of total rainfall for each of the four quarters of the year and identify the corresponding 80th percentile.[14] For a given rainfall station and a given quarter-year, I say that a *rainfall shock* occurred if rainfall was above the 80th percentile of the rainfall distribution of the corresponding quarter.[15] This definition takes into account seasonality at the station and quarter level: a shock occurs if, in a given quarter-year, rainfall is particularly high compared to the historical raininess recorded at that station during that quarter.[16] I define an additional measure of rainfall shock using the 90th percentile as the threshold. My preferred rainfall shock variable for a given loan is the number of shocks in the first year after loan disbursement. Therefore the shock variable can take a value of 0, 1, 2, 3, or 4. I focus on shocks during the first year of the loan's life cycle for simplicity and because I observe loans of different maturities, sometimes of less than one year.[17]

*Loan Characteristics*: for each loan, I observe the date of disbursement, size (in Colombian pesos), interest rate, maturity, loan program, loan type, and loan line. The loan program

[14]For each station, I have 35 observations per quarter, that is, 35 observations for the first quarter of the year (January, February, and March), 35 for the second quarter and so forth.

[15]I focus on excessive rainfall shocks since coffee growing in Colombia is more sensitive to periods of excessive rain than to periods of lack of rain. See Gast et al. (2013), Boucher and Moya (2015), and Turbay et al. (2014).

[16]Jayachandran (2006) and Kaur (2019) define shocks at the district-year level similarly, using the historical rainfall distribution and the 80th percentile to define excessive rainfall shocks. As Kaur (2019) points out, this definition captures the non-linear relationship between rainfall and crop productivity.

[17]For example, considering shocks in the second year of the loan tenure would require to leave one year maturity loans out of the sample.

indicates whether the loan is administered by a special government program. The loan line and type are general classifications used by the bank to categorize loans in groups with similar characteristics. I discuss in more detail the loan lines and types below.

*Loan Outcomes*: for each loan, I have monthly information on loan repayment. In particular, I observe the number of days the customer was late with her payments (overdue days). I also observe the *Loan Score*, which tracks the producer's standing with the bank and that is reported each month to credit bureaus and financial authorities; it is based on loan repayment. Each loan starts with a Loan Score of A and can fall to B, C, D, or E. To study repayment behavior at the loan level, I use monthly overdue days to construct a default indicator. My preferred outcome is a dummy equal to one for loans that were *at any point in time* 30 days (or more) past due. I define an analogous dummy for 60 days past due. I also define two measures of loan performance based on the Loan Score: a dummy equal to one if the score ever fell from A and a dummy equal to one if the score fell to E, the lowest score possible.

Table 1 presents summary statistics of the variables in the Loan Sample. Panel A presents statistics of the characteristics of the farm and the coffee plot linked to the loan. The average distance to the rainfall station is 6.4 km (median 5.9 km).[18] The average farm size is 2.8 hectares, 1.6 of which are cultivated with coffee. The coffee trees are 6.5 years old on average and there are about 5,000 coffee trees per hectare. On average 12% of the coffee plantations are shaded and 60% are of the Caturra variety.

Panel B of Table 1 presents statistics of the weather shocks in the Loan Sample. The average number of shocks in the first year of the loan tenure is 1.5, and at least 75% of the loans had two shocks in that year. This finding is likely due to three reasons. First, the shocks I consider are fairly common, since they correspond to rainfall above the 80th percentile in a given quarter. Second, the study period was particularly rainy: it coincided with the "La Niña" climatic phenomenon in 2010 and 2011.[19] Third, since climate change

[18]Panel A of Appendix Figure A1 plots the distribution of coffee farms in the SICA across Colombia for the years 2006–2014. Panel B shows the distribution of rainfall stations in the IDEAM data. The figure shows that the distribution of rainfall stations is dense in the country's coffee-growing areas, which suggests that the closest rainfall station provides a good measure of the extreme weather events that coffee producers face.

[19]La Niña [The Girl] is an ocean-atmospheric phenomenon generated by below-average temperatures of the Eastern-Central Pacific Ocean that leads to higher levels of rain and cloudiness in some South American countries including Colombia, Ecuador, and Peru (Gast et al., 2013). Its counterpart, "El Niño" [The Boy], is associated with above-average temperatures and has the opposite effect on rainfall. The occurrence of La Niña during the study period likely leads to additional variation that is useful for estimating the effect of rainfall shocks. A potential concern is that the recovery results presented below do not apply to periods with no Niña. However, this is unlikely to be the case, since if anything, La Niña leads to more severe weather.

increases the occurrence of periods of atypical weather (Trenberth, 2006), such events are more likely to be concentrated at the end of the time series, which corresponds to my period of analysis. The fact that weather extreme events are becoming more frequent because of climate change enhances the relevance of my results.

To get a sense of the spatial and temporal variation in my shocks measure, consider the "Zona Cafetera" [The Coffee Zone] a geographical region of Colombia that is well known for producing coffee. Appendix Figure A2 plots the Sica farms (in blue) and the IDEAM rainfall stations (in red) in the region. Black lines represent municipality borders. The figure shows that there are multiple rainfall stations in the region and that coffee farms are close to different rainfall stations. Appendix Figure A3 depicts the shocks that occurred in the Zona Cafetera for each quarter-year between 2008-q1 and 2010-q4. Red triangles represent stations that experienced a shock and blue circles represent stations without one. Two points are worth highlighting. First, a given rainfall station experiences shocks in some quarters and not in others. Second, in many quarter-years some stations experience shocks while others that are close in distance do not. Therefore, this figure suggests that there is enough variation both in the cross-section and in time to identify the effect of the shocks.

Panel C of Table 1 presents summary statistics on the loan characteristics: the average value disbursed is 3.4 million pesos of 2010 (US$ 1,790 at the 2010 average daily exchange rate, equal to 1,899 Colombian pesos per US$) at a yearly interest rate of 10% and maturity of 3.3 years. 9% of the loans were from special government programs and 98% were agricultural loans. Over half of the loans were to finance working capital like fertilizer (Line 1), 22% were to replace old coffee trees (Line 2), 16% to invest in new coffee trees (Line 3), and 6% to invest in equipment and commercialization (Line 4). Together, these four loan lines constitute 99% of the loans in my sample. Summary statistics of the loan outcomes are presented in Panel D. 15% of the loans were at least 30 days overdue, and 11% were at least 60 days late. Overdue payments translate into changes in the Loan Score: 18% of the loans in the sample had a score that fell below A; for 9% of the loans, it fell to the worst possible score, E. Yet most loans (more than 82%) were never overdue or had a score lower than A.

## 3.2 Empirical Strategy

I estimate by ordinary least squares (OLS) the following linear-probability model in the Loan Sample, where each observation corresponds to a loan:

$$y_{i,j,m,t} = \beta Shocks_{j,t} + X_i + \mu_{m,t} + \delta_j + \epsilon_{i,j,m,t} \tag{1}$$

$y_{i,j,m,t}$ is a measure of default of loan $i$ (for example a dummy equal to one if the loan was 30 days past due at any point in time). $j$ denotes the rainfall station linked to the loan and $t$ the quarter-year of the loan disbursement. I divide loans into two groups according to their maturity: loans with maturities below 3 years (and denote them by $m = 0$) and loans with maturities above three years ($m = 1$). $Shocks_{j,t}$ is the number of shocks in the first year of the loan.[20] As discussed previously and for simplicity, I focus on the effect of rainfall shocks during the first year of the loan. $\beta$ captures the effect of an additional rainfall shock in the probability that the loan defaults and is expected to be positive. $X_i$ is a vector of predetermined variables at the loan level that include farm and loan characteristics (listed in panels A and C of Table 1). The term $\mu_{m,t}$ denotes a fixed effect for loans with maturities above three years ($m = 1$) interacted with quarter-year of disbursement fixed effects. $\delta_j$ denotes rainfall station fixed effects and $\epsilon_{i,j,m,t}$ is a mean-zero error term. I cluster the error term at the municipality level when estimating equation (1).[21]

The inclusion of $\mu_{m,t}$ and $\delta_j$ implies that the coefficient $\beta$ is identified based on variation in $Shocks_{j,t}$ over time within the same rainfall station, for loans of similar maturity. Therefore $\beta$ 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 the repayment of outstanding loans. Formally, I assume that:

$$E[\epsilon_{i,j,m,t} | Shocks_{j,t}, X_i, \mu_{m,t}, \delta_j] = 0 \tag{2}$$

This assumption is reasonable given my definition of rainfall shocks. $Shocks_{j,t}$ captures atypical rainfall levels for a given year at a particular station. This variation is likely to be uncorrelated with other time-varying factors that affect repayment. To study the plausibility of this assumption, I conduct a covariate balance test. I estimate equation (1) using as the left-hand variable the elements of $X_i$ (loan and farm characteristics that are predetermined when the loan is disbursed, listed in Panels A and C of Table 1). For ease of presentation, before running the regression I standardize both the left-hand variable and $Shocks_{j,t}$ by subtracting their sample mean and dividing by their standard deviation. Figure 1 plots the estimates of $\beta$ with 90% confidence intervals. Out of 15 covariates, 9 are precise zeros, three coefficients are marginally significant (Coffe Plot Density, Loan Line 1, and Loan Line 3)

[20]For simplicity, and since I define rainfall shocks using quarter-years, I also handle other dates at the quarter-year level.

[21]Within most municipalities in Colombia, geographic and climatic conditions are similar (IGAC, 2004). Clustering at the municipality level allows for correlation of the error term across observations associated with different rainfall stations but located in places with similar weather.

and three are significant (interest rate, and the dummies for special programs and Loan Line 4). Still, the magnitude of these three estimates is small, ranging from -0.023 to 0.015 standard deviations for a one-standard-deviation increase in $Shocks_{j,t}$. The results from Figure 1 suggest that rainfall shocks are not systematically correlated with loan- and farm-level characteristics. Still, I include controls for predetermined characteristics in my main specifications. In Section 3.3, I also present estimates of equation (1) omitting the set of predetermined controls $X_i$ and the results remain practically unchanged, which provides additional support for the identification assumption in equation (2).

There are two additional concerns regarding the identification assumption. First, it is possible that weather shocks affect the prices paid to coffee producers. But in Colombia the FNC regulates the minimum price for "parchment" (dried) coffee sold by farmers. This price closely follows the international price and is the same across the country (see for e.g. Macchiavello and Florensa-Miquel, 2019). Second, rainfall shocks can also affect the wages paid to the coffee pickers who pick the coffee beans from the trees at the time of the harvest. This can in turn affect the profits of the coffee farmer. But in the event of a rainfall shock that reduces coffee production, the demand for the labor of coffee pickers will be lower and their wages will fall, increasing profits. Therefore, my estimate of $\beta$ in equation (1) is a lower bound for the parameter that I would obtain if I was able to control for the wage level.

### 3.3 Results

In this section, I present and discuss the results of the effect of shocks on current loan outcomes. I find that shocks increase the probability of default and lower the Loan Score reported to credit bureaus and financial authorities.

Table 2 presents the results from estimating different versions of equation (1). In Panel A, the outcome is a measure of default. Panel B uses outcomes related to the Loan Score. Each column of each panel corresponds to a different regression. I include predetermined controls at the time of loan disbursement in all specifications (variables in Panels A and C of Table 1). The first column of Panel A shows the estimated effect of the number of shocks in the first year of the loan on the probability of being 30 days past due. Each additional shock increases this probability by 0.004 points or about 3% of the sample mean (of 0.15). This implies that loans with two or more shocks in their first year (about 25% of the loans in the sample) were 6% more likely to be overdue compared to the average loan. Column (2) of Panel A shows that the effect is similar if I use the 90th percentile of the rainfall distribution to define the shock. If I use a dummy equal to 1 for loans that had two or more shocks as my measure of weather variability, I find a similar effect (column (3)). Column (4) uses a dummy for loans that were 60 days or more overdue as the outcome and the estimated effect is the same as that of column (1). If I restrict the sample to loans with a distance to the rainfall station below the 75th percentile, or if I look only at loans with short maturities, the effect remains unchanged (columns (5) and (6), respectively). Column (7) reports the effect of the number of rainfall shocks for the sample of loans with maturities over 3 years. Although the effect is non-significant, the point estimate is similar to the one in column (1).

The results of Panel A show that shocks lead to lower rates of repayment. But does this translate into negative credit reports? Panel B of Table 2 shows the effect of shocks on the Loan Score reported by the BAC to credit bureaus. The main outcome is a dummy equal to 1 for loans with scores that fell from the initial standard A rating. Columns (1) to (3) show that irrespective of how the shock is defined, a higher number of shocks increases the probability that the Loan Score falls. The outcome of column (4) is a dummy for loans that had a score of E (the lowest possible) at any point in time. Column (5) presents the estimates of the effect on the Loan Score for farms close to the rainfall station. Columns (6) and (7) restrict the sample to loans of short and long maturities, respectively. In all cases, more shocks in the first year increase the probability that the Loan Score falls. This shows that exogenous and transitory shocks lead to negative information being reported to credit bureaus.

Appendix Table A1 reports the estimates of equation (1) without the inclusion of predetermined controls. The results are very similar to those in Table 2. For Panel A the results are practically identical. In Panel B the effect of the shock variable on the probability that the Loan Score will decrease is non-significant in columns (2), (4), and (6), but the point estimate is still positive. The finding that the inclusion of predetermined controls does not seem to substantially affect the estimates of equation (1) gives support to the identifying assumption.

# 4 Shocks and Future Loan Applications

In this section, I estimate the effect of shocks on loan application outcomes. I find that producers who experience a transitory and exogenous shock during a loan tenure have lower Bureau Scores and a higher probability of rejection in their next loan application. Using a panel of coffee producers, I show that these effects can last up to five years.

## 4.1 The Effect of Shocks on the Next Application

I use administrative records from the BAC to obtain a sample of applications. Recall from Section 2 that when the coffee producer applies for a loan in a bank branch, a bank employee enquires about his standing with the credit bureau. This query produces a report with a *Bureau Score* and, based on BAC policies, it also indicates if the application process can continue. The reports are available from 2010 to 2015 and I refer to them as the Bureau Reports. To estimate the effect of shocks on applications, I obtain the *Application Sample* based on the information in the Bureau Reports.

