Paying the Poor to Save: The Long-Lasting Effect of Premiahorro

Pavel Luengas-Sierra[†]

September 29, 2023

(Find the latest version here)

Abstract

Can paying people to save help them to save and does the effect persist once payment stops? I assess the effect of Premiahorro (2009-2015), a large-scale program for low-income people in Mexico. To help people save, Premiahorro combined three features: financial training, a match, and a flexible commitment savings strategy. Identification relies on monthly administrative records from the bank that offered the program. I estimate differences-in-differences on a sample of participants matched with an equal number of non-participants. To choose the propensity score model, I use post-double lasso selection to parse over 276 variables that account for a wide-array of savings constraints. While it was active, Premiahorro increased savings balances by 48 percent and the number of deposits by 200 percent. After it was phased out, the effects persist, savings balances increasing by 66 percent and the number of deposits by 38 percent. Participants kept at the bank two-thirds of what they saved and received in matches. As they were taught, they saved most of the money, increasing their resilience against shocks. By combining three features instead of providing each on its own, Premiahorro helped people to save. And once it was phased out, its positive effect persisted.

Keywords: Matched savings; Financial literacy training; Commitment savings strategies; Post-double lasso selection; Rigorous lasso; Regression discontinuity design; High-frequency and high-volume financial records.

JEL Codes: C23, C53, C55, D14, I22

[†]University of Oxford. email pavel.luengas-sierra@economics.ox.ac.uk, pavel@pavellsierra.com

1 Introduction

Savings help people to smooth consumption, to resist shocks, and to invest timely (Karlan et al., 2014). Given how important savings are, can public policies help people to save? And when they do, and once they are phased out, can their positive effects persist? I evaluate the effect of Premiahorro, a matched savings program for low-income individuals created and implemented by The National Savings and Financial Services Bank (Bansefi), a Mexican government institution.[1] From 2009 to 2015, around 180,000 people participated in the program. To help people save, Premiahorro provided three features. First, it provided a match, a payment once participants reached a savings goal. Second, it provided financial training. And third, by the way it worked, it provided a flexible commitment savings strategy. Using differences-in-differences on a sample of participants matched with an equal number of non-participants, I estimate the effect before and after Premiahorro ended. Outcomes are monthly savings balances and number of deposits from a census financial records. Matched non-participants come from all bank clients in branches that did not offer the program. Matching relies on post-double lasso selection. To select the propensity score model, post-double lasso selection parses over a long list of carefully selected variables. Matching also relies on trends of outcomes of the 18 months preceding participation. While it was active, I find, Premiahorro increased savings balances by 48 percent and the number of deposits by 200 percent. After it was phased out, the effects persist, savings savings balances increasing by 66 percent and the number of deposits by 38 percent. Premiahorro helped bank clients to save. And once it was phased out, its positive effect persisted.

Premiahorro worked in calendar-year cycles. Each trimester within a cycle, participants had to deposit at least 15 USD in a locked savings account. If they did, the bank deposited in the account a match. At the end of the trimester the bank allowed participants to opt-out. If they did, facing no penalty, they dropped from the calendar-year cycle and withdrew their savings and the match. But if they opted-in for a new trimester, the match increased. By locking savings within trimesters, Premiahorro was a hard commitment savings strategy. But by allowing participants to leave between trimesters, it was a flexible one. At the end of the calendar-year cycle, a participant deposited at least 60 USD and received 37 USD in matches, equal to a 62 percent interest rate.

By combining the match with financial training and with flexible commitment, Premiahorro

[1] Bansefi is a large bank; it managed around 5-7 million accounts in 2010. The same year, two large banks, BBVA Bancomer and Banco Azteca, managed 7 million and 10 million of accounts.

helped people to save. A long strand of economic literature finds that commitment devices can help people to achieve their goals (Bryan et al., 2010). Karlan et al. (2014) argue that few people use commitment savings devices because the cost of commitment outweighs the benefit, especially for those who are poor and who may need their savings in case of an emergency. Further, commitment savings devices can reduce welfare; for example, when people pay the cost of commitment but cannot save (John, 2020). Premiahorro charged neither fees for participation nor penalties for dropping-out. In this way, it reduced possible negative effects on welfare and the cost of commitment. It also raised the benefit of commitment in two ways: by granting the flexibility of dropping out and by increasing each trimester the match. By reducing the cost and raising the benefit of commitment, Premiahorro was popular. Most participants took part in three or more calendar-year cycles. Bansefi each year offered 65,000 accounts, the maximum its annual budget allowed. By late February, almost all accounts had been opened.

The most salient feature of Premiahorro was its generous match. The match equals a 62 percent annual interest rate, one well beyond the market interest rate. Such a market distortion is only warranted if it has a long-term effect, if it leads to the habit of saving (Karlan et al., 2014). Schaner (2018) finds that large interest rate subsidies can create lasting welfare impacts that continue years after the subsidy has been removed. For Premiahorro, financial training helped on this score. Evidence on the positive effects of financial training on financial literacy and behavior is increasing (Kaiser and Menkhoff, 2017). Simple and focused financial training can be effective when combined with offers of financial products (Karlan et al., 2014). Premiahorro participants received simple and focused financial training. A 90-minute financial training workshop covered how Premiahorro works, what to save for (emergencies) and how much to save (the fund for emergencies should be at least three times monthly expenses). Before and after the workshop, trainees were surveyed. According to information from the surveys, financial training was effective. It primed saving for emergencies—precautionary savings—as the main goal of Premiahorro. Financial training helped to cement the habit of saving.

Identification relies on administrative records. From August 2007 to March 2017, the monthly-level records cover all participants in branches that offered the program and all clients in branches that did not. Bansefi offered Premiahorro in half of its branches, using for assignment an arbitrary, strictly enforced rule. Based on the population of the village or city in which the branch is placed, the rule aids identification.[2] Similar to Lara Ibarra

[2] All publicity of the program was at the village or city of the branches that offered the program—all publicity

et al. (2017), I estimate a difference-in-differences in a sample of participants matched to an equal number of non-participants.[3] The sample of participants are all who had a savings account with Bansefi for at least 18 months before they participated in Premiahorro (n=17,000). Each participant is matched with one non-participant counterpart from a pool of 190,000 clients of Bansefi in branches that did not offer the program. I calculate the propensity score using the post-double least absolute shrinkage and selection operator (lasso), a regularized regression method proposed by Belloni et al. (2014). To choose the tuning parameters, I use 'rigorous lasso' (Belloni et al., 2012). Post-double lasso parses over 276 carefully selected variables. Selection of the variables follows a taxonomy of savings constraints proposed by Karlan et al. (2014). Among the variables are distance to the bank, length of time as a client, and community characteristics like poverty, education, access to financial services, and proxies for female bargaining power. To select each nearest neighbor, I use both the propensity score and the trends of outcomes of the 18 months preceding participation.

Premiahorro helped people to save. From the 18 months preceding participation to up to 86 months after it, Premiahorro increased savings balances by 2119 Mexican pesos ($SE = 151$), an amount equal to 158 USD.[4] The increase equals 52 percent of the mean of savings of matched counterfactuals (effect size=0.16). For the number of deposits, the effect is larger. Premiahorro increased the number of deposits per month by 0.21 ($SE = 0.01$), a large increase of 162 percent (effect size=0.44). I then adapt the estimating equation to divide the effect of Premiahorro between the one it had when it was active and the one it had after it was phased out. Results for savings show that when it was active, Premiahorro increased savings balances by 145 USD ($SE = 10$; equal to 48 percent; effect size=0.15). After it was phased out, the effect is higher at 204 USD ($SE = 16$; equal to 66 percent; effect size=0.21). Results for the number of deposits show that the effect waned but persisted.[5]

was local. Spillover effects are unlikely. Around 30 kilometers is the median Euclidean distance from each branch that did not offer the program to the closest branch that did. In comparison, around 4 kilometers is the median distance between each bank client and her branch.

[3] Lara Ibarra et al. (2017) study the effect of financial training on responsible use of credit, measured by responsible use of credit cards. They find no results in their experimental setting. Their randomized control trial lacks power owing to a very low take-up rate, a pervasive problem in financial training studies. To overcome the problem, they exploit their data, administrative records with high frequency and high volume. Using regularized regressions, they match people who were offered and received training with people who were not offered training. They then estimate difference-in-differences in the matched sample. They find that trainees were more likely to pay their cards on time and to pay more than the minimum.

[4] The exchange rate is 13.4 Mexican pesos per dollar. It corresponds to the mean exchange rate between 2009 and 2015.

[5] This is consistent with the design of the program. Financial training taught participants to save a specific amount for emergencies. Once participants reached the target, they likely stopped making deposits or made fewer.

When it was active, Premiahorro increased the number of deposits by 0.26 (SE = 0.01; equal to 200 percent; effect size=0.54). After it was phased out, the effect careened to 0.05 (SE = 0.01; equal to 38 percent; effect size=0.11). While it was active, Premiahorro helped bank clients to save. After it was phased-out, its positive effects on both savings and the number of deposits persisted.

The Premiahorro account could have crowded out the use of other savings accounts at the bank. For this reason I estimate, as additional result, the effect it had on active use of other savings accounts. I use two measures of active use of accounts: making at least one deposit over the last six months or making at least two. In the first definition (≥1 deposits), Premiahorro increased active use of other savings accounts by 14 percentage points (SE = 1.1). The increase equals 64 percent over matched counterfactuals, of which 22 percent used the account actively (effect size=0.33). In the second (≥2 deposits), it increased active use by 9 percentage points (SE = 0.9). The increase equals 60 percent over matched counterfactuals, of which 15 percent used the account actively (effect size=0.26). Effects during and after the program are similar. Point estimates are lower after the program was phased out (14 vs 12; 9 vs 7) but confidence intervals largely overlap. When it was active or when it was phased-out, Premiahorro increased active use of other savings accounts at the bank.

Results from a regression discontinuity design strategy complements the main results. Branch-level assignment follows a sharp regression discontinuity design. Bansefi offered the program only in branches placed in villages or cities with fewer than 50,000 inhabitants. I estimate the effect of Premiahorro on the threshold of the assignment rule. I find at the branch-level large increases in the number of accounts and transactions. The results are consistent with the individual-level results. However, branch-level results are imprecise and not statistically different from zero. Because the number of branches is relatively low (around 400), and because many of them are away from the threshold, the design is under-powered. Power calculations I present show that the sample needed to detect with 0.80 power plausible effects in the number of accounts and of transactions far exceeds the size of the current sample. It is two to three times more than all branches the bank has—and six to nine times the number of branches around the threshold.

Matched subsidies, financial training, and commitment savings strategies—all in their own can help people to save. The paper contributes to the literature by showing that combining the three features could overcome low take-up rates when each feature is offered on its own. Another contribution is showing that the positive effects of offering a match—a costly feature—can persist after it is phased-out. Implications of the results, however, are limited. An open question is which combination of the three features is more effective. And the results are limited to a specific sub-sample—current bank clients who have been with the bank for more than a year-and-a-half—and to a specific outcome—formal savings. Results do suggest that Premiahorro likely strengthen a downstream outcome, resilience against shocks. Participants on average took part in 3 cycles, accumulating 290 USD of which 110 USD they received in matches. Estimates of the effect on the period after the program was phased-out suggest they saved 200 USD they would have not saved had they not participated.[6] As they were taught, they saved most of the money.

The paper proceeds as follows. Section 2 explains eligibility requirements and the three features of the program—the match, flexible commitment, and financial literacy training. It closes by providing descriptive statistics for the adoption of the program. Section 3 describes information sources (financial records at the branch- or at the individual-level) and additional information used to calculate indicators for the villages or cities where branches are placed or where people live. It also describes the sample of branches and of individuals. Section 4 details the difference-in-differences identification strategy. It describes the estimating equation, outcomes, and control variables; also, it carefully explains the process to select counterfactuals. Section 5 provides the main results— results at the individual-level from the difference-in-differences—while section 6 provides supporting results—results at the branch-level from the regression discontinuity design. Section 7 concludes.