To obtain this sample I proceed in the following way. For each coffee producer in the Loan Sample (the sample that I used in the previous section) I keep only loans with a date of disbursement of 2008 or later and take the most recent loan. I then find the first application after the end of that loan. This leaves me with one application per farmer that is preceded by a previous loan.[22] I am interested in the effect of shocks that happened *during the first year of the initial loan* on the Bureau Score and the probability of rejection of this "follow-up" application. Panel E of Table 1 shows summary statistics of the application outcomes in the Application Sample. For 70% of the initial loans, I find a subsequent application, which happens on average 1.25 years after the end of the first loan. The average Bureau Score is 935, and 10% of the loans were denied at this stage.[23]

I then estimate the same equation of the previous section (equation (1)) but on the Application Sample. The variable of interest is the number of shocks in the first year of the loan and the outcomes are the Bureau Score of the subsequent application or a dummy equal to one if the application was denied. Table 3 presents the results. In column (1) the outcome is a dummy equal to 1 if the coffee producer applied for a new loan after the initial one. Therefore, the estimated coefficient is the effect of shocks on the probability of applying for a new loan. The estimate is negative and marginally significant, which implies that shocks during the first loan reduce the probability that the producer applies for a new one.[24] Since shocks have a negative effect on repayment and applying for a new loan entails some costs, farmers who experience a shock and still apply for a new loan are positively

[22] I start the search for the next application with loans disbursed in 2008 or after because the Bureau Reports start in 2010. In addition, in 2008 Colombia introduced a law that requires negative information (past defaults) to be erased from credit bureau data sets after four years. Note that this law does *not* imply that the effect of shocks on credit access can last up to four years. For example, suppose a farmer is hit by a shock in year 1, which causes overdue payments up to year 3. The record of these overdues can then last up to year 7 because of the law. In this example, the shock affects credit access for seven years.

[23] The average Bureau Score comes from about 26,500 loans and is not representative of all BAC loans.

[24] The estimated effect is small, representing only a decrease of around 1% of the sample mean of the probability of applying for a new loan.

selected. Therefore, this selection works against finding a negative effect of shocks on the credit score of the next application or a positive effect on its probability of denial.

Column (2) reports the effect of shocks on the probability of being 30 days overdue in the first loan.[25] This is equivalent to column (1) of Table 2, but in this case, the sample of loans corresponds to those that are followed by a subsequent application. Again, rainfall shocks cause worse repayment outcomes. The estimated effect is much larger than in Table 2. In this sample, each additional shock increases the probability of being 30 days overdue by 12% relative to the mean, which is considerably larger than the 3% increase reported in Table 2. Column (3) of Table 3 shows that each additional shock leads to a decrease in the Bureau Score of about 3.4 points in the next application. Relative to the mean, this is effect is not large (since the mean is close to 1,000) but it still represents 3% of a standard deviation of the outcome. Finally, column (4) shows the effect of shocks on a dummy for application denial. Each additional shock in the first year of the initial loan leads to an increase in the probability of denial of around 0.01. This effect is large: it corresponds to a 7.5% increase in the probability of denial relative to the sample mean.

The results reported in Table 3 indicate that shocks lead to a lower Bureau Score and to a higher probability that future applications will be rejected. In other words, transitory and exogenous shocks exclude coffee producers from future access to credit.

## 4.2 The Persistent Effect of Transitory Shocks

In this section, I study the persistence of the effect of shocks on future loan applications. Using a panel at the individual-year level, I show that the effect of transitory shocks on scores and access to credit lasts at least five years.

I use the Bureau Reports data of 2010–2015 described in the previous section to obtain a panel at the individual-year level with two outcomes: 1) the average Bureau Score (for each individual I average across all the observed reports in year $t$) and 2) a dummy equal to 1 if any of the applications in $t$ were rejected.[26] I use OLS to estimate the following model:

$$y_{i,j,t} = \theta_i + \eta_t + \sum_{k=0}^{5} \gamma_k \, Shocks_{j,t-k} + u_{i,j,t} \tag{3}$$

where $y_{i,j,t}$ is the average Bureau Score of individual $i$ in year $t$ or a dummy equal to 1 if any

[25]Note that the number of observations falls from column (1) to column (2). This is due to the fact that not all the coffee producers in the sample of column (1) apply for a new loan. The sample of columns (2) to (4) consists only of farmers who apply for a second loan, after the initial one.

[26]The panel includes only coffee producers in the Loan Sample. See the Data Appendix for details.

of individual $i$'s applications were rejected in year $t$. As before, $j$ denotes rainfall stations and $Shocks_{j,t}$ the number of shocks in rainfall station $j$ in year $t$. $\theta_i$ and $\eta_t$ denote individual and year fixed effects, respectively. $u_{i,j,t}$ is a mean zero-error term, which I cluster at the municipality level.[27]

Since I only observe the producer if she applies for a loan, the resulting panel is unbalanced, and selection is a concern. But this concern is not a big one, for the following reasons. First, in the previous section, I showed that the probability of applying for a new loan is negatively affected by shocks. Therefore producers that remain in the sample after a shock are positively selected so that selection works against finding a negative effect of shocks on the Bureau Score and against finding a positive effect on the probability of loan denial. Second, in the results reported below, I present estimates for samples that differ in the number of years that each individual is observed, and the results are similar. Third, if selection is fixed at the individual level, it is captured by the individual fixed effect $\theta_i$ of equation (3).

Table 4 presents the results of estimating equation (3). Columns (1) and (2) use a sample with at least two observations per individual in the years 2010–2015. Columns (3) and (4) use a sample of individuals that I observe in all six years. Column (1) shows that shocks have a negative and persistent effect on the Bureau Score. With the exception of lag $t-4$, all lags are statistically significant. In particular, the result for lag $t-5$ implies that shocks affect credit scores even five years later. Column (2) shows the effect on the probability that at least one application is rejected in year $t$. This effect closely follows the impact on the Bureau Score. Exogenous shocks increase the probability of rejection in lags $t-1$, $t-2$, $t-3$ and $t-5$ which implies that the effect is persistent. Importantly, the effect of $t-5$ is sizable: each additional shock in $t-5$ increases this probability by 0.002 points, equal to 3.4% of the sample mean. Columns (3) and (4) show the same estimates but restricting the sample to individuals who applied for a loan every year from 2010 to 2015. The results are very similar to those of columns (1) and (2). This suggests that selection is not an important concern in this exercise. Furthermore, the effects on loan denial of column (4) are larger than those reported in column (2). For example, each additional shock in lag $t-5$ increases the probability of denial in $t$ in 0.003 points, which equals 8.1% of the sample mean. Since these results are obtained from an estimation that includes individual fixed effects, they are identified from within-individual variation. Therefore, holding borrower type fixed, shocks exclude producers from credit markets for a period of time of at least five years.

[27]Note that equation (3) is unconditional on whether the farmer has an outstanding loan in year $t$, unlike equation (1) in the previous section.

An alternative way to study persistence is to consider the sample of loans that are followed by a subsequent application. In Section 4.1, I showed that rainfall shocks during the first year of a loan decrease the Bureau Score and increase the probability that a future loan application will be rejected (Table 3). In Appendix Table A2, I estimate the same regression but restricting the sample according to the amount of time that elapsed between the end of the initial loan and the next application. Columns (1) and (2) report the effects of rainfall shocks experienced during the first loan on the Bureau Score and the probability that the following loan is rejected, with a time-lapse of one year or more. The results show that shocks decrease the Bureau Score and increase rejection rates. For a time-lapse of two years (or more) and three years (or more), the effect on the Bureau Score is negative and significant (columns (3) and (5)) but it is not significant on the probability of denial (columns (4) and (6)). Still, the point estimate is positive and large in the case of column (4) (although imprecisely measured). Therefore, in this sample, the effect on the Bureau Score can last at least three years. The effect on denial can last at least two years.

Overall, the results of this section show that exogenous and transitory shocks have a persistent effect on credit bureau scores and on the probability of loan denial. The estimates from the panel show that these effects can last for at least five years.

# 5 The Recovery from Shocks

In this section, I study the recovery of income and repayment rates of coffee producers after a shock. First, I show that shocks in the year preceding the harvest have important negative effects on farmers' incomes but shocks older than one year and a half do not have a robust negative effect. It follows that producers' incomes recover one year and a half after the shock. Furthermore, I use data on long-term loans to measure default over a long period of time and show that in years 4 and 5 after a shock, there is no effect on repayment. This result implies that repayment has completely recovered when credit scores and credit access are still suppressed (according to the results of Section 4.2).

## 5.1 Income Recovery

Excessive rainfall in the year before a harvest affects the productivity of coffee trees, but if the weather returns to normal the next harvest is not affected (Arcila et al., 2007; Gast et al., 2013). This suggests that shocks only have a short-term effect on agricultural income as long as they are not catastrophic.

I test this idea using data from the Colombian Agricultural Census which was administered by DANE and contains information on the quantities produced in the harvest of 2013 by all farms in the country. I use a price of 3,730 pesos per kilo of coffee to measure income from coffee production.[28] In the Data Appendix, I describe in detail the cleaning of the data. My estimation sample consists of about 175,000 farms that produced coffee in 2013. As before, I restrict the sample to small and medium-sized producers (up to 10 ha). Appendix Table A3 shows the summary statistics of the census sample of coffee producers. The average farm is 2.5 ha and the average distance to the rainfall station is 6.5 km, which is close to the average distance of the SICA farms linked to the BAC loan data of the previous sections. The mean income from coffee production in the 2013 harvest was 3.9 million pesos (US$ 2,090 at the 2013 average daily exchange rate).

I estimate by OLS the following equation:

$$r_{i,j} = \alpha_0 + \sum_{k=-4}^{12} \alpha_k \, Shock_{j,(2013\ q1-k)} + Z_i + \phi_d + u_{i,j} \tag{4}$$

where $r_{i,j}$ is farm $i$'s coffee income (in millions of pesos) from the 2013 harvest. As before, $j$ denotes the rainfall station closest to producer $i$. The variables of interest, denoted by $Shock_{j,(2013\ q1-k)}$, are dummies equal to 1 if there was a shock in rainfall station $j$ in quarter 2013 $q1 - k$, where $k$ is a lead and lag indicator in relation to the quarter 2013-q1.[29] I consider shocks in 17 quarters, 12 before the harvest (2010-q1 to 2012-q4) and 5 during or after the harvest (2013-q1 to 2014-q4). $\alpha_0$ is a constant term and $Z_i$ is a vector of farm-level controls. Finally, $\phi_d$ denotes department fixed effects and $u_{i,j}$ is a mean zero-error term (that I cluster at the municipality level).[30] In this case the identification assumption requires $u_{i,j}$ to be uncorrelated with the $Shock_{j,(2013\ q1-k)}$ variables. This condition is stronger than the assumption in equation (2) since it requires the shock terms to be uncorrelated with the error term in the cross-section. Below I show that the results do not change importantly with the set of controls, which gives support to the assumption. Furthermore, the inclusion

[28] The FNC regulates the minimum price for "parchment" (dried) coffee sold by farmers in Colombia. This price closely follows the international price and is the same across the country. I take the simple average of the daily price in 2013 to obtain the price of 3,730 pesos per kilo.

[29] In regressions not reported here, when I aggregate quarterly shocks to the yearly level I find no effects on income. This is perhaps due to noisier estimations as a result of the aggregation.

[30] The department is the largest administrative unit in Colombia. I follow Muñoz-Mora (2016), who controls for coffee region fixed effects in some of his estimations. There are four coffee regions in Colombia with broadly similar characteristics, like altitude and soil quality, that affect coffee production. In Muñoz-Mora (2016) coffee regions correspond to groups of departments, so the department fixed effects of equation (4) absorb the coffee region fixed effects.

of leads in equation (4) constitutes a placebo test for its plausibility.

In the baseline exercises, $Z_i$ includes farm size and different characteristics of the farm household.[31] I use dummies to denote the educational attainment of the household head as well as households with access to the following: health insurance; electricity, sewerage, and an aqueduct; high-quality walls; and high-quality floors.[32] These controls, although in principle predetermined at the time of the shocks, are arguably less exogenous than farm size.

Figure 2 plots the estimated coefficients $\alpha_k$ with 90% confidence intervals. The dashed line indicates quarter $k = 0$ (2013-q1, the first quarter of the year of the harvest). The main feature of the figure is that shocks two or three quarters before the harvest (lags 2 and 3) have a large and negative effect on income. The effect of lag 1 is negative but marginally significant (p = 0.106). The point estimates of lags 2 and 3 are -1.23 and -0.85, respectively. It follows that those producers who experienced atypically high rainfall levels two quarters before the harvest saw their coffee income decrease by 32%. If the shock was experienced three quarters before the harvest, income decreased by 22%. The second main feature of the figure is that besides the coefficients of lags 1, 2, and 3, no other coefficient is negative and significant. The only exception is the coefficient of lag 6, which equals -0.56 (although this result is not robust to the set of controls, as discussed below). Shocks older than one year and a half (lag 7 onward) have no negative effect on income, which indicates that income recovers from shocks during that time. The third feature of the figure that stands out is that the estimated coefficients for the leads are close to zero and statically insignificant. The results from this placebo test lend support to the assumption that the shocks are exogenous.

Appendix Figure A4 presents the same figure but with a different set of controls. Panel A includes no controls, and Panel B only controls for farm size. Both graphs are similar to Figure 2 and the estimated effect of lags 2 and 3 are similar in magnitude. The effect of lag 6 is close to 0 and non-significant in Panel A. Therefore, according to this figure, only shocks in lags 2 and 3 have a negative effect on income.

In conclusion, Figure 2 and Appendix Figure A4 show that shocks that happen at most three quarters before the harvest have a negative and significant effect on coffee producers' income. This finding is consistent with the one reported in Section 3.3 – that shocks in the first year of the loan increase the probability of default. Furthermore, shocks older than one and a half years do not have a negative effect on current income. Therefore, income recovers

[31]Some farms have many households. See the Data Appendix for details on how I handle this issue.

[32]Floor and wall quality are frequently used as proxies of household wealth. See for example Camacho and Conover (2011).

from transitory and exogenous shocks much faster than credit bureau scores or credit access, which are affected for at least five years.

### 5.2 Repayment Recovery

It is difficult to determine whether a coffee producer would have been able to repay a loan she was denied because of an exogenous shock. As documented previously, rainfall shocks cause higher rates of rejection of subsequent loan applications, so the sample of farmers who get their next loan approved is selected. Furthermore, I do not observe the repayment rates of farmers who are denied a loan. To get around this problem, I use long-term loans (with maturities of at least five years) to proxy for the evolution of repayment over a long period of time and I estimate the effect of shocks during the first year on repayment in subsequent years. I find that shocks increase the probability of default in years 2 and 3 but that this effect decreases in year 4 and dies out in year 5. This implies that repayment has fully recovered five years after the shock.