2 The Matched Savings Product (2009-15)

2.1 Design: Match and Commitment Features

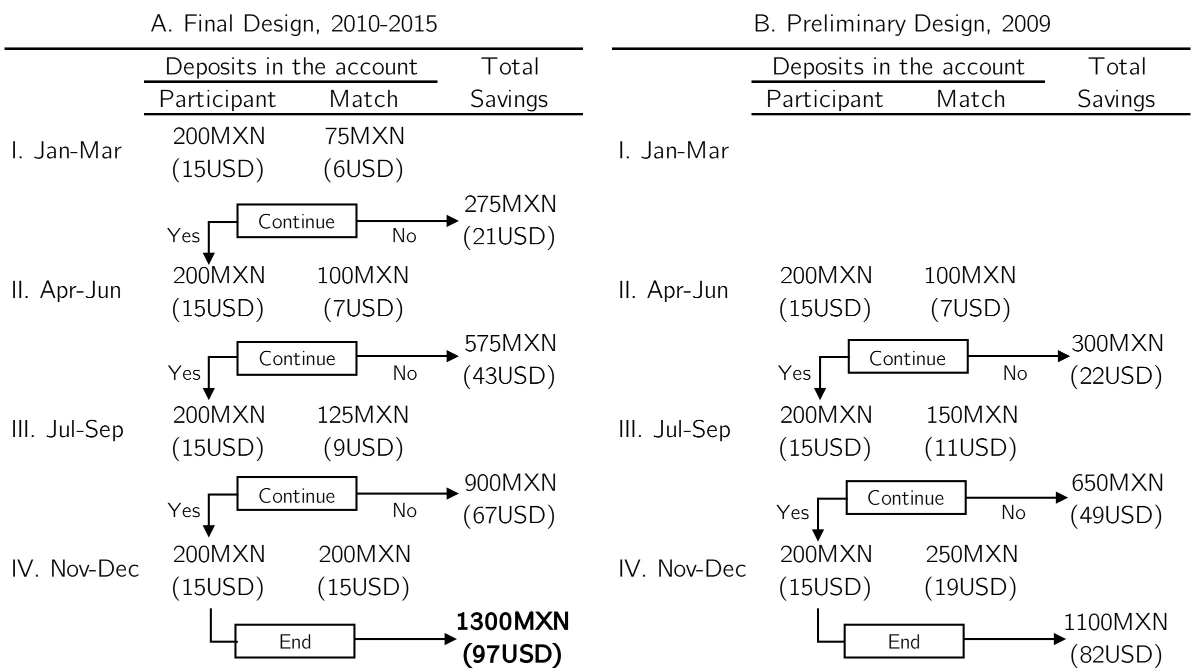
Premiahorro worked in calendar-year cycles, which worked in four trimester cycles. In January selected branches began offering accounts, beginning the year cycle. The participant received her account after receiving program details and providing proof of identity and residency. Once she deposited at least 200 MXN (around 15 USD), the bank deposited in her account a match, a fixed amount regardless of how much she deposited. The account locked her deposits and the match until the trimester lapsed. At the end of the first trimester, in early April, she decided either to exit or to continue. To continue, she deposited at least another 200

[6] Refer to column 2 in table 7. The point estimate of 2738.04 MXN equals 204.33 USD using an exchange rate of 13.4 Mexican pesos per dollar (mean between rate 2009 and 2015).

MXN within the first 15 days of the second trimester; the bank deposited a second, higher match.[7] To exit, she made no deposit. At the end of the second trimester, in early July, she once more decided either to exit or to continue. If she continued, the bank deposited a third, higher match. At the end of the third trimester, in early October, she made a last decision. If she continued, she received a fourth and last, higher match. At the end of the fourth trimester, the calendar-year cycle ended.

Figure 1 describes the design of Premiahorro. Panel a describes the final design it had from 2010 to 2015, the year it ended. Panel b describes a preliminary design it had in 2009. From 2010 to 2015, a participant locked savings for a year, saved at least 800 MXN (60 USD), and received 500 MXN (37 USD) in matches, equal to an interest rate of 62 percent. The 1300 MXN (97 USD) in savings is 11 percent of the monthly income of a typical household. But for households in the bottom quintile of income, it is 46 percent.[8]

Figure 1: Design of Premiahorro



Exchange rate: 13.4 Mexican pesos per dollar (mean between 2009 and 2015). The program started with a preliminary design in May 2009.

Premiahorro for a year rendered savings illiquid, making it a hard commitment savings

[7] If natural disasters or local emergencies struck the community, the 15-day grace period was longer.

[8] Own estimations using the National Survey of Household Income and Expenditure 2010 (ENIGH).

strategy. But it was a flexible strategy. If at the end of a trimester a participant opted-out, she faced neither penalties nor restrictions on participation in further cycles. And she could withdraw anytime after the grace period the deposits and matches accrued. After opting-out, the account lay dormant. She could leave money on it as long as she wished, but making deposits was only possible until the next calendar-year cycle. Leaving money in the account kept, free of charge, savings safe.

2.2 Eligibility: Anyone Could Open an Account at Selected Branches

Anyone could open an account, but only at selected branches. A prospective participant had to meet a few 'know-your-client' requirements. To join, she had to be at least 18 years old, to show a valid ID (such as a voting ID), and to provide proof of residence (such as a bill dated within the last three months). After submitting the documents, the participant opened the account. Making deposits and withdrawals was only possible at the branch that provided the account. Bansefi allowed only one account per year cycle and imposed a crucial requirement: participants ought to attend its financial literacy training.

Half of the branches offered Premiahorro accounts. To select which half, Bansefi followed an arbitrary rule. Bansefi based the rule on the population of the village or city in which the branch is placed.[9] Table 1 summarizes the rule. In the final design (panel a), a branch offers Premiahorro, PA_b, if the population of the village or city (x_b) in which the branch is placed is strictly below 50,000. In the preliminary design (panel b), the population threshold was lower at 20,000, but not all branches offered Premiahorro. Instead, an unknown function ($g(.)$) of poverty (P_b) of the village or city mediated which branches did.[10] Of the 487 branches of Bansefi, 109 offered in 2009 the preliminary version. The program expanded to 149 branches in 2010. From 2010 to 2015, 258 branches offered Premiahorro.

2.3 Financial Literacy Training: Priming for Precautionary Savings

Bansefi provided financial literacy training participants ought to have attended. When a participant opened her account, she signed a commitment letter. The letter prescribed her to attend training that would occur in the village or city of the branch. A participant who

[9] Villages or cities are Mexico's third political division (states are the first; municipalities, the second). Each municipality comprises 'localities', which according to population either are villages or cities. Villages are localities with fewer than 2,500 habitants. The official definition of rural areas is the population inhabiting villages.

[10] Offering no details, official documents declare that "priority was given to branches placed in poor villages or cities."

Table 1: Regression Discontinuity Rule
Probability a Branch of the Bank Offered Premiahorro

A. Final Design, 2010–2015	B. Preliminary Design , 2009
$PA_b = \begin{cases} 1 & \text{if } x_b < 50,000 \\ 0 & \text{if } x_b \geq 50,000 \end{cases}$	$\Pr(PA_b = 1) = \begin{cases} g(P_b) & \text{if } x_b < 20,000 \\ 0 & \text{if } x_b \geq 20,000 \end{cases}$

x_b is the population of the village or city in which the branch is placed and $g(.)$ is an unknown function of poverty of the village or city (P_b).

wanted to enroll in a new calendar-year cycle would have to submit a stub given at the workshop. (Whether Bansefi requested the stub is unclear.) Bansefi provided workshops for 227 of the 258 Premiahorro branches.[11] From 2009 to 2013, 1,800 workshops per year took place in villages or cities with Premiahorro branches. In public squares and in other central locations, 90-minute workshops trained between 20 to 50 people. People learned to use Premiahorro, to reduce expenses, and to match financial needs with credit, insurance, and savings products. Around 263,000 people received financial literacy training.

Although anyone could attend training, most attendees were bank clients. Bansefi advertised the workshops locally. It asked bank tellers to tell clients and hired drivers to advertise with megaphones. At the beginning of each workshop, Bansefi collected information from attendees. Besides providing name and date of birth, attendees told whether they were bank clients and which savings products they used. Using information from the 2010 workshops (the only year for which I received information), I find that most attendees were bank clients. Of all attendees, 52 percent already had a Premiahorro account; 23 percent, an Oportunidades account; 21 percent; other account. Only 4 percent had no account with Bansefi.

Financial training should be short, simple, and focused. Karlan et al. (2014) argue that this type of financial training can be effective, especially when attached to offers of financial products. Kaiser and Menkhoff (2017) provide updated evidence. Their meta-analysis of 76 randomized control trials finds that financial training improves financial literacy and financial behavior. Of the financial behaviors they analyze, saving and budgeting improved the most. The workshops were short, simple, and focused. The appendix details the protocol of the 90-

[11] To lower costs, Bansefi avoided remote villages or cities, unless they had many Premiahorro participants. Using enrollment information, I find that 97 percent of participants opened accounts in the 227 branches receiving workshops.

minute workshop (section A.1.2). In the first half-hour, attendees learned about Premiahorro and type of expenses. The facilitator imprinted on attendees a nugget of information: the best way to save is to reduce expenses. In the second half-hour, they learned how to set savings goals and how to match expenses with financial products. The facilitator imprinted a second nugget of information: set an emergency fund that equals three times your expenses in a typical month. In the third and last half-hour, they learned how life insurance works.[12]

Training informed attendees effectively.[13] After training, attendees were more likely to say they would reduce expenses and would set a emergency savings fund. Facilitators surveyed in each workshop between 3 and 5 attendees at random, asking them six multiple-choice questions. Attendees responded to the survey twice, at the beginning and at the end of the workshop. Figure 2 shows descriptive statistics from the surveys of the 2010 workshops. The two nuggets of information primed attendees successfully. Before the workshops, fewer than 33 percent said the best way to save is to reduce expenses. Also 33 percent said an emergency fund should equal three times expenses in a typical month. After the workshops, the proportions increased to 81 and to 88 percent. Before the workshop, 66 percent knew how Premiahorro worked. After it, all did. The workshop changed what attendees expected to do with their savings in Premiahorro. It halved the percent who said would use the savings to cover expenses (from 25 to 13 percent). It raised from 45 to 65 percent the percent who said would use the savings for emergencies.