To obtain the sample for this analysis, I start with the Loan Sample of Section 3.3 and keep only loans with maturities of five years or more. Long-term loans are used mostly to plant new coffee trees of better varieties. Coffee producers generally have many plots with different characteristics on the same farm. Although the plots associated with the long-term loan are used to plant new coffee trees (which take at least two years to produce coffee and generate income), the farmer can repay the loan with income from the other plots. Therefore, it is sensible to assume that her ability to repay the long-term loan is based on the income from the whole farm and not only the plots with new trees. Appendix Table A4 presents the summary statistics of farms linked to long-term loans. The median number of plots is 3 and the median age is 4.5 years. To make sure that the farm has many plots and that income is generated even if some of the plots only have new trees, in the baseline exercises of this section I only include loans linked to farms with a number of plots and a mean age above the corresponding median. I refer to this sample as the *Baseline Long-Term Loan Sample*.[33] I estimate using OLS the following equation:

$$y^k_{i,j,t} = \beta_k \; Shocks_{j,t} + X_i + \psi_t + \iota_j + \nu_{k,i,j,t} \tag{5}$$

where $i$ indexes loans, $j$ the rainfall station linked to the loan, and $t$ the quarter-year of the

[33]Results without this restriction on the sample are reported in the appendix and I refer to that sample as the *Full Long-Term Loan Sample*.

loan disbursement. $y^k_{i,j,t}$ is a dummy equal to one if the loan was 30 days past due (or more) during its year $k$. I estimate equation 5 separately for $k \in \{1, 2, 3, 4, 5\}$. For example, in the case $k = 3$, the outcome is $y^3_{i,j,t}$, a dummy equal to 1 for loans that were overdue 30 days (or more) during their third year. Equation (5) is identical to equation (1) of Section 3.3 but the outcome depends on the year where the loan can be overdue. $Shocks_{j,t}$ denotes the number of shocks experienced by the coffee producer during *the first year* of the loan. $\psi_t$ and $\iota_j$ denote quarter-year of disbursement and rainfall station fixed effects respectively. $\nu_{k,i,j,t}$ is a mean-zero error term that I cluster at the municipality level.

Some of the loans in the estimation sample were restructured by the BAC, and the rights to some delinquent loans were purchased by the government through the Fondo de Solidaridad Agropecuario (FONSA) [Fund for Agricultural Solidarity].[34] Since I do not observe future repayment for either restructured or FONSA loans, I estimate equation (5) in three different ways. First, I ignore the problem and estimate it on the full sample. Second, I control for loans using a dummy equal to one for restructured or FONSA loans. Finally, I drop restructured and FONSA loans.[35]

Table 5 shows statistics of the "paths" of the loans depending on the years with overdue payments, for different samples. Each line corresponds to a different path. For example, the first line (path 17) consists of loans that were 30 days past due in years 2 to 5 but not in year 1.[36] Columns (1) and (2) show the number and frequency of loans in each path of the Baseline Long-Term Loan Sample. There are 24,513 loans, 83% of which were never overdue; 3,932 (16.5%) were overdue by at least one year. Of these, 256 (about 1%) followed path #24 with overdues only in year 2. This path is consistent with repayment recovery in the sense that its loans were overdue at some point in time but were not at the end (in year 5). Seven paths (denoted by an even number in the table, except for #32) are consistent with repayment recovery. These paths account for 7.9% of the loans in this sample and for 48%

[34]Loan restructuring occurs when the BAC agrees with the client to start a new obligation that incorporates previous obligations. The new obligation usually has different conditions than the previous one (for example, a different payment schedule). Loan restructuring alternatives are sometimes offered to clients who are overdue or claim they cannot pay. Out of 105,229 observations in the sample of long-term loans, 10,552 (roughly 10%) were restructured by the BAC or purchased by FONSA.

[35]It is possible that shocks cause loans to be restructured or purchased by FONSA. Appendix Table A5 shows results of estimating equation (5) using as an outcome a dummy for loans that were restructured and a dummy for loans that were purchased by FONSA. Only for the sample with all the farms do shocks increase the probability of a FONSA purchase (column (2) of Appendix Table A5). But the effect is small and only marginally significant. These results suggest that selection into restructuring or FONSA is not an important concern.

[36]I omit from the table paths 1 to 16, which start with an overdue in year 1. Only 87 loans in the Baseline Long-Term Loan Sample correspond to these paths since the vast majority of long-term loans have a grace period in the first year.

of the loans that were overdue at some point. Therefore, at least half of all overdue loans follow a path that is consistent with repayment recovery. Although the number of loans in recovery paths decreases if I omit from the sample restructured and FONSA loans (columns (3) and (4)) the number is still large (5.2% of all loans and 48% of overdue loans). Columns (5) to (8) contain statistics of the paths of the Full Long-Term Loan Sample (that is, without the restriction on the number of plots and the age of the farm). Again, the number of loans in recovery paths is large. Now I turn to the results on how shocks in the first year affect these overdue trajectories.

Figure 3 plots the estimated coefficients $\beta_k$ of equation (5) separately for $k \in \{1, 2, 3, 4, 5\}$ in the Baseline Long-term Loan Sample, along with 90% confidence intervals. Panel 1 shows the results without controlling for FONSA or restructured loans. The shock has no effect on overdues in year 1, which is expected since long-term loans have a grace period. Then, the effect increases for year 2 and peaks at year 3 and decreases for years 4 and 5. In year 5 the point estimate is close to 0. The evolution of the effect of the shock shows that repayment recovers in years 4 and 5 from a shock in year 1. Panel 2 presents the results after controlling with a dummy for loans that were either restructured or bought by FONSA. The results are practically unchanged. The effect follows a similar pattern if I drop loans that were restructured or bought by FONSA (the third panel), although in this case, the estimate of $\beta_3$ is only marginally significant at the 10% level. In Appendix Figure A5, I present as a robustness exercise the evolution of $\beta_k$ but obtained from estimating equation (5) with the Full Long-Term Loan Sample. The results are similar to those of Figure 3. Again, in all three cases (that is, independently of how I control for restructured and FONSA loans), shocks have an effect at the beginning of the loan tenure but no effect in year 5. In the third graph of the figure, the effect does not peak in year 3; instead, the point estimate is similar for years 2 to 4.

Two additional pieces of evidence support the relation between income and repayment. First, BAC's repayment schedules are timed to match the farmers' harvests, suggesting that income from farm production is related to loan repayment. Second, although there is little empirical evidence on this matter, Chirwa (1997) finds that small farmers' crop sales are correlated with higher loan repayments in Malawi, and Acquah and Addo (2011) find that higher fishing incomes are correlated with higher repayment of loans to fishermen.

In sum, the results of Sections 5.1 and 5.2 imply that producer's income and repayment recover faster from exogenous and transitory shocks than credit scores and credit access. Recall from Section 4.2 that shocks have a negative effect on the Bureau Score and access to credit even after five years. Since income recovers at most one year and a half after the shock, the effect on scores and credit access far outlasts the impact on income. After a shock, repayment starts to recover in year 4 and has fully recovered in year 5 but credit scores and credit access are still depressed in those years. This shows that some producers who can repay a loan are denied credit during this time.

One important concern remains. I estimated the patterns of income recovery and repayment in samples that are different from the ones that I use to show that shocks lower credit access. So in order to claim that income and repayment recover faster from shocks than credit scores and credit access, I need to assume that the patterns of recovery have some degree of external validity and, in particular, are applicable to the producers in the bank data. This is likely the case since data on income comes from a census that contains information on all agricultural producers in the country. Furthermore, the number of long-term loans in the repayment recovery exercise is large and the results are robust to different ways of defining the sample.

# 6 Conclusion

Using the case of coffee production in Colombia, this paper documented that transitory and exogenous shocks can have persistent effects on access to credit. Furthermore, it showed that income and repayment can recover faster from shocks than credit access. These results imply costs to both the producer and the lender and show that the interplay of exogenous and transitory shocks and credit reporting systems can exclude profitable producers (those who can repay a loan) from credit markets.

These results yield immediate policy implications. For instance, insurance– in particular, agricultural insurance– can mitigate the impact of exogenous shocks on income, repayment, credit scores, and future access to credit. But there are additional policy instruments that may mitigate the negative channels I document. On the one hand, contingency-dependent repayment schemes that link repayment to the occurrence of exogenous shocks could attenuate the scores downgrades. Repayment plans in this spirit are already in place in some settings. For example, reimbursement schemes that link repayment to borrower's earnings have been implemented in the US student loan market (see e.g. Abraham et al., 2020) and the microfinance literature has explored the use of flexible repayment schemes that allow borrowers to postpone payments during the loan cycle (see e.g. Barboni and Agarwal, 2018). On the other hand, credit scores could include information on exogenous shocks constructed, for example, with data on exchange rates, prices, or weather.[37]

Finally, my results could motivate the deletion of negative credit reports. If the reports were caused by transitory and exogenous shocks then their deletion could facilitate access to credit of customers who can repay a loan. In fact, many countries with credit reporting systems also have policies that require reports to be erased after some time. One of their motivations is the observation that individuals can default because of negative shocks out of their control and that the resulting exclusion from credit markets can be unfair (Steinberg, 2014; Liberman et al., 1986).[38] The results of this paper give empirical support to the idea that the deletion of old negative reports can be welfare-enhancing in some circumstances.

[37]The lender and credit bureau analyzed in this study do not do incorporate information on weather shocks in their credit scores. Nor do other banks in Latin America working with the same credit bureau. I am able to measure with precision rainfall shocks at the loan level using a data set with geographical coordinates that is not available to the lender of this study. I can only do this for one crop – coffee. At the time of this study, the bank has no GPS location or a precise address of the farms. For research on how to improve credit scores in other ways see, for example, Rocha Sousa et al. (2016) and Wei et al. (2015).

[38]In a theoretical model of repeated borrowing and lending with moral hazard and adverse selection, Elul and Gottardi (2015) show that under plausible circumstances some level of "forgetting" can be optimal. Bos and Nakamura (2014) show evidence that is consistent with welfare gains of a policy in Sweden that deleted negative reports while Liberman et al. (1986) estimate negative welfare effects of a policy in Chile that erased reports of a large fraction of the population.

# References

Abraham, K. G., Filiz-Ozbay, E., Ozbay, E. Y., and Turner, L. J. (2020). Framing effects, earnings expectations, and the design of student loan repayment schemes. *Journal of Public Economics*, 183:104067.

Acemoglu, D. and Zilibotti, F. (2001). Productivity differences. *The Quarterly Journal of Economics*, 116(2):563–606.

Acquah, H. and Addo, J. (2011). Determinants of loan repayment performance of fishermen: Empirical evidence from ghana. *Cercetari Agronomice in Moldova*, 44(4).

Arcila, J. P., Farfán, J., Moreno, A. B., Salazar, L. F., and Hincapié, E. (2007). Sistemas de producción de café en colombia. Technical report.

Avery, R. B., Calem, P. S., and Canner, G. B. (2004). Consumer credit scoring: Do situational circumstances matter. *Journal of Banking and Finance*, 28:835–856.

Barboni, G. and Agarwal, P. (2018). Knowing what's good for you: Can a repayment flexibility option in microfinance contracts improve repayment rates and business outcomes? *Working Paper*.

Berger, A. N., Frame, W. S., and Miller, N. H. (2005). Credit scoring and the availability, price, and risk of small business credit. *Journal of Money, Credit and Banking*, pages 191–222.

Berman, N., Martin, P., and Mayer, T. (2012). How do different exporters react to exchange rate changes? *The Quarterly Journal of Economics*, 127(1):437–492.

Bos, M., Breza, E., and Liberman, A. (2018). The labor market effects of credit market information. *The Review of Financial Studies*, 31(6):2005–2037.

Bos, M. and Nakamura, L. I. (2014). Should defaults be forgotten? evidence from variation in removal of negative consumer credit information.

Boucher, S. and Moya, A. (2015). Gestión de riesgo: Recomendaciones para el diseño e implementación de un seguro indexado para pequeños productores cafeteros. In Echavarría, J. J., Esguerra, P., McAllister, D., and Robayo, C. F., editors, *Misión de Estudios para la Competitividad de la Caficultura en Colombia*. Fondo Nacional del Café, Bogotá.

Camacho, A. and Conover, E. (2011). Manipulation of social program eligibility. *American Economic Journal: Economic Policy*, 3(2):41–65.

Carter, M. R. and Olinto, P. (2003). Getting institutions "right" for whom? credit constraints and the impact of property rights on the quantity and composition of investment. *American Journal of Agricultural Economics*, 85(1):173–186.

Casaburi, L. and Willis, J. (2018). Time versus state in insurance: Experimental evidence from contract farming in kenya. *American Economic Review*, 108(12):3778–3813.

Castro, C. and Garcia, K. (2014). Default risk in agricultural lending: the effects of commodity price volatility and climate. *Agricultural Finance Review*, 74(4):501–521.

Chirwa, E. W. (1997). An econometric analysis of the determinants of agricultural credit repayment in malawi. *African Review of Money Finance and Banking*, 2(1):107–122.

Cole, S., Giné, X., Tobacman, J., Topalova, P., Townsend, R., and Vickery, J. (2013). Barriers to household risk management: Evidence from india. *American Economic Journal: Applied Economics*, 5(1):104–35.

Cole, S., Giné, X., and Vickery, J. (2017). How does risk management influence production decisions? evidence from a field experiment. *The Review of Financial Studies*, 30(6):1935–1970.

Conning, J. and Udry, C. (2007). Rural financial markets in developing countries. *Handbook of Agricultural Economics*, 3:2857–2908.

de Janvry, A., McIntosh, C., and Sadoulet, E. (2010). The supply-and semand-side impacts of credit market information. *Journal of Development Economics*, 93(2):173–188.

de Luna-Martínez, J. d. and Vicente, C. L. (2012). Global survey of development banks. World Bank, Policy Research Working Paper, 5969.

de Olloqui, F. e. (2013). *Bancos Públicos de Desarrollo, Hacia un Nuevo Paradigma?* Banco Interamericano de Desarrollo.

Dell, M., Jones, B. F., and Olken, B. A. (2014). What do we learn from the weather? the new climate economy literature. *Journal of Economic Literature*, 52(3):740–798.

Djankov, S., McLiesh, C., and Shleifer, A. (2007). Private credit in 129 countries. *Journal of Financial Economics*, 84(2):299–329.

DNP and FINAGRO (2014). Sistema nacional de crédito agropecuario, propuesta de reforma.

Dobbie, W., Goldsmith-Pinkham, P., Mahoney, N., and Song, J. (2020). Bad credit, no problem? credit and labor market consequences of bad credit reports.

Einav, L., Jenkins, M., and Levin, J. (2012). Contract pricing in consumer credit markets. *Econometrica*, 80(4):1387–1432.