2.4 Participation: Adoption Was Fast

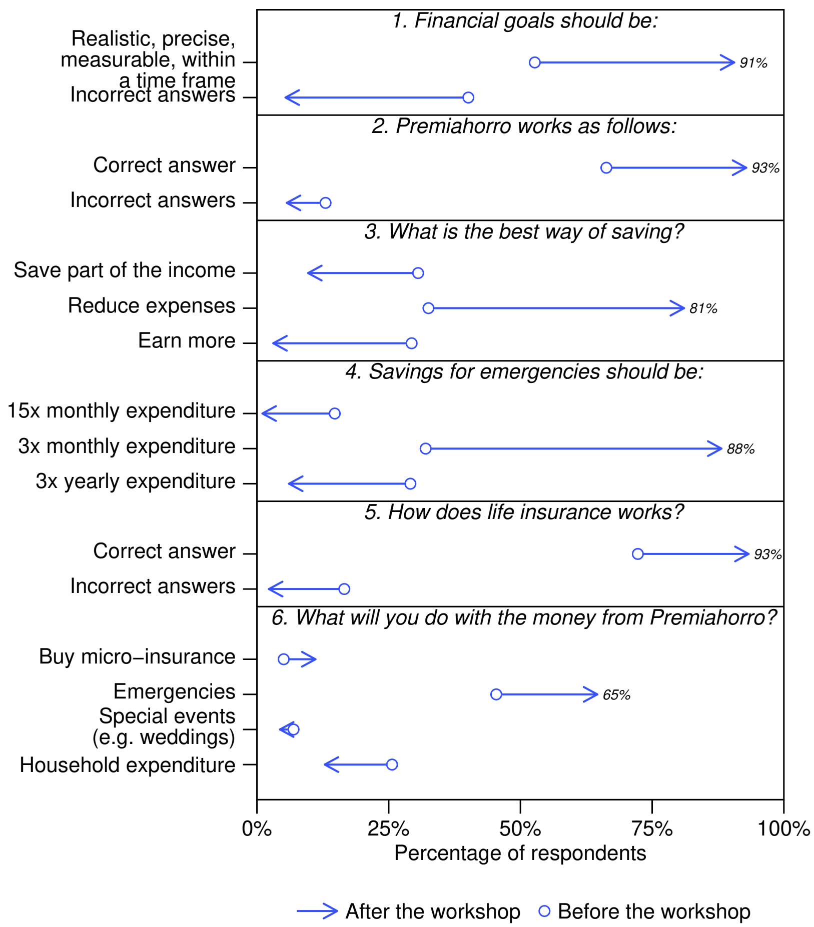
Around 184,000 people used Premiahorro. Figure 3 shows descriptive statistics of participation by year. Panel a shows by year the number of accounts opened. Bansefi offered around 65,000 accounts per year, except for 2009 when it offered 45,000 accounts. Across years the number of accounts offered was similar because the yearly budget was similar.[14] Between 2010 and 2015, fewer than 14 percent who opened an account in the year participated that year only. Most people participated three or more years, suggesting they were satisfied with the program. Panel b shows the year and trimester in which participants opened their first

[12] Why a third of the training focused on life insurance is unclear.

[13] The effect of financial literacy training cannot be disentangled from the other features of Premiahorro. Because 263,000 people attended the workshops while 184,000 people opened the premiahorro account (and because attending was a requirement), I assume that almost every participant in Premiahorro received training. In principle, attendees can be matched to participants of Premiahorro. Bansefi recorded the names and the date of birth of all who attended the workshops. However, only the information for the 2010 workshops was digitized and made available for my analysis.

[14] Bansefi closely monitored the number of accounts being opened. Once the number approached what the yearly budget allowed for, it stopped opening accounts.

Figure 2: Financial Literacy Workshop
Survey to Participants of the 2010 Workshops



The survey sampled at random of 7,042 (14.6 percent) of the 48,242 people who attended the workshops in 2010.

account. Of the 139,000 people who only joined the final design of the program (2010–15), 38 percent opened their first account in the first trimester of 2010. Across years, entering after the first trimester was rare.[15]

Adoption (take-up)

Evaluations of financial products often discuss take-up. For example, the take-up rate for

[15] Two reasons explain why. First, Bansefi followed the calendar-year cycle strictly. A person who opened an account on the second trimester received matches as she had entered during the first. Her cycle, however, also ended in December. She would receive at most 300 MXN (= 75 + 100 + 125) instead of the normal match of 500 MXN (= 75 + 100 + 125 + 200); see figure 1. Second, owing to a limited budget and to a high demand, Premiahorro stopped opening accounts early. Someone trying to enter late might not enter at all.

saving products with commitment features is low, at around 20 to 30 percent (Karlan et al., 2014). Take-up of matched savings accounts also is low. Duflo et al. (2006) find take-up to be 14 percent for an individual retirement account that matched contributions by 50 percent. Take-up of accounts with subsidized interest rates is higher. Schaner (2018) finds take-up to be 50 percent for a savings account that offered an annual interest rate of 20 percent. For Premiahorro, take-up is not well-defined. Bansefi each year offered around 65,000 accounts regardless of demand. And accounts were not offered to a specific group of people. Publicity was local, but anyone could open an account.[16]

Instead of take-up, I explore how fast Premiahorro was adopted. Figure 4 depicts Kaplan-Meier cumulative hazard functions.[17] Panel a refers to the 139,000 participants of the final design and panel b refers to the 45,000 participants of the preliminary design. Adoption was brisk. A participant of the final desing, on average, adopted Premiahorro in 63 days. Had Bansefi offered the program to a specific group of people, such a fast adoption suggests a high take-up rate.

[16] Premiahorro is more likely to be adopted by people living near the branches that offered it, owing to its local publicity and diffusion. According to official documents and interviews, Bansefi used four methods of publicity. The first two are the same used for the financial literacy workshops: through bank tellers and hired drivers. The third was banners in the village or city (Bansefi never used publicity on radio, television, or internet). The fourth was the network of beneficiaries of Oportunidades. During the first years of Oportunidades, Bansefi worked with a beneficiary in the village, a local leader, to help to disburse cash payments. The leader organized all beneficiaries, bringing them to a local point, at a certain time, to receive the cash payment. Bansefi relied on these leaders to diffuse information about Premiahorro.

[17] Lacking a well-defined population at risk, the estimates instead use all people who eventually opened an account that year.

Figure 3: Participation in Premiahorro

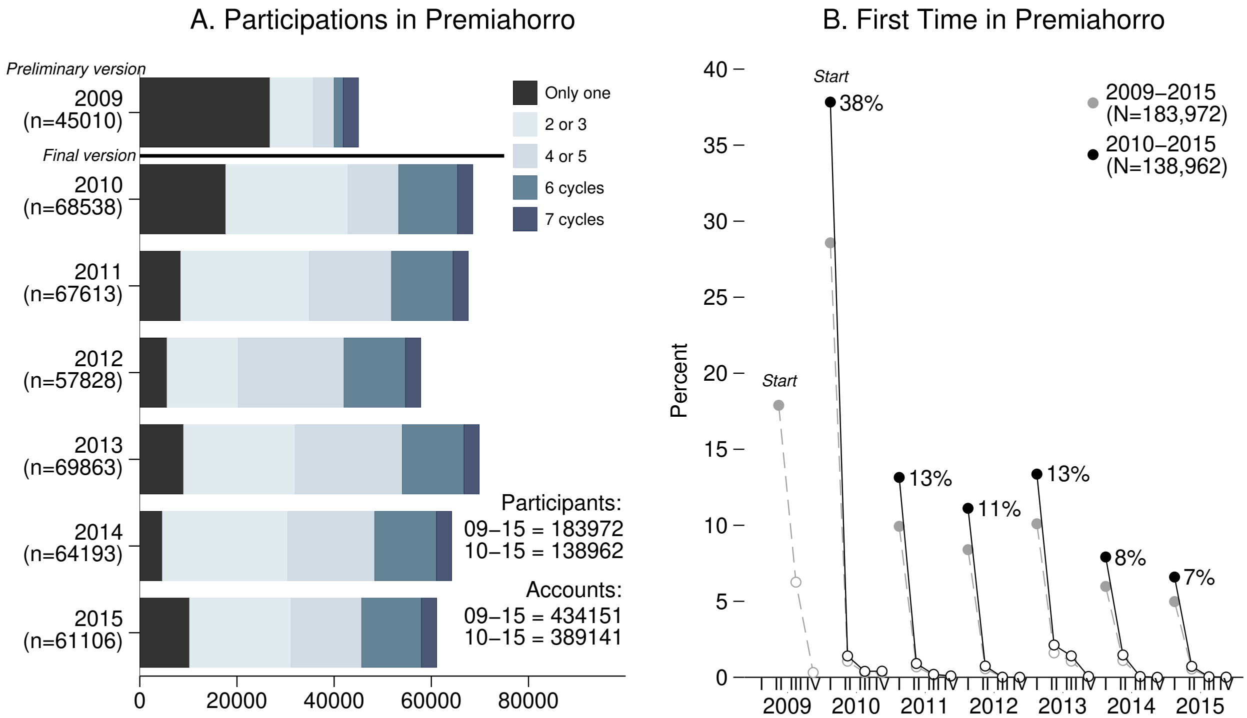
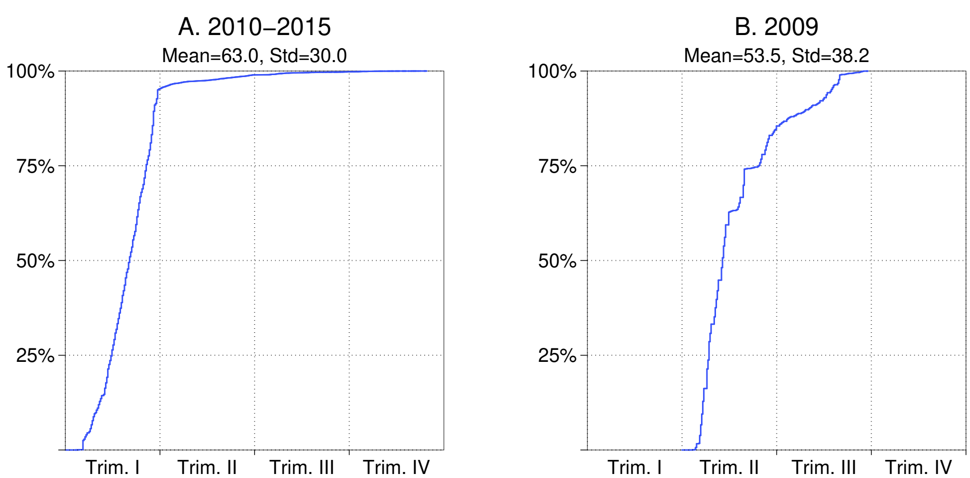


Figure 4: Days to Adoption
Cumulative Hazard Functions



In 2009 the program began in the second trimester.

3 Data

The main source of information is financial records at the branch or at the individual level. To complement the records, I create indicators at the village or city level that I match to the financial records. Information to create the indicators comes from population censuses, administrative records from Oportunidades, and administrative records from the banking and securities regulator (CNBV). Besides detailing sources of information, this section explains how I match village or city level indicators to the financial records.

3.1 Branches

Summaries of financial indicators are the main source of information at the branch level. Bansefi provided me a panel dataset with information for the 468 branches operating between 2010 and 2017.[18] The monthly panel covers all fifteen products of Bansefi from January 2008 to June 2017. Among all fifteen products, the main three products are: (i) a basic savings account, (ii) a basic savings account linked to a debit card, and (iii) a time-based commitment savings account.[19] Just as for Premiahorro, anyone can request any of three products (appendix A.1.1 briefly describes them). Of the secondary twelve products, eight link to social programs, Oportunidades being the main one.[20] The remaining secondary products are Premiahorro, an investment account, and two savings products, one for children and another for groups. Every product in the dataset has three indicators: total savings balances, number of accounts, and number of transactions.[21] For Premiahorro, savings balances exclude the match. I only change the information to remove outliers. For each product, indicator, and month, I winsorize the top 1 percent.

Premiahorro was a small product. It amounted to fewer than 5 percent of all accounts and to a paltry 0.6 percent of all savings balances of Bansefi. Figure 5 shows descriptive statistics at the branch level, subsuming the fifteen products into five categories. From top to bottom the categories are: (i) Premiahorro; (ii) the three main savings products; (iii) investment products, group savings, and children savings; (iv) savings accounts linked to

[18] Of the 487 branches operating in December 2008, 19 branches closed between 2009 and 2016.

[19] Starting in 2007, Bansefi began offering a basic savings accounts linked to a debit card. The new product, 'Debicuenta,' did not replace the traditional basic savings accounts, named 'Cuentahorro.' Clients with the traditional product had to request the upgrade, which only was available in selected branches.

[20] Bansefi opened and managed accounts for beneficiaries of Oportunidades. Besides accounts, Bansefi provided debit cards for beneficiaries. In rural areas, beneficiaries received a card they could use to withdraw the cash transfer at community-owned stores in remote areas. In urban and semi-urban areas, they began receiving in 2009 the basic savings accounts linked to a debit card.