Einav, L., Jenkins, M., and Levin, J. (2013). The impact of credit scoring on consumer lending. *The RAND Journal of Economics*, 44(2):249–274.

Elul, R. and Gottardi, P. (2015). Bankruptcy: Is it enough to forgive or must we also forget? *American Economic Journal: Microeconomics*, 7(4):294–338.

Fafchamps, M. (1992). Cash Crop Production, Food Price Volatility, and Rural Market Integration in the Third World. *American Journal of Agricultural Economics*, 74(1):90–99.

Fletschner, D., Guirkinger, C., and Boucher, S. (2010). Risk, credit constraints and financial efficiency in peruvian agriculture. *The Journal of Development Studies*, 46(6):981–1002.

Garmaise, M. J. and Natividad, G. (2017). Consumer default, credit reporting, and borrowing constraints. *The Journal of Finance*, 72(5):2331–2368.

Gast, F., Benavides, P., Sanz, J. R., Herrera, J. C., Ramírez, V. H., Cristancho, M. A., and Marín, S. M. (2013). *Manual del Cafetero Colombiano, Investigación y Tecnología para la Sostenibilidad de la Caficultura*. Federación Nacional de Cafeteros, Cenicafé.

Giné, X., Goldberg, J., and Yang, D. (2012). Credit market consequences of improved personal identification: Field experimental evidence from malawi. *The American Economic Review*, 102(6):2923–2954.

Guirkinger, C. and Boucher, S. R. (2008). Credit constraints and productivity in peruvian agriculture. *Agricultural Economics*, 39(3):295–308.

IGAC (2004). Guía para el deslinde y amojonamiento: Entidades territoriales de la república de colombia.

Jappelli, T. and Pagano, M. (2002). Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking & Finance*, 26(10):2017–2045.

Jayachandran, S. (2006). Selling labor low: Wage responses to productivity shocks in developing countries. *Journal of Political Economy*, 114(3):538–575.

Jensen, R. (2000). Agricultural volatility and investments in children. *American Economic Review*, 90(2):399–404.

Karlan, D., Osei, R., Osei-Akoto, I., and Udry, C. (2014). Agricultural decisions after relaxing credit and risk constraints. *The Quarterly Journal of Economics*, 129(2):597–652.

Kaur, S. (2019). Nominal wage rigidity in village labor markets. *American Economic Review*, 109(10):3585–3616.

Kazianga, H. and Udry, C. (2006). Consumption smoothing? livestock, insurance and drought in rural burkina faso. *Journal of Development Economics*, 79(2):413–446.

Liberman, A., Neilson, C., Opazo, L., and Zimmerman, S. (1986). The equilibrium effects of information deletion: Evidence from consumer credit markets. NBER working paper no. 25097.

Limodio, N. (2021). Liquidity risk and long-term finance: Evidence from a natural experiment. *Review of Economic Studies*. Forthcoming.

Macchiavello, R. and Florensa-Miquel, J. (2019). Buyer-driven upgrading in gvcs: The sustainable quality program in colombia. *Working Paper*.

Mester, L. J. (1997). What's the point of credit scoring? *Business Review*, 3:3–16.

Muñoz-Mora, J. C. (2016). Fighting while ‘talking’: Assessing the impact of civil war on agriculture. Doctoral Dissertation Chapter.

Musto, D. K. (2004). What happens when information leaves a market? evidence from postbankruptcy consumers. *The Journal of Business*, 77(4):725–748.

Pagano, M. and Jappelli, T. (1993). Information sharing in credit markets. *The Journal of Finance*, 48(5):1693–1718.

Pelka, N., Musshoff, O., and Weber, R. (2015). Does weather matter? how rainfall affects credit risk in agricultural microfinance. *Agricultural Finance Review*, 75(2):194–212.

Petrick, M. (2004). A microeconometric analysis of credit rationing in the polish farm sector. *European Review of Agricultural Economics*, 31(1):77–101.

Rocha Sousa, M., Gama, J., and Brandao, E. (2016). A new dynamic modeling framework for credit risk assessment. *Expert Systems With Applications*, 45:341–351.

Rosenzweig, M. R. and Wolpin, K. I. (1993). Credit market constraints, consumption smoothing, and the accumulation of durable production assets in low-income countries: Investments in bullocks in india. *Journal of Political Economy*, 101(2):223–244.

Steinberg, J. (2014). Your privacy is now at risk from search engines even if the law says otherwise. *Forbes*.

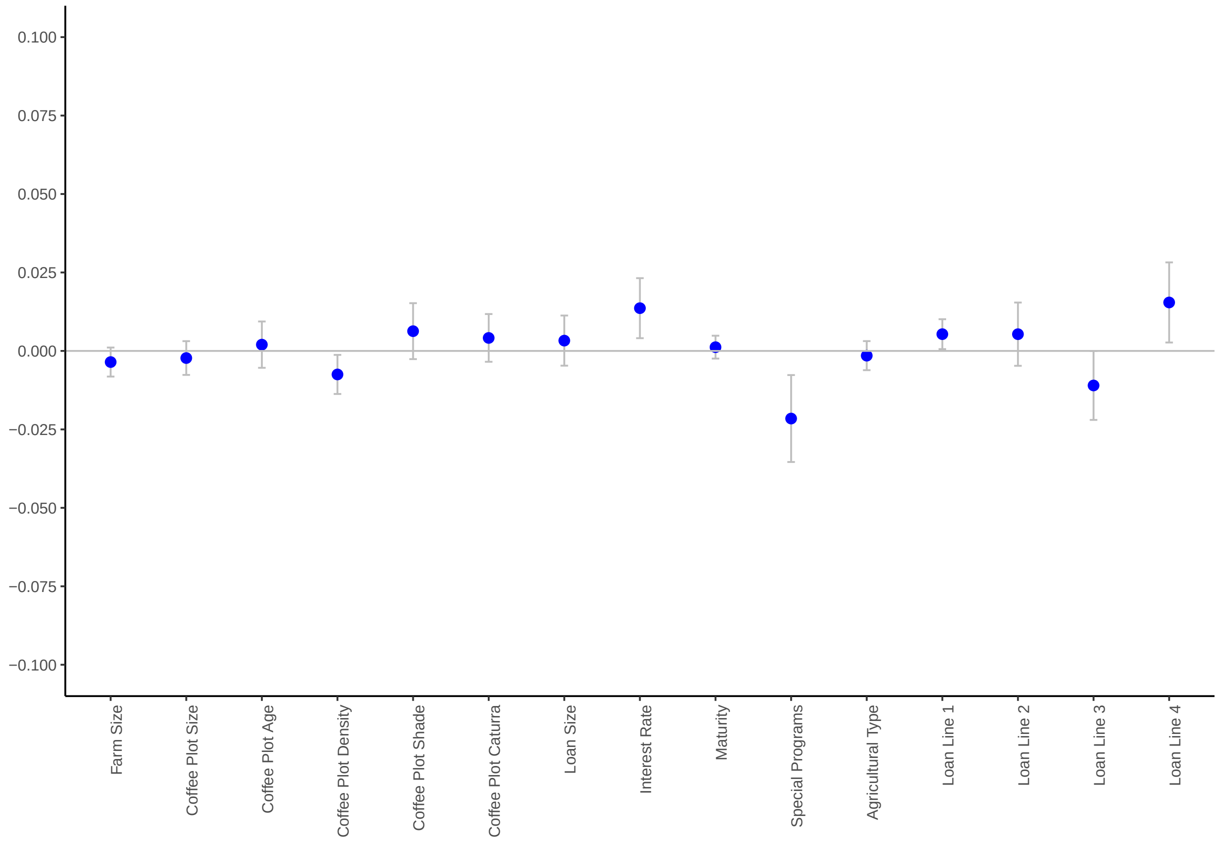
Trenberth, K. E. (2006). The impact of climate change and variability on heavy precipitation, floods, and droughts. *Encyclopedia of Hydrological Sciences*.

Turbay, S., Nates, B., Jaramillo, F., Vélez, J. J., and Ocampo, O. L. (2014). Adaptación a la variabilidad climática entre los caficultores de las cuencas de los ríos porce y chinchiná, colombia. *Investigaciones Geográficas, Boletín del Instituto de Geografía*, 2014(85):95–112.

Wei, Y., Yildirim, P., Van den Bulte, C., and Dellarocas, C. (2015). Credit scoring with social network data. *Marketing Science*, 35(2):234–258.

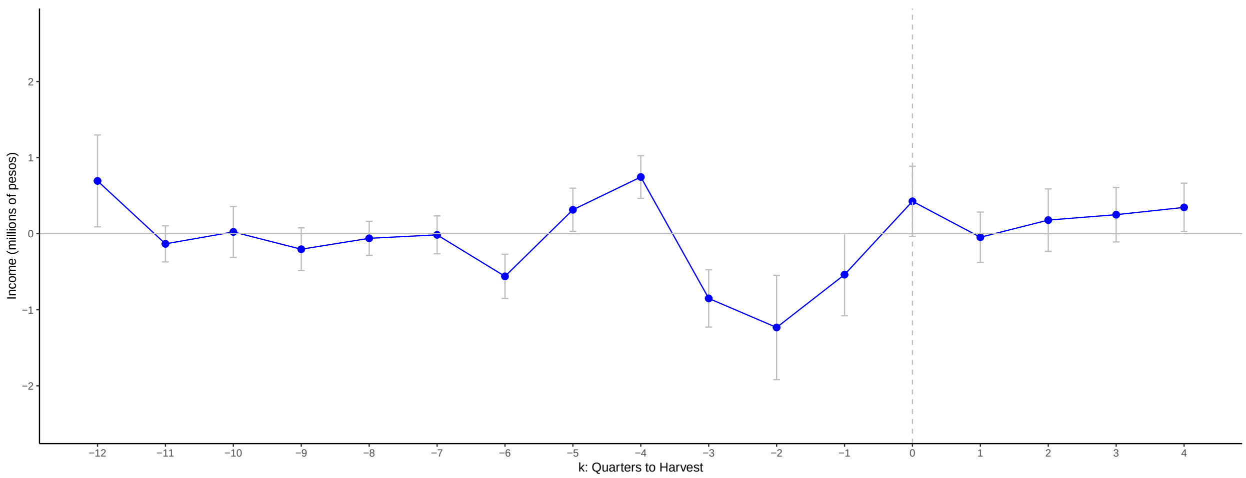
World Bank (2007). World development report 2008: Agriculture for development. Technical report.

**Figure 1:** Covariate Balance Test



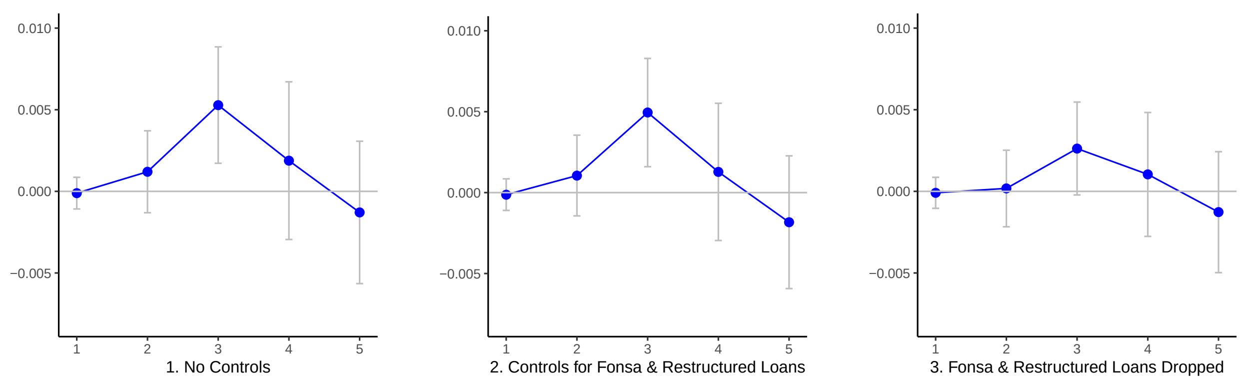

*Notes*: This figure displays estimates of the effect of exogenous shocks on predetermined covariates at the loan level and 90% confidence intervals. Each estimate corresponds to the covariate listed in the horizontal axis. The covariates are defined in the main text and the Data Appendix. The outcome and shocks variables are standardized (by subtracting the sample mean and dividing by the sample standard deviation) before estimating equation (1) using the variable listed in the horizontal axis as the outcome.

## **Figure 2:** Income Recovery




*Notes*: This figure displays the effect of exogenous shocks on income from the 2013 coffee harvest. All coefficients are obtained from the same regression – equation (4) in the text. Each dot shows the estimated effect of a shock during a lead or a lag quarter relative to the quarter of the harvest, 2013-q1. Leads and lags are denoted by $k$ in the horizontal axis with positive values corresponding to leads and negative values corresponding to lags. The dashed line indicates lag 0 (2013-q1). The vertical lines represent 90% confidence intervals. Income is measured in millions of pesos. The regression includes the following controls: department fixed effects; farm size; and dummies for the educational attainment of the household head; health insurance; access to electricity, sewerage, aqueduct; and households with high-quality walls and high-quality floors.

## **Figure 3:** Repayment Recovery




*Notes*: This figure displays the effect of a shock in the first year of the loan on repayment in subsequent years. Each coefficient is obtained from a different estimation of equation (5) using as an outcome the default rate in the year of the loan denoted in the horizontal axis. For example, the outcome used to obtain the coefficient above label "2" is a dummy equal to 1 if the loan was overdue (by 30 days or more) in year 2 of its life cycle. The variable of interest is the number of rainfall shocks in year 1 of the loan life cycle. All regressions include the controls listed in Panels A and C of Table 1. Vertical lines represent 90% confidence intervals. The sample of loans includes only long-term loans (with maturities of at least five years) linked to farms with three or more plots and with coffee trees aged 4.5 years or more on average. Panel A includes no controls for loans the BAC restructured or for loans FONSA purchased. Panel B includes a dummy for loans either restructured by the BAC or purchased by FONSA. Panel C drops loans the BAC restructured and loans FONSA purchased.