[21] The number of transactions is the sum of deposits and withdrawals made by the client. The dataset lacks separate series for deposits and withdrawals.

socials programs, except Oportunidades; and (v) Oportunidades. Panel a focuses on the number of accounts; panel b, on savings balances. Premiahorro only had 5 percent of the accounts (16 percent of the accounts in the first two categories) and a paltry 0.6 percent of the balances (0.8 percent). Between 2010 and 2015, Oportunidades had most of the accounts (54 percent) while the basic savings accounts anyone can request had most of the balances (72 percent).[22]

Branches assigned to offer Premiahorro: Regression Discontinuity Design

In December 2008, five months before Premiahorro began, Bansefi served clients through 487 branches. Figure 6 depicts whether each end up offering Premiahorro. Panel a focuses on the preliminary design; panel b, on the final one. The x-axis represents for the village or city in which the branch is placed the population in the census 2005, the variable Bansefi used for assignment. Because most branches pile-up either in villages or in smaller cities, the x-axis uses a log scale. The y-axis denotes whether each branch offered Premiahorro. A horizontal line on each panel demarcates the population thresholds Bansefi used, 20,000 for the preliminary design and 50,000 for the final design.

Bansefi followed the arbitrary assignment rule strictly (see table 1). In the final design, all branches that could have offered Premiahorro offered it. Below the population threshold of 50,000 habitants are 263 branches. From 2010 to 2015, all but five branches offered it. (Bansefi closed these five branches in 2010.)[23] Above the threshold are 224 branches. None offered Premiahorro. In the preliminary design, half of the branches that could have offered Premiahorro offered it. Which branches did aligns well with prioritizing branches in poor villages or cities.[24]

Sample of branches

The sample of branches used in regression analysis is 395. Table 2 details the sample. I make two sample restrictions to the sample of 487 branches serving clients in December 2008.

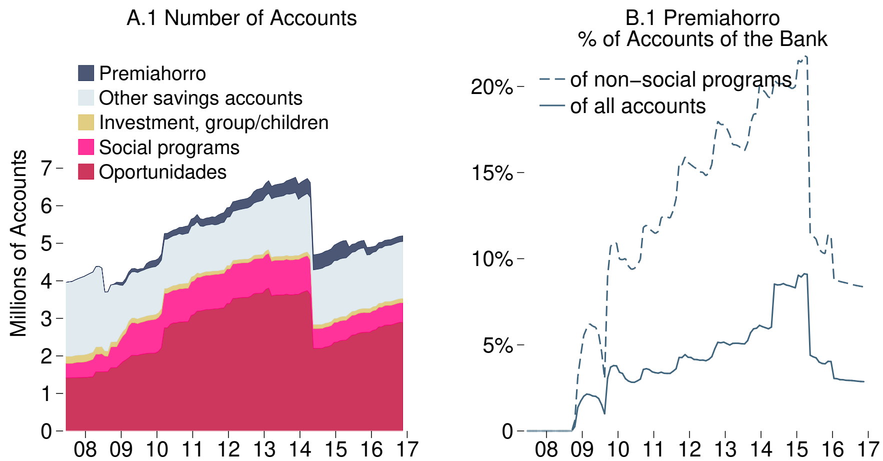
[22] Between 2014 and 2015, a new federal administration restructured Oportunidades and rebranded it Prospera. The restructure reflects on the financial records, which show a large decrease in the number of accounts while showing a similar total balance.

[23] Three closed in February, one in April, and one in November.

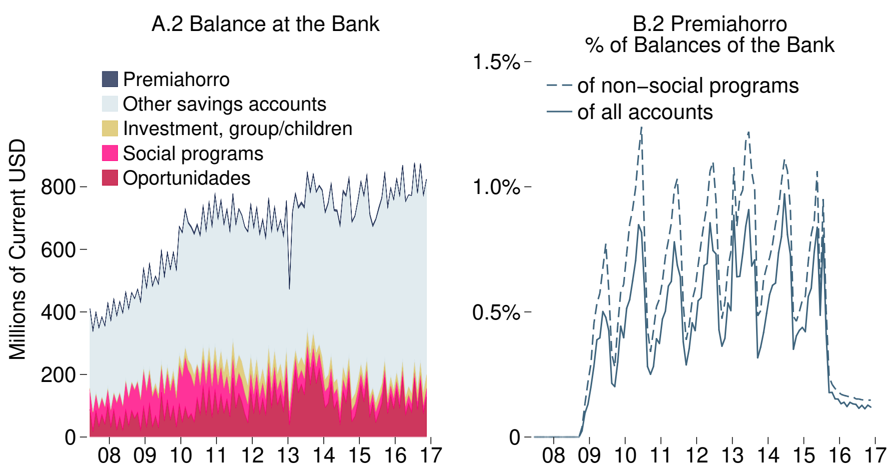
[24] A simple linear probability model with a poverty score from the 2005 census as the sole regressor explains 55 percent of the variation. An increase of one standard deviation in the poverty score correlates with an increase of 37 percentage points in the probability the branch offered Premiahorro (the mean is 52 percent, half of the branches). See figure 6.

Figure 5: Branch Financial Records
Number of Accounts and Total Savings Balances

A. Number of Accounts



B. Total Savings Balances

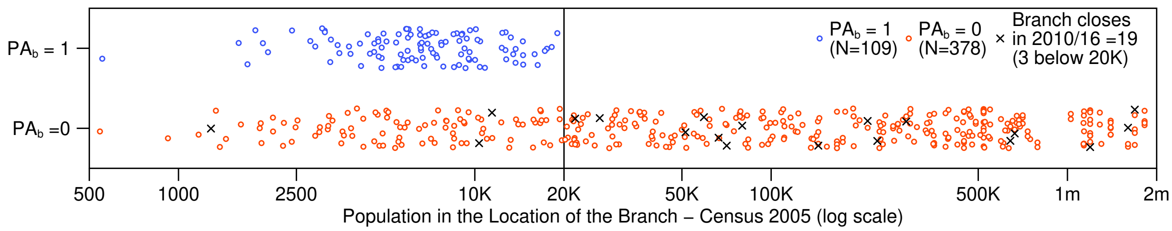


'Non-social programs' refers to Premiahorro and to the three main savings products of the bank, labeled in the graph as 'other savings accounts.'