**Table 1:** Summary Statistics, Loan Sample

| | Mean | St. Dev. | Min | Pct. 25 | Median | Pct. 75 | Max |
|---|---|---|---|---|---|---|---|
| | A. Farm and Coffee Plot Characteristics | | | | | | |
| Dist. Rain. Stat. (km) | 6.37 | 3.76 | 0.02 | 3.80 | 5.86 | 8.18 | 40.81 |
| Farm Size (ha) | 2.84 | 2.22 | 0 | 1.1 | 2.1 | 4 | 10 |
| Coffee Plot Size (ha) | 1.56 | 1.20 | 0.00 | 0.70 | 1.20 | 2.03 | 10.00 |
| Coffee Plot Age | 6.50 | 5.50 | 0.00 | 3.37 | 5.10 | 7.82 | 79.75 |
| Coffee Plot Density | 5.05 | 1.13 | 0.00 | 4.35 | 5.01 | 5.63 | 47.58 |
| Coffee Plot Shade | 0.12 | 0.27 | 0 | 0 | 0 | 0 | 1 |
| Coffee Plot Caturra | 0.59 | 0.41 | 0.00 | 0.00 | 0.67 | 1.00 | 1.00 |
| | B. Exogenous Shocks | | | | | | |
| Shocks | 1.47 | 1.13 | 0 | 1 | 1 | 2 | 4 |
| Shocks (p. 90) | 0.94 | 0.95 | 0 | 0 | 1 | 2 | 4 |
| Shocks $\geq 2$ | 0.46 | 0.50 | 0 | 0 | 0 | 1 | 1 |
| | C. Loan Characteristics | | | | | | |
| Loan Size | 3.36 | 2.42 | 0.00 | 1.55 | 2.58 | 4.74 | 28.92 |
| Interest Rate | 10.34 | 2.50 | 2.00 | 9.13 | 10.28 | 11.66 | 43.98 |
| Maturity (years) | 3.31 | 2.62 | 0 | 1 | 2 | 6 | 30 |
| Special Programs | 0.09 | 0.28 | 0 | 0 | 0 | 0 | 1 |
| Agricultural Type | 0.98 | 0.15 | 0 | 1 | 1 | 1 | 1 |
| Loan Line 1 | 0.55 | 0.50 | 0 | 0 | 1 | 1 | 1 |
| Loan Line 2 | 0.22 | 0.42 | 0 | 0 | 0 | 0 | 1 |
| Loan Line 3 | 0.16 | 0.36 | 0 | 0 | 0 | 0 | 1 |
| Loan Line 4 | 0.06 | 0.23 | 0 | 0 | 0 | 0 | 1 |
| | D. Loan Outcomes | | | | | | |
| Overdue 30 | 0.15 | 0.35 | 0 | 0 | 0 | 0 | 1 |
| Overdue 60 | 0.11 | 0.31 | 0 | 0 | 0 | 0 | 1 |
| Loan Score Fell | 0.18 | 0.39 | 0 | 0 | 0 | 0 | 1 |
| Loan Score Fell to E | 0.09 | 0.29 | 0 | 0 | 0 | 0 | 1 |
| | E. Application Outcomes | | | | | | |
| Applied | 0.69 | 0.46 | 0 | 0 | 1 | 1 | 1 |
| Time Lapse | 1.25 | 1.14 | 0.08 | 0.25 | 1.00 | 1.75 | 7.00 |
| Bureau Score | 934.97 | 107.25 | 11.00 | 929.00 | 964.00 | 983.00 | 1,051.00 |
| Denial | 0.10 | 0.30 | 0 | 0 | 0 | 0 | 1 |

*Notes*: This table displays summary statistics of the variables in the Loan Sample (Panels A to D) and in the Application Sample (Panel E). The Loan Sample consists of loans disbursed to coffee farmers by the BAC from 2005 to 2011 that are linked to a SICA farm and a rainfall station. The Application Sample consists of loan applications preceded by a loan disbursed in 2008–2011. Panel A presents summary statistics of the characteristics of the farm and of the coffee plot. Panel B presents statistics of the number of rainfall shocks in the first year of the loan. Panel C presents loan characteristics. The loan size is measured in millions of pesos of 2010. The Loan Line denotes a group of loans with similar characteristics (purchases of fertilizer, for example) and the loan line variables are the corresponding dummies. Panel D presents statistics of loan outcomes. Overdue 30 and 60 are dummies equal to one if the loan was 30 or 60 days past due at any point in time. Panel E presents statistics of application outcomes. See the main text and the Data Appendix for definitions of the variables and details of the construction of the samples.

**Table 2:** Effect of Exogenous Shocks on Current Loan Outcomes

| | A. Default Outcomes | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 30 Days Overdue | | 60 Days Overdue | Distance $<$ p. 75 | Matu. $\leq 3$ | Matu. $> 3$ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Shocks | 0.004*** | | | 0.004*** | 0.004*** | 0.004*** | 0.003 |
| | (0.001) | | | (0.001) | (0.002) | (0.001) | (0.002) |
| Shocks (p. 90) | | 0.004** | | | | | |
| | | (0.001) | | | | | |
| Shocks $\geq 2$ | | | 0.009*** | | | | |
| | | | (0.003) | | | | |
| Outcome Mean | 0.147 | 0.147 | 0.147 | 0.109 | 0.141 | 0.097 | 0.209 |
| Observations | 242,332 | 242,332 | 242,332 | 242,332 | 182,172 | 134,064 | 108,268 |
| Adjusted $R^2$ | 0.083 | 0.083 | 0.083 | 0.076 | 0.076 | 0.048 | 0.077 |
| | B. Loan Score Outcomes | | | | | | |
| | | Score Fell | | Score Fell to E | Distance $<$ p. 75 | Matu. $\leq 3$ | Matu. $> 3$ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Shocks | 0.006*** | | | 0.003*** | 0.007*** | 0.006*** | 0.006** |
| | (0.002) | | | (0.001) | (0.002) | (0.002) | (0.002) |
| Shocks (pct90) | | 0.004** | | | | | |
| | | (0.002) | | | | | |
| Shocks $> 2$ | | | 0.013*** | | | | |
| | | | (0.003) | | | | |
| Outcome Mean | 0.185 | 0.185 | 0.185 | 0.092 | 0.179 | 0.103 | 0.285 |
| Observations | 242,332 | 242,332 | 242,332 | 242,332 | 182,172 | 134,064 | 108,268 |
| Adjusted $R^2$ | 0.104 | 0.104 | 0.104 | 0.077 | 0.098 | 0.050 | 0.062 |

*Notes*: This table displays estimates of the effect of exogenous shocks on current loan outcomes (equation (1) in the text). The sample consists of loans for coffee production disbursed by the BAC in 2005–2011 to coffee farmers and linked to a rainfall station. Each coefficient corresponds to a different regression. The shock measure is listed at the left of each line: *Shocks* is the number of shocks in the first year of the loan, using the 80th percentile of the calendar quarter distribution of rainfall to define shocks. *Shocks (p. 90)* uses the 90th percentile to define shocks. *Shocks* $\geq 2$ is a dummy equal to 1 if *Shocks* is greater than or equal to 2. In all columns of Panel A (except in column (4)), the outcome is a dummy equal to 1 for loans that were 30 days overdue at any point in time. The outcome of column (4) is a dummy equal to 1 for loans that were 60 days overdue at any point in time. Column (5) includes only loans of farms with a distance to the rainfall station below the 75th percentile of the sample of column (1). Column (6) considers only loans with maturity $\leq 3$ years, and column (7) considers loans with maturity $> 3$. In all columns of Panel B (except column (4)) the outcome is a dummy equal to 1 for loans with a Loan Score that fell from A at any point in time. In column (4) the outcome is a dummy for loans with a Loan Score that fell to E, the lowest possible score, at any point in time. All regressions include the controls listed in Panels A and C of Table 1. Except for columns (6) and (7), all columns include quarter-year of disbursement times maturity fixed effects, where the maturity fixed effect is a dummy equal to one for loans with a maturity of 3 or more years. Columns (6) and (7) only include date of disbursement fixed effects. Standard errors, clustered at the municipality level, are reported in parentheses.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

**Table 3:** Effect of Exogenous Shocks on Future Loan Applications

| | Applied | Initial Loan Overdue | Bureau Score | Application Denial |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Shocks | $-0.008^{*}$ | $0.015^{***}$ | $-3.036^{***}$ | $0.008^{***}$ |
| | (0.004) | (0.004) | (0.986) | (0.003) |
| Outcome Mean | 0.688 | 0.129 | 935 | 0.097 |
| Observations | 41,425 | 26,569 | 26,176 | 26,569 |
| Adjusted $R^2$ | 0.152 | 0.126 | 0.073 | 0.047 |

*Notes*: This table displays estimates of the effect of exogenous shocks on application outcomes. The sample of column (1) consists of the most recent loan in the period 2008–2011 for each coffee producers in the Loan Sample (the sample of Table 2). The sample of columns (2) to (4) consists of applications preceded by a loan disbursed in 2008–2011. *Shocks* is the number of shocks in the first year of the *initial* loan. The outcome of column (1) is a dummy equal to one for individuals who applied for a new loan after the end of the initial loan. The outcome of column (2) is a dummy equal to one for initial loans that were at least 30 days overdue, at any point in time. The outcome of column (3) is the application's Bureau Score, and the outcome of column (4) is a dummy indicating rejected applications. All regressions include the controls listed in Panels A and C of Table 1. All columns include quarter-year of disbursement times maturity fixed effects according to the initial loan, where the maturity fixed effect is a dummy equal to one for loans with a maturity of 3 or more years. Standard errors, clustered at the municipality level, are reported in parentheses.
$^{*}$ $p < 0.1$; $^{**}$ $p < 0.05$; $^{***}$ $p < 0.01$.

**Table 4:** Persistence of Effect of Shocks on Future Loan Applications

| | Two Years in Sample | | All Years in Sample | |
|---|---|---|---|---|
| | Score | Denial | Score | Denial |
| | (1) | (2) | (3) | (4) |
| Shocks, t | $-1.083$ | $0.002^{*}$ | $-0.383$ | $0.004^{*}$ |
| | (0.684) | (0.001) | (1.057) | (0.002) |
| Shocks, t-1 | $-2.263^{***}$ | $0.005^{***}$ | $-3.077^{***}$ | $0.009^{***}$ |
| | (0.810) | (0.001) | (1.091) | (0.002) |
| Shocks, t-2 | $-2.893^{***}$ | $0.006^{***}$ | $-2.901^{***}$ | $0.007^{***}$ |
| | (0.760) | (0.001) | (1.049) | (0.002) |
| Shocks, t-3 | $-2.045^{***}$ | $0.004^{***}$ | $-2.227^{**}$ | $0.005^{**}$ |
| | (0.695) | (0.001) | (0.962) | (0.002) |
| Shocks, t-4 | $-1.222^{*}$ | 0.001 | $-0.957$ | 0.002 |
| | (0.650) | (0.001) | (0.972) | (0.002) |
| Shocks, t-5 | $-1.212^{**}$ | $0.002^{*}$ | $-2.133^{*}$ | $0.003^{*}$ |
| | (0.588) | (0.001) | (1.097) | (0.002) |
| Outcome mean | 933.1 | 0.058 | 938.6 | 0.037 |
| Observations | 235,965 | 242,188 | 16,710 | 16,710 |
| Adjusted $R^2$ | 0.301 | 0.229 | 0.345 | 0.210 |

*Notes*: This table displays estimates of the effect of the lags of exogenous shocks on the Bureau Score and on the rejection probability of loan applications in year $t$, using an individual-year level panel. Each column corresponds to one regression that includes five lags ($t-1$ to $t-5$) of the shocks variable (equation (3) in the text). The sample consists of coffee producers in the Bureau Reports who applied for a loan in at least two years in 2010–2015 (columns (1) and (2)) or who applied in all years in 2010–2015 (columns (3) and (4)). The outcome of columns (1) and (3) is the average Bureau Score in year $t$. The outcome of columns (2) and (4) is a dummy equal to 1 if any of the applications in year $t$ were rejected. All regressions include individual and year-fixed effects. Standard errors, clustered at the municipality level, are reported in parentheses.
$^{*}$ $p < 0.1$; $^{**}$ $p < 0.05$; $^{***}$ $p < 0.01$.

**Table 5:** Long-Term Loans Repayment Paths

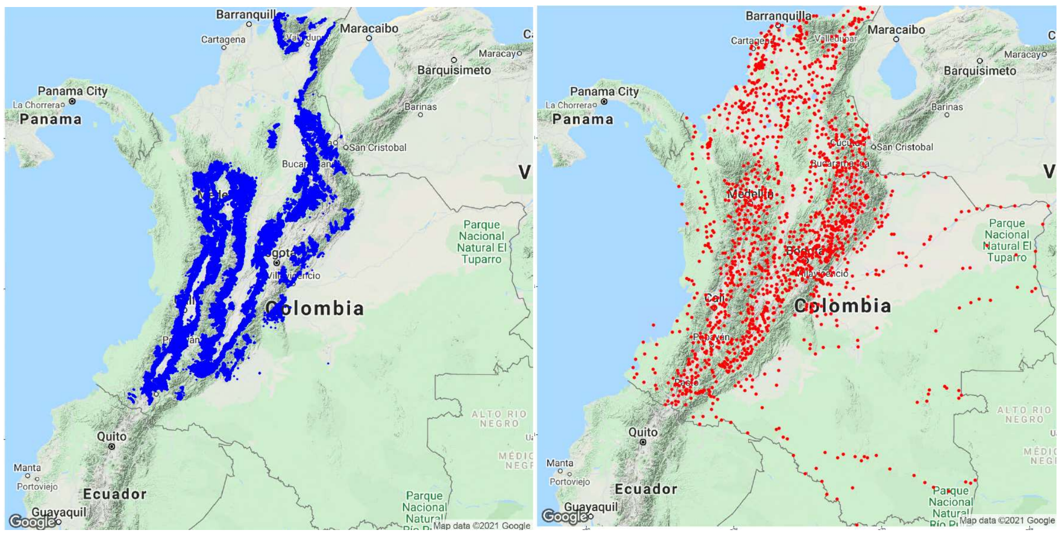
| | | Baseline Sample | | Baseline (no R. no F.) | | Full Sample | | Full Sample (no R. no F.) | |
|---|---|---|---|---|---|---|---|---|---|
| | | # | freq | # | freq. | # | freq. | # | freq. |
| **path** | y1 y2 y3 y4 y5 | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| **17**: | 0 - 1 - 1 - 1 - 1 | 127 | 0.005 | 89 | 0.004 | 597 | 0.006 | 446 | 0.005 |
| **18**: | 0 - 1 - 1 - 1 - 0 | 72 | 0.003 | 18 | 0.001 | 381 | 0.004 | 113 | 0.001 |
| **19**: | 0 - 1 - 1 - 0 - 1 | 8 | 0 | 6 | 0 | 19 | 0 | 12 | 0 |
| **20**: | 0 - 1 - 1 - 0 - 0 | 73 | 0.003 | 34 | 0.002 | 232 | 0.002 | 114 | 0.001 |
| **21**: | 0 - 1 - 0 - 1 - 1 | 17 | 0.001 | 9 | 0 | 81 | 0.001 | 39 | 0 |
| **22**: | 0 - 1 - 0 - 1 - 0 | 28 | 0.001 | 9 | 0 | 103 | 0.001 | 42 | 0 |
| **23**: | 0 - 1 - 0 - 0 - 1 | 20 | 0.001 | 11 | 0.001 | 69 | 0.001 | 33 | 0 |
| **24**: | 0 - 1 - 0 - 0 - 0 | 256 | 0.01 | 161 | 0.007 | 928 | 0.009 | 625 | 0.007 |
| **25**: | 0 - 0 - 1 - 1 - 1 | 203 | 0.008 | 111 | 0.005 | 933 | 0.009 | 542 | 0.006 |
| **26**: | 0 - 0 - 1 - 1 - 0 | 218 | 0.009 | 69 | 0.003 | 970 | 0.009 | 306 | 0.003 |
| **27**: | 0 - 0 - 1 - 0 - 1 | 29 | 0.001 | 16 | 0.001 | 116 | 0.001 | 83 | 0.001 |
| **28**: | 0 - 0 - 1 - 0 - 0 | 415 | 0.017 | 210 | 0.01 | 1423 | 0.014 | 745 | 0.008 |
| **29**: | 0 - 0 - 0 - 1 - 1 | 759 | 0.031 | 392 | 0.018 | 4273 | 0.041 | 2281 | 0.024 |
| **30**: | 0 - 0 - 0 - 1 - 0 | 893 | 0.036 | 629 | 0.029 | 4432 | 0.042 | 3317 | 0.035 |
| **31**: | 0 - 0 - 0 - 0 - 1 | 941 | 0.038 | 600 | 0.027 | 4204 | 0.04 | 2806 | 0.03 |
| **32**: | 0 - 0 - 0 - 0 - 0 | 20367 | 0.831 | 19475 | 0.89 | 86111 | 0.818 | 82962 | 0.876 |
| Total loans | | 24513 | | 21892 | | 105229 | | 94677 | |
| Ovd. loans recov. | | 1955 | 7.9% | 1130 | 5.2% | 8469 | 8.1% | 5262 | 5.5% |
| Ovd. loans non-recov. | | 2104 | 8.5% | 1234 | 5.6% | 10292 | 9.9% | 6242 | 6.6% |
| Loans never overdue | | 20367 | 83.1% | 19475 | 89.0% | 86111 | 81.8% | 82962 | 87.6% |

*Notes*: This table shows the distribution of long-term loans across repayment paths. A path is defined by a combination of default outcomes in years 1 to 5. Each line corresponds to one of the 32 possible paths. For example, path 26 corresponds to loans that were overdue (by 30 days or more) in years 3 and 4 but not overdue in years 1, 2, or 5. Paths 1 to 16 start with an overdue in year 1 and are omitted from the table. The table presents the number of loans in each path and the corresponding frequency in four different samples: in columns (1) and (2) the Baseline Long-Term Loan Sample (loans of maturities of at least five years, linked to farms with three or more plots and with coffee trees aged 4.5 years or more on average), in columns (3) and (4) the Baseline Long-Term Loan Sample excluding loans that the BAC restructured or that FONSA purchased, in columns (5) and (6) the Full Sample of Long-Term Loans (all loans of maturities of at least five years), and in columns (7) and (8) the Full Sample of Long-Term Loans excluding those the BAC restructured or those FONSA purchased. The line *Total loans* reports the total number of loans in each sample. The line *Overdue loans recovery* reports the number and frequency of loans in a recovery path (not overdue in year 5 and overdue in at least one of years 2, 3, and 4). The line *Overdue loans non-recovery* reports the number and frequency of loans that were overdue but not in a recovery path (overdue in year 5). The line *Loans never overdue* reports the number of loans that were never overdue.

# Appendix for Online Publication

## A. Appendix Figures and Tables

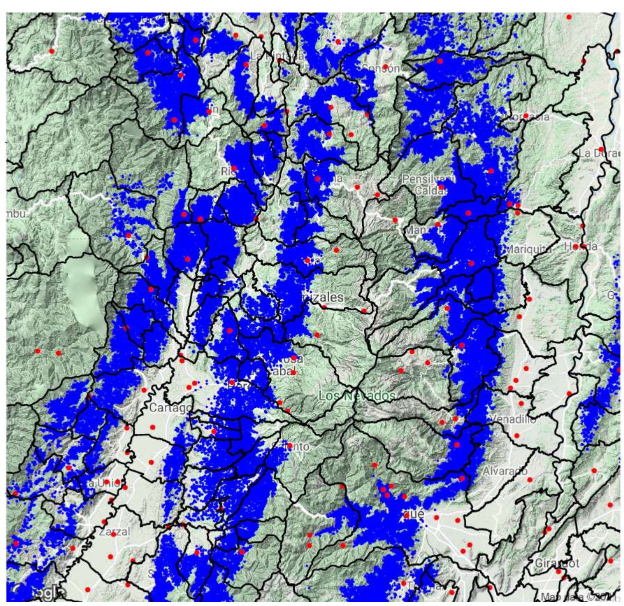
**Figure A1:** Coffee Farms and Rainfall Stations in Colombia




A. SICA Farms B. IDEAM Rainfall Stations

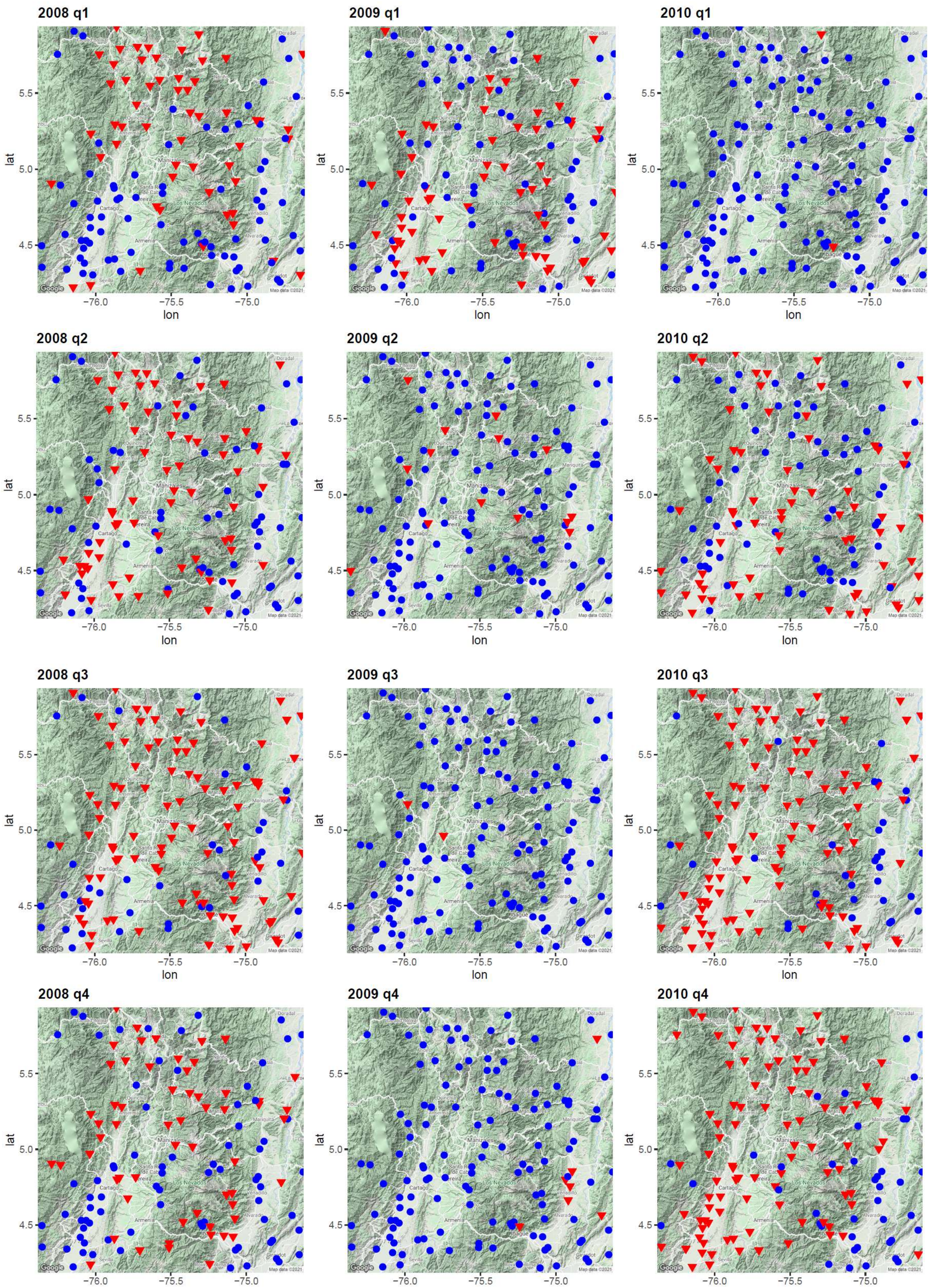
*Notes*: This figure shows the distribution of SICA farms (Panel A) and IDEAM rainfall stations (Panel B) in Colombia. Panel A includes farms that appear in the SICA data for the years 2006–2014.

**Figure A2:** Coffee Farms and Rainfall Stations in the "Zona Cafetera"



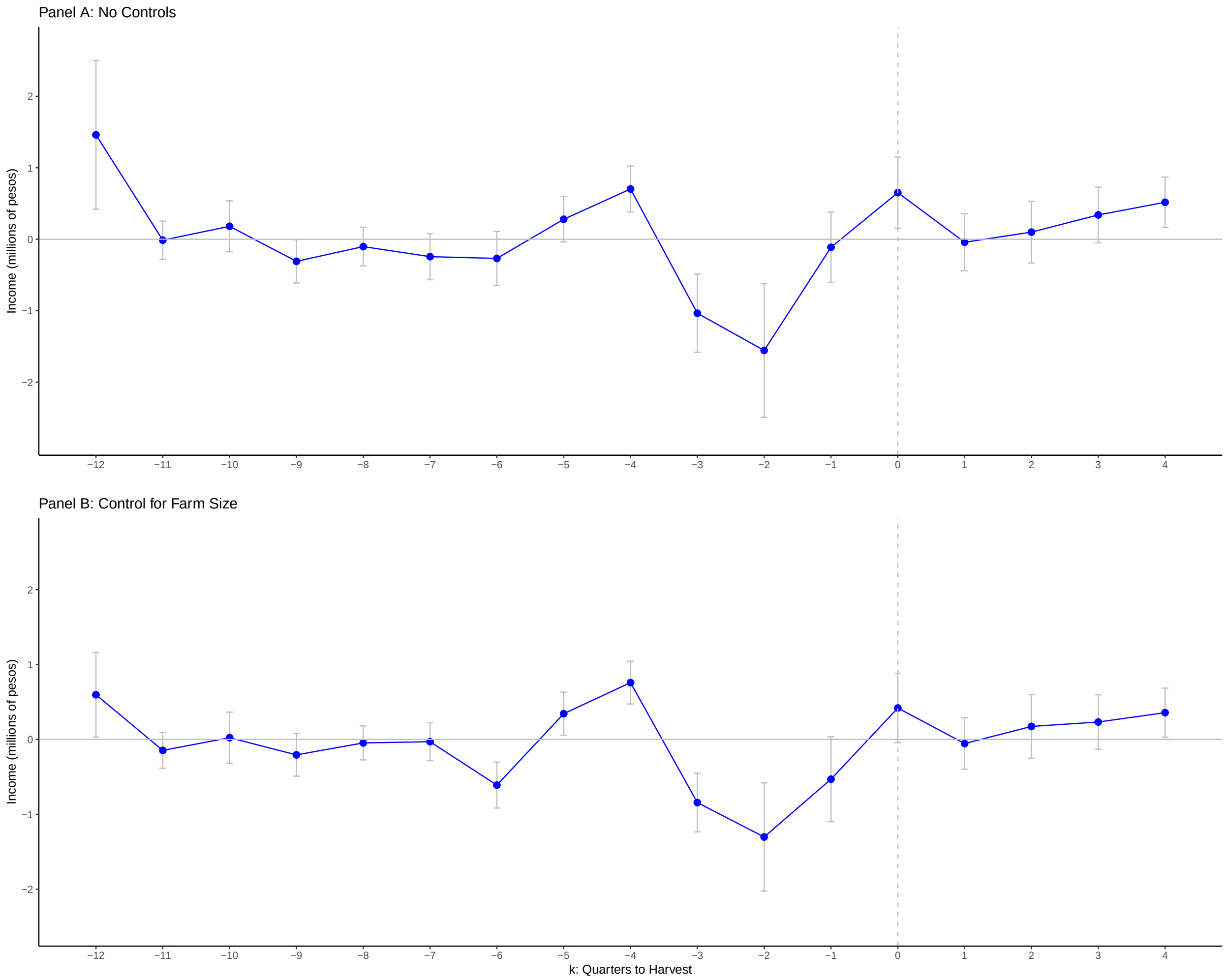

*Notes*: This figure shows the distribution of SICA farms (in blue) and IDEAM rainfall stations (in red) in the "Zona Cafetera" (the Coffee Zone) of Colombia. Black lines illustrate municipality boundaries.

**Figure A3:** Shocks in the "Zona Cafetera" (2008-2010)



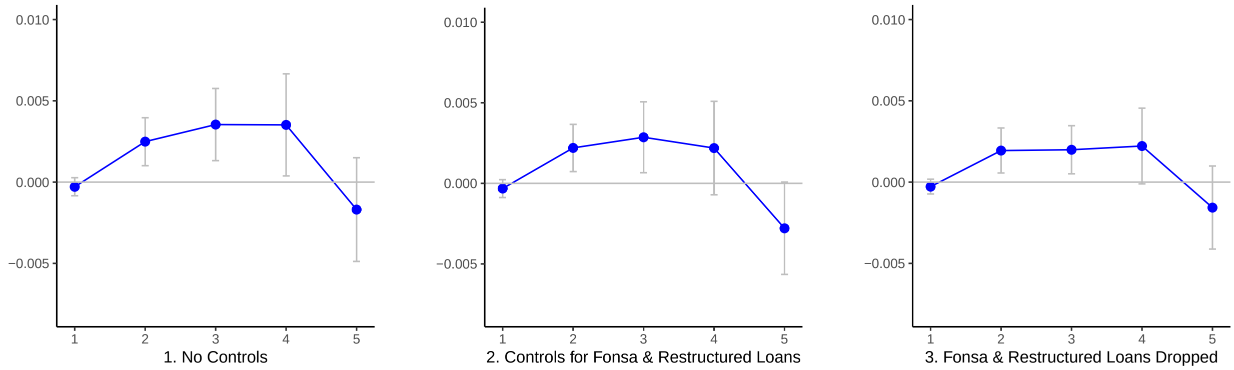

*Notes*: This figure shows quarter-year shocks for the period 2008-q1 to 2010-q4 for the rainfall stations of the "Zona Cafetera". A red triangle represents a rainfall station that experienced a shock in the corresponding quarter-year. A blue circle represents a rainfall station that did not experience a shock.

**Figure A4:** Income Recovery, Robustness




*Notes*: This figure displays the effect of exogenous shocks on income from the 2013 coffee harvest. The figure is analogous to Figure 2 in the main text but with different sets of farm-level controls. Panel A does not include controls and Panel B includes farm size only. In each panel, all coefficients are obtained from the same regression – equation (4) in the text. Each dot shows the estimated effect of a shock during a lead or a lag quarter relative to the quarter of the harvest, 2013-q1. Leads and lags are denoted by $k$ in the horizontal axis with positive values corresponding to leads and negative values corresponding to lags. The dashed line indicates lag 0 (2013-q1). The vertical lines represent 90% confidence intervals. Income is measured in millions of pesos.

**Figure A5:** Repayment Recovery, All Long-Term Loans




*Notes*: This figure displays the effect of a shock in the first year of the loan on repayment in subsequent years. It is analogous to Figure 3 in the text but it includes all long-term loans (maturities of five years or more). Each coefficient is obtained from a different estimation of equation (5) using as an outcome the default rate in the year of the loan denoted in the horizontal axis. For example, the outcome used to obtain the coefficient above label "2" is a dummy equal to 1 if the loan was overdue (by 30 days or more) in year 2 of its life cycle. The variable of interest is the number of rainfall shocks in year 1 of the loan life cycle. All regressions include the controls listed in Panels A and C of Table 1. Vertical lines represent 90% confidence intervals. Panel A includes no controls for loans the BAC restructured or for loans FONSA purchased. Panel B includes a dummy for loans either restructured by the BAC or purchased by FONSA. Panel C drops loans the BAC restructured and loans FONSA purchased.

**Table A1:** Effect of Exogenous Shocks on Current Loan Outcomes, No Controls

| | A. Default Outcomes | | | | | | |
|---|---|---|---|---|---|---|---|
| | | 30 Days Overdue | | 60 Days Overdue | Distance < p. 75 | Matu. $\leq 3$ | Matu. $> 3$ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Shocks | 0.004** | | | 0.003*** | 0.004** | 0.004*** | 0.002 |
| | (0.001) | | | (0.001) | (0.002) | (0.001) | (0.002) |
| Shocks (pct90) | | 0.004** | | | | | |
| | | (0.002) | | | | | |
| Shocks > 2 | | | 0.008*** | | | | |
| | | | (0.003) | | | | |
| Outcome mean | 0.152 | 0.152 | 0.152 | 0.113 | 0.146 | 0.104 | 0.212 |
| Observations | 249,684 | 249,684 | 249,684 | 249,684 | 187,886 | 138,448 | 111,236 |
| Adjusted $R^2$ | 0.077 | 0.077 | 0.077 | 0.070 | 0.070 | 0.051 | 0.068 |
| | B. Loan Score Outcomes | | | | | | |
| | | Score Fell | | Score Fell to E | Distance < p. 75 | Matu. $\leq$ | Matu. $>$ |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Shocks | 0.005** | | | 0.001 | 0.005** | 0.003 | 0.005** |
| | (0.002) | | | (0.002) | (0.002) | (0.002) | (0.002) |
| Shocks (pct90) | | 0.002 | | | | | |
| | | (0.002) | | | | | |
| Shocks > 2 | | | 0.011*** | | | | |
| | | | (0.004) | | | | |
| Outcome mean | 0.2 | 0.2 | 0.2 | 0.107 | 0.195 | 0.127 | 0.291 |
| Observations | 249,684 | 249,684 | 249,684 | 249,684 | 187,886 | 138,448 | 111,236 |
| Adjusted $R^2$ | 0.103 | 0.103 | 0.104 | 0.086 | 0.099 | 0.089 | 0.057 |

*Notes*: This table displays estimates of the effect of exogenous shocks on current loan outcomes (equation (1) in the text). It is analogous to Table 2 in the text but without controls. The sample consists of loans for coffee production disbursed by the BAC in 2005–2011 to coffee farmers and linked to a rainfall station. Each coefficient corresponds to a different regression. The shock measure is listed at the left of each line: *Shocks* is the number of shocks in the first year of the loan, using the 80th percentile of the calendar quarter distribution of rainfall to define shocks. *Shocks (p. 90)* uses the 90th percentile to define shocks. *Shocks* $\geq 2$ is a dummy equal to 1 if *Shocks* is greater than or equal to 2. In all columns of Panel A (except in column (4)), the outcome is a dummy equal to 1 for loans that were 30 days overdue at any point in time. The outcome of column (4) is a dummy equal to 1 for loans that were 60 days overdue at any point in time. Column (5) includes only loans of farms with a distance to the rainfall station below the 75th percentile of the sample of column (1). Column (6) considers only loans with maturity $\leq 3$ years, and column (7) considers loans with maturity $> 3$. In all columns of Panel B (except column (4)) the outcome is a dummy equal to 1 for loans with a Loan Score that fell from A at any point in time. In column (4) the outcome is a dummy for loans with a Loan Score that fell to E, the lowest possible score, at any point in time. All regressions include the controls listed in Panels A and C of Table 1. Except for columns (6) and (7), all columns include quarter-year of disbursement times maturity fixed effects, where the maturity fixed effect is a dummy equal to one for loans with a maturity of 3 or more years. Columns (6) and (7) only include date of disbursement fixed effects. Standard errors, clustered at the municipality level, are reported in parentheses.
* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

**Table A2:** Persistence of Effect of Shocks on Future Applications

| | Time Lapse: 1 Year | | Time Lapse: 2 Years | | Time Lapse: 3 Years | |
|---|---|---|---|---|---|---|
| | Bureau Score (1) | Denial (2) | Bureau Score (3) | Denial (4) | Bureau Score (5) | Denial (6) |
| Shocks | $-4.539^{***}$ | $0.011^{***}$ | $-6.308^{**}$ | 0.007 | $-9.211^{**}$ | 0.001 |
| | (1.492) | (0.004) | (2.615) | (0.007) | (3.713) | (0.010) |
| Outcome mean | 924.2 | 0.112 | 914.3 | 0.134 | 909.4 | 0.14 |
| Observations | 13,633 | 13,889 | 6,011 | 6,106 | 2,779 | 2,786 |
| Adjusted $R^2$ | 0.056 | 0.048 | 0.053 | 0.026 | 0.013 | 0.028 |

*Notes*: This table displays estimates of the effect of exogenous shocks on application outcomes in the Application Sample. It is analogous to Table 3 but different samples are considered depending on the time lapse between the end of the initial loan and the application. The Application Sample consists of applications preceded by a loan disbursed in 2008–2011. *Shocks*, the variable of interest, is the number of shocks in the first year of the *initial* loan. The outcome of columns (1), (3), and (5) is the application's Bureau Score and the outcome of columns (2), (4), and (6) is a dummy indicating rejected applications. All regressions include controls, listed in Panels A and C of Table 1. Standard errors, clustered at the municipality level, are reported in parentheses.
$^{*}$ $p < 0.1$; $^{**}$ $p < 0.05$; $^{***}$ $p < 0.01$.

**Table A3:** Summary Statistics, Agricultural Census

| | Mean | St. Dev. | Min | Pct. 25 | Median | Pct. 75 | Max |
|---|---|---|---|---|---|---|---|
| Distance | 6.55 | 3.80 | 0.02 | 3.89 | 6.07 | 8.47 | 71.28 |
| Farm Size (ha) | 2.51 | 2.42 | 0.01 | 0.61 | 1.67 | 3.71 | 10.00 |
| Health Insurance | 0.10 | 0.30 | 0 | 0 | 0 | 0 | 1 |
| Edu. No Degree | 0.15 | 0.36 | 0 | 0 | 0 | 0 | 1 |
| Edu. Primary Mid. | 0.77 | 0.42 | 0 | 1 | 1 | 1 | 1 |
| Edu. High School | 0.06 | 0.24 | 0 | 0 | 0 | 0 | 1 |
| Edu. Technical | 0.01 | 0.10 | 0 | 0 | 0 | 0 | 1 |
| Edu. College | 0.01 | 0.10 | 0 | 0 | 0 | 0 | 1 |
| Edu. Graduate | 0.002 | 0.04 | 0 | 0 | 0 | 0 | 1 |
| Wall Quality | 0.85 | 0.36 | 0 | 1 | 1 | 1 | 1 |
| Floor Quality | 0.64 | 0.48 | 0 | 0 | 1 | 1 | 1 |
| Aqueduct | 0.45 | 0.50 | 0 | 0 | 0 | 1 | 1 |
| Sewerage | 0.06 | 0.24 | 0 | 0 | 0 | 0 | 1 |
| Electricity | 0.91 | 0.28 | 0 | 1 | 1 | 1 | 1 |
| Coffe Income | 3.88 | 5.61 | 0 | 0.3 | 1.6 | 5.0 | 30 |

*Notes*: This table displays the summary statistics of coffee producers in the agricultural census. Distance is the distance in kilometers to the closest rainfall station. Health insurance and education variables correspond to the maximum level across household heads on the farm. Wall and floor quality are dummies equal to 1 if at least one dwelling on the farm has high-quality walls or floors. Aqueduct, sewerage, and electricity are dummies equal to 1 if at least one of the dwellings is connected to the aqueduct or has sewerage or electricity service. Coffee income is the total income generated from coffee production in the 2013 harvest. I use a price of 3,730 pesos per kilo of parchment coffee to convert quantities to income. See the Data Appendix for details on the variables and the construction of the sample.

**Table A4:** Summary Statistics, Long-Term Loans

| | Mean | St. Dev. | Min | Pct. 25 | Median | Pct. 75 | Max |
|---|---|---|---|---|---|---|---|
| | A. Farm and Coffee Plot Characteristics | | | | | | |
| Farm Size (ha) | 3.11 | 2.30 | 0.00 | 1.39 | 2.50 | 4.20 | 10.00 |
| # of Coffee Plots | 3.73 | 2.60 | 1 | 2 | 3 | 5 | 29 |
| Coffee Plot Size (ha) | 1.60 | 1.20 | 0.00 | 0.76 | 1.27 | 2.10 | 10.00 |
| Coffee Plot Age | 5.81 | 5.50 | 0.00 | 2.65 | 4.50 | 7.03 | 79.75 |
| Coffee Plot Density | 5.12 | 1.09 | 0.00 | 4.46 | 5.10 | 5.67 | 47.58 |
| Coffee Plot Shade | 0.11 | 0.26 | 0 | 0 | 0 | 0 | 1 |
| Coffee Plot Caturra | 0.53 | 0.41 | 0 | 0 | 0.5 | 1 | 1 |
| | B. Loan Characteristics | | | | | | |
| Loan Size | 4.76 | 2.64 | 0.00 | 2.89 | 4.53 | 6.00 | 28.92 |
| Interest Rate | 9.04 | 2.58 | 2.65 | 7.58 | 8.77 | 10.44 | 17.03 |
| Maturity (years) | 6.18 | 0.93 | 5 | 5 | 7 | 7 | 30 |
| | C. Restructured and FONSA | | | | | | |
| Restructured | 0.051 | 0.22 | 0 | 0 | 0 | 0 | 1 |
| FONSA | 0.049 | 0.22 | 0 | 0 | 0 | 0 | 1 |
| | D. Overdue 30 | | | | | | |
| All Years | 0.21 | 0.41 | 0 | 0 | 0 | 0 | 1 |
| Years 1 | 0.003 | 0.06 | 0 | 0 | 0 | 0 | 1 |
| Year 2 | 0.02 | 0.15 | 0 | 0 | 0 | 0 | 1 |
| Year 3 | 0.05 | 0.21 | 0 | 0 | 0 | 0 | 1 |
| Year 4 | 0.11 | 0.32 | 0 | 0 | 0 | 0 | 1 |
| Year 5 | 0.10 | 0.30 | 0 | 0 | 0 | 0 | 1 |

*Notes*: This table displays the summary statistics for the sample of loans with maturities of five years or more. The variables in Panels A are B are the same as in Panels A and C of Table 1. In Panel C, Restructured is a dummy to denote loans that were restructured by the BAC and FONSA is a dummy for loans that FONSA purchased. Panel D presents default outcomes. All Years is a dummy equal to 1 if the loan was 30 days overdue at any point. Year 1 is a dummy for loans that were 30 days overdue in year 1 and so on. See the main text and the Data Appendix for definitions of the variables and details on the construction of the samples.

**Table A5:** Effect of Shocks on Restructuring and FONSA

| | A. Baseline Sample | |
|---|---|---|
| | Restructured | FONSA |
| | (1) | (2) |
| Shocks | 0.001 | 0.001 |
| | (0.003) | (0.002) |
| Outcome mean | 0.063 | 0.044 |
| Observations | 24,513 | 24,513 |
| Adjusted $R^2$ | 0.092 | 0.046 |
| | B. All Farms | |
| Shocks | 0.002 | 0.002* |
| | (0.002) | (0.001) |
| Outcome mean | 0.051 | 0.049 |
| Observations | 105,229 | 105,229 |
| Adjusted $R^2$ | 0.079 | 0.050 |

*Notes*: This table displays the estimated effect of exogenous shocks on the probability of loan restructuring (column (1)) and the probability that FONSA purchased the loan (column (2)). In Panel A, the sample consists of long-term loans (with maturities of at least five years) linked to farms with three or more plots and with coffee trees aged 4.5 years or more on average. The sample for Panel B is all long-term loans. The variable of interest is the number of rainfall shocks in year 1 of the loan tenure.

# B. Data Appendix

This appendix provides details on the variables and the construction of the samples used in the paper.

### B.1 Construction of the Samples

I use five different samples to study the effect of rainfall shocks: 1) the *Loan Sample*, which I mainly use in Section 3.3, 2) the *Application Sample*, used in Section 4.1, 3) the individual panel of applications, used in Section 4.2, 4) the sample of coffee farms from the CNA that I use to study income recovery in Section 5.1 and 5) the sample of long-term loans that I use to study repayment recovery in Section 5.2. In this section of the Data Appendix, I explain the construction of each of these samples in detail.

**Loan Sample:** This sample consists of all the loans for coffee production disbursed by the BAC in the period of 2005–2011. Although I have information on loans disbursed in 2011–2015, in order to construct loan default outcomes I need to observe repayment behavior in subsequent years. Therefore, I restrict the sample to loans disbursed no later than 2011 to observe repayment at least five years after the loan was disbursed.

To obtain the Loan Sample, I first define a set of coffee farmers using the SICA. I consider an individual to be a coffee farmer if she is in the SICA farm data for 2006–2014. I merge the BAC and SICA data to obtain a set of loans disbursed to coffee farmers in 2005–2011. Of the resulting loans, I keep only those of individuals with a farm in the SICA in the year the loan was disbursed. If I observe multiple farms for the same person, I assign the largest farm to the loan. Farms and rainfall stations are geo-referenced. I find the closest rainfall station to the farm associated with the loan using the Euclidean distance.

To search for the closest rainfall station, I need to define a sub-set of stations from the IDEAM data. For each station, I observe monthly rainfall from 1982 to 2016. But not all rainfall stations have data going back to 1982, and information is missing in some months. The set of rainfall stations that I use in the paper have at least one observation in every year in 2006–2012 and at least one observation in 16 different years in 1982–2012. I also require that information on the latitude and longitude of the station is available.[39] The first restriction is helpful since most of the loans in the data were disbursed in the period 2006–2012. The second restriction guarantees that for the selected rainfall stations, there

[39] I lack geographical coordinates for only one rainfall station in the IDEAM raw data.

is sufficient historical information to construct a rainfall distribution that reflects long-term rainfall levels. There are 1,606 rainfall stations that meet these conditions.

Finally, from the set of loans linked to a rainfall station, I keep loans related to coffee production. There were a total of 498,000 loans of producers in the SICA in the period 2005–2011. Around 302,000 (61%) of these loans were related to coffee production and close to 274,000 can be linked to a rainfall station. Of these, 250,000 loans were granted to small or medium size farmers. The final sample has about 242,000 loans with non-missing information for all the variables of interest and about 129,000 farmers. Of the 1,606 selected IDEAM rainfall stations, around 1,200 are closest to at least one farm from the Loan Sample.

**Application Sample**: This sample consists of loan applications that follow an initial loan. To construct it, for each individual in the Loan Sample I find the most recent loan of those disbursed in 2008–2011. This results in a single loan for each coffee producer. Then I find the first application in the Bureau Reports data, which spans 2010-2015. For the date of the end of the initial loan, I use the date of loan disbursement plus the term of the loan. The resulting sample consists of a single observation for each coffee producer which consists of a loan and a subsequent loan application. Recall from Section 4.1 in the main text, that the Bureau Reports contain information on the results from queries to credit bureaus, done by BAC officers when a customer applies. The queries result in a report with a *Bureau Score* and, based on BAC policies, it also indicates if the application process can continue. These are the main variables of interest of the Application Sample.

**Individual-Level Panel of Applications**: This sample is an individual panel at a yearly frequency with information on application outcomes. To construct the panel, I start with the sample of coffee producers in the Loan Sample and find all their reports in the Bureau Reports data. For each individual in each year, I average across reports to obtain an average credit bureau score. I also define a dummy that equals 1 if any of the individual's applications were rejected. The resulting panel is unbalanced, so I keep only farmers who have at least one observation in two years.

**Sample of Coffee Farms from the CNA**: This sample is a census of coffee farms surveyed in 2014 by DANE. The unit of observation of the census is the Unidad de Producción Agropecuaria (UPA) [Unit of Agricultural Production]. A UPA is broadly defined by a plot of land that produces some type of agricultural output and has a unique decision-making entity. An UPA can have many dwellings, and a dwelling can be made up of several households. These distinctions are important for defining the variables below. I keep only UPAs of less than 10 ha that produced coffee in the 2013 harvest. I also require the UPA to have at least one dwelling and at least one household.[40] In the census data there are about 958,000 UPAs with crops. Of these, about 386,000 cultivate coffee and 249,000 have a household and a dwelling. Finally, about 208,000 UPAs have fewer than 10 ha and about 175,000 have no missing values in all the control variables that I use in this paper. This is my final sample of coffee farms from the CNA. To obtain rainfall shock measures, I use the geographical coordinates provided by DANE for each UPA to find the closest rainfall station from the set of 1,606 stations defined above.

**Sample of Long-Term Loans**: The sample of long-term loans is obtained by restricting the Loan Sample to loans with terms of five years or more. I use two versions of this sample: the Baseline Long-Term Loan sample that requires the farm linked to the loan to have at least three plots and coffee trees with an average age of over 4.5 years and the Full Long-Term Loan Sample that includes all loans of five years or more.

### B.2 Variable Definitions

In this section, I describe the variables of all the samples. See Table 1 for the list of variables in the Loan Sample and Application Sample, and Table A3 for the variables in the Sample of Coffee Farms from the CNA. The variables in the samples of long-term loans are the same as those in the Loan Sample.

#### Variables in the Loan Sample

*Distance to Rainfall Station (km)*: The distance in kilometers from the farm linked to the loan and the closest rainfall station in the IDEAM data

*Farm Size (ha)*: The size (in hectares) of the farm linked to the loan in the year of loan disbursement. This variable is obtained from the SICA data. If the farmer has multiple farms in the year of disbursement, I take the largest farm.

*Coffee Plot Size (ha)*: Hectares of the farm cultivated with coffee. I add the plots to obtain the total coffee area.

[40] I need characteristics of the dwelling and the household to define controls.

*Coffee Plot Age*: Average age of the farm's coffee trees. The age is available for each plot of the farm. I average age across plots.

*Coffee Plot Density*: The number of trees (in thousands) per hectare of the farm.

*Coffee Plot Shade*: For each plot of the farm I observe the fraction cultivated near larger trees that provide shade to coffee trees. I take the average across plots to obtain the average fraction of shade of the farm.

*Coffee Plot Caturra*: Average fraction of the farm cultivated with the Caturra coffee variety. Coffee tree varieties differ in terms of productivity, susceptibility to diseases, etc. I take the average fraction of Caturra across plots.

*Shocks*: The number of rainfall shocks in the first year of the loan life cycle. To construct this variable, I start with the monthly IDEAM precipitation data from the closest rainfall station to the farm linked to the loan. I then add monthly rainfall to obtain total rainfall for each quarter-year from 1982 to 2016. This results in about 35 observations (depending on data availability) for each quarter (q1, q2, q3, q4) and each rainfall station. For each quarter, I obtain the 80th percentile of rainfall at each station. For each quarter-year and each rainfall station, I define a shock as having occurred if the rainfall was above the 80th percentile of the corresponding quarter distribution (which is specific to the rainfall station). *Shocks* ranges from 0 to 4.

*Shocks (p. 90)*: This variable is analogous to *Shocks*, but I use the 90th percentile of the rainfall distribution to define rainfall shocks.

*Shocks* $\geq$ *2*: A dummy equal to 1 if *Shocks* is 2 or more.

*Loan Size*: The total amount of Colombian pesos of 2010 disbursed by the BAC and received by the farmer.

*Interest Rate*: The annual interest rate of the loan.

*Maturity*: The maturity (or term) of the loan which is predetermined at the time of the loan disbursement. Depending on whether the farmer repays the loan on time, the loan may be observed in the bank data for more or less time than its maturity.

*Special Programs:* Some of the bank's loans are disbursed as part of government programs created to help farmers, for instance during periods of low prices or adverse weather conditions. Usually, they involve different conditions than normal loans (for example, lower interest rates). The Special Programs dummy takes a value of 1 if the loan belongs to one of these programs.

*Agricultural Type:* The bank lends money for activities other than agriculture. A small fraction of the loans in the Loan Sample (2%, according to Table 1) are non-agricultural loans. Agricultural is an indicator dummy for loans that fall into this category.

*Loan Lines*: Loan lines are broad groups of loans with similar characteristics such as interest rate or maturity. I observe more than ten lines in the data, but four lines make up 99% of the loans in the sample: "Working Capital Line" (Loan Line 1), "Replacement of Old Coffee Trees" (Loan Line 2), "Investment in New Coffee Trees" (Loan Line 3) and "Investment in Equipment and Commercialization" (Loan Line 4). I group the remaining loan lines into the same group. In the baseline regressions (for example, in Table 2), I include dummies for each of the four main loan lines.

*Overdue 30 and 60*: dummies equal to 1 for loans that were overdue, at any point in time, by 30 days or more or 60 days or more.

*Loan Score Fell*: The financial authorities require the bank to report a score for each outstanding loan to credit bureaus every month. This score is based mainly on the repayment behavior of the customer, starting with A (highest) when the loan is disbursed and ranging down to E (lowest). *Loan Score Fell* is a dummy for loans that had a score other than A at any point during the loan tenure.

*Loan Score Fell to E*: A dummy equal to 1 for loans that had a Loan Score that fell to the lowest score possible, E.

**Variables in the Application Sample**

*Applied*: A dummy for loans followed by a subsequent application in the Bureau Reports. To construct this variable, for each farmer I find the most recent loan disbursed in the period 2008–2011. This dummy is equal to 1 if I observe any application in the credit bureau data in the period 2010–2015 that is after the end date of the initial loan. The end date of the initial loan is the date of disbursement plus the term of the loan.

*Time Lapse*: The number of years between the end of the initial loan and the next application in the Bureau Reports.

*Bureau Score*: The score reported by the credit bureau at the time of the application.

*Denial*: A dummy equal to 1 if the loan application was rejected.

**Variables in the Individual-Level Panel of Applications**:

*Bureau Score*: The individual-level panel is a yearly panel. In this sample, the bureau score is the average of the bureau scores across an individual's applications in a given year.

*Denial*: A dummy equal to 1 if any of the applications submitted that year were rejected.

**Variables in the Sample of Coffee Farms from the CNA**:

*Distance to Rainfall Station*: Each UPA in the CNA census data is georeferenced. *Distance* is the distance in kilometers to the closest rainfall station in the IDEAM data.

*Farm Size*: The total size of the UPA in hectares.

*Health Insurance*: As described above, an UPA can have multiple dwellings and households. Health Insurance is a dummy equal to 1 if at least one household head in an UPA has any type of health insurance.

*Education Dummies*: Among household heads of the UPA, I find the one with the highest level of education. Then I create a set of dummies to indicate her education level: no degree, primary and middle school, high school, technical education, college, and graduate studies.

*Wall Quality*: A dummy equal to 1 if at least one of the household's dwellings has high-quality walls, defined as constructed primarily of one or more of the following materials: concrete block, bricks, stone, polished wood, "tapia pisada", adobe, or "bahareque".

*Floor Quality*: A dummy equal to 1 if at least one of the household's dwellings has high-quality floors, defined as constructed primarily of one or more of the following materials: carpet, marble, parquetry, polished wood, tiles, brick, ceramic, concrete or gravel.

*Aqueduct*: A dummy equal to 1 if at least one of the dwellings of the UPA is connected to an aqueduct.

*Sewerage*: A dummy equal to 1 if at least one of the dwellings of the UPA is connected to a sewerage system.

*Electricity*: A dummy equal to 1 if at least one of the dwellings of the UPA is connected to an electricity grid.

*Coffee Income*: Total income generated from coffee production. It is the total quantity (in kilos) of coffee produced by the UPA in the 2013 harvest multiplied by a price of 3,730 pesos.

## C. Exogenous Shocks in a Model of Borrower Screening

In this appendix, I show with a simple model of borrower screening that the inclusion of information of exogenous and transitory shocks in a "credit score" leads the lender to make mistakes less frequently. A mistake is defined as lending to a non-profitable borrower or denying credit to a profitable one. This framework has common features with the model of de Janvry et al. (2010).

Consider a scenario with two periods, $t-1$ and $t$, 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 she grants this subsequent loan. The borrower is characterized by a level of profitability $\pi_0$ which is not observed by the lender.

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

$$\pi_{t-1} = \pi_0 + z + \epsilon \tag{A1}$$

where $z$ is an exogenous shock (in the sense that it is independent of $\pi_0$) and transitory (in the sense that it only affects repayment in $t-1$). $z$ is potentially observable to the lender. $\epsilon$ is another exogenous component (independent of both $\pi_0$ and $z$) but unobservable to the lender. I assume that $z \sim N\big(0, \sigma_z^2\big)$ and $\epsilon \sim N\big(0, \sigma_\epsilon^2\big)$. 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, she 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 her decision) so that $\pi_t = \pi_0$. Therefore, the lender makes a positive profit in the second loan if $\pi_0 > 0$ and a 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, she forms a "credit score" based on $\pi_{t-1}$. In the case where the exogenous shock $z$ is not observed, this credit score is given by:

$$E[\pi_t|\pi_{t-1}] = \pi_{t-1} \tag{A2}$$

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 two facts documented in the paper, that shocks affect repayment of current loans and that they only have a transitory effect on the level of income and the repayment ability of coffee producers. 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 the lender grants a loan to an unprofitable 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 shock $z$ is not included in the credit score and one in which it is.

**Scenario 1: $z$ is not included in the credit score**

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 $N\big(0, \sigma_z^2 + \sigma_\epsilon^2\big)$ 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

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

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, $\pi_0$.

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:

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

Given that $\pi_0 > 0$, this expression is also increasing in $\sqrt{\sigma_z^2 + \sigma_\epsilon^2}$. Again, the probability that the 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 included in the credit score**

Suppose $z$ is observable and can be included in the credit score by the lender. Then her prediction of $\pi_t$ changes. In particular, since she knows the process generating $\pi_{t-1}$ (given by equation (A1)) she 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 $P_u = P(E[\pi_t|\pi_{t-1}, z] \geq 0) = P(\pi_0 + \epsilon \geq 0)$ with $\pi_0 < 0$. Therefore,

$$P_u = \Phi\left(\frac{\pi_0}{\sigma_\epsilon}\right) \tag{A5}$$

with $\pi_0 < 0$. The probability of denying a loan to a profitable borrower, $P_d$ in this scenario is given by $P(\pi_0 + \epsilon \leq 0)$ with $\pi_0 > 0$, that is:

$$P_u = \Phi\left(\frac{-\pi_0}{\sigma_\epsilon}\right) \tag{A6}$$

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, $P_u$, is larger in Scenario 1 (equation A4) than in Scenario 2 (equation A6). The difference in the mistake probabilities is larger the larger is the variance of the exogenous 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 $z$ is included in the credit score.

These results imply that the inclusion of exogenous shocks in the credit score increases its precision and reduces the probability of lender mistakes. The gains from this inclusion are larger if the variance of the shocks is large.