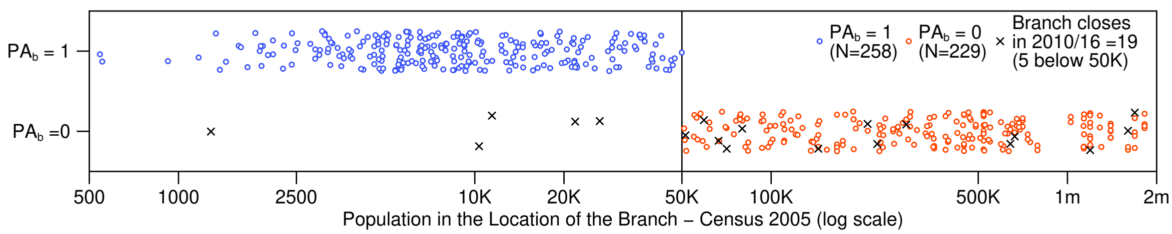
First, I drop branches Bansefi closed. Between 2010 and 2017, Bansefi closed 19 branches, all but four in 2010. Because it was a small product, Premiahorro likely is unrelated to Bansefi's decision to close the branches. The first restriction lowers the sample from 487 to 468 branchers. Second, I drop branches that sent or received accounts from other branches. In August 2015, Bansefi reallocated all accounts from 48 branches to another 25 branches. Reallocations muddle identification of the effect of the program, which had its last cycle in 2015. For example, in 2015 participants from reallocated branches had to travel to a different

Figure 6: Probability a Branch would Offer Premiahorro According to Population of the Village or City in Which the Branch is Placed

A. Preliminary design, 2009



B. Final design, 2010–2015



Sample: branches serving clients in December of 2008 (N=487). Between January 2009 and December 2016, 19 branches closed. Of the 208 branches placed in villages or cities below 20,000 habitants, 109 offered Premiahorro (52 percent) in 2009. Relation, below the threshold, of poverty with the probability of offering Premiahorro in 2009:
$PA_b = \underset{(0.11)}{1.62} + \underset{(0.09)}{1.03}\, Poverty_b + \varepsilon_{it} \quad R^2 = 0.55 \quad N = 205$, the poverty index ($Poverty_b$) has a mean of -1.05 ($SD = 0.36$). The sample excludes the three branches that closed. In brackets are standard errors.

branch, affecting their savings behavior. Further, savings balances in branches that received reallocated accounts artificially increased. The second restriction lowers the sample from 468 to 395 branchers. Because only 5 of the 48 branches that moved accounts and 3 of the 25 that received them offered it, Premiahorro likely is unrelated to the attrition of the 73 branches.[25] On whether the branch closed or reallocated or received accounts, Premiahorro barely loomed.

[25] Most branches that moved (30 of 48) or received accounts (11 of 25) were placed in cities with over 1 million habitants, a population range far away from the 50,000 population threshold.

Table 2: Sample of Branches for Analysis

Population (2005)	a. Original Sample: Branches	a. Original Sample: Unique mass points	b. Branches discarded: b1. Closures: Closed	b. Branches discarded: b2. Reallocations: Moved accounts	b. Branches discarded: b2. Reallocations: Received accounts	c. Sample for Analysis: Branches	c. Sample for Analysis: Unique mass points
[1 - 10,000)	157	157	1	5	0	151	151
[10,000 - 20,000)	51	51	2	0	1	48	48
[20,000 - 50,000)	55	54	2	0	2	51	51
[50,000 - 100,000)	39	38	5	0	1	33	33
[100,000 - 250,000)	49	37	3	4	4	38	33
[250,000 - 500,000)	52	31	1	9	6	36	28
[500,000 - 1m)	43	23	2	13	5	23	19
≥1m	41	11	3	17	6	15	8
Total	487	402	19	48	25	395	371

Unique mass points: Villages or cities with more than one branch are counted as a single observation. Identification of the supporting results relies on unique mass points because regression discontinuity design rely on them, making them its effective sample size.

3.2 Individuals

Monthly records of transactions and balances are the main source of information at the individual level. Bansefi provided me a panel dataset stripped from any confidential information. The monthly panel covers from August 2007 to March 2017 all accounts of Premiahorro and of the three main products of Bansefi. To identify each account, the panel includes anonymized account and client numbers. Individual-level information is more detailed than branch-level information. It includes start and end of the month savings balances as well as the number and amount of withdrawals and of deposits. All accounts report in separate variables payments made by Bansefi, including Premiahorro's match. Besides transactions and balances, for each client I received information on gender, date of birth, and postal code. The postal code corresponds to the address of the client when she opened an account at the bank. For each account, I received the account opening date.

I adjust the information in three ways. First, I discard accounts without a single transaction (either a deposit or a withdrawal) across all months in the long time series. Second, I recode from missing to zero all variables from the month an account closes to the end of the time series. Bansefi closed inactive accounts with a balance below the minimum (25 Mexican pesos). A value of zero for closed accounts means not having savings or not making deposits at the bank. Third, same as branch-level data, I change the information to remove outliers. For the universe of accounts and for each product, variable, and month, I winsorize the top 1 percent.

Participants in Premiahorro differ from typical clients of the bank

Participants in Premiahorro and typical clients of Bansefi differ widely. Figure 7 illustrates the stark difference. It compares two characteristics of the places where people live: poverty and distance to the branch. People in the sample the figure uses come from the 395 branches I examine. Participants are Premiahorro participants from the 250 branches below the cutoff (n=179,000).[26] Typical clients are all basic savings account holders (n=428,000) from the 145 branches above it. Panel a refers to poverty. It reports the proportion of people living in villages or cities within each of five levels of poverty. It compares three series: one for participants, another for typical clients, and a reference category for all people in Mexico. Panel b refers to distance to the branch. The distance corresponds to the Euclidean distance, a proxy for actual travel distance. The panel compares two densities: one for participants and another for typical clients. Because most people live close to the branch, the x-axis uses a log scale.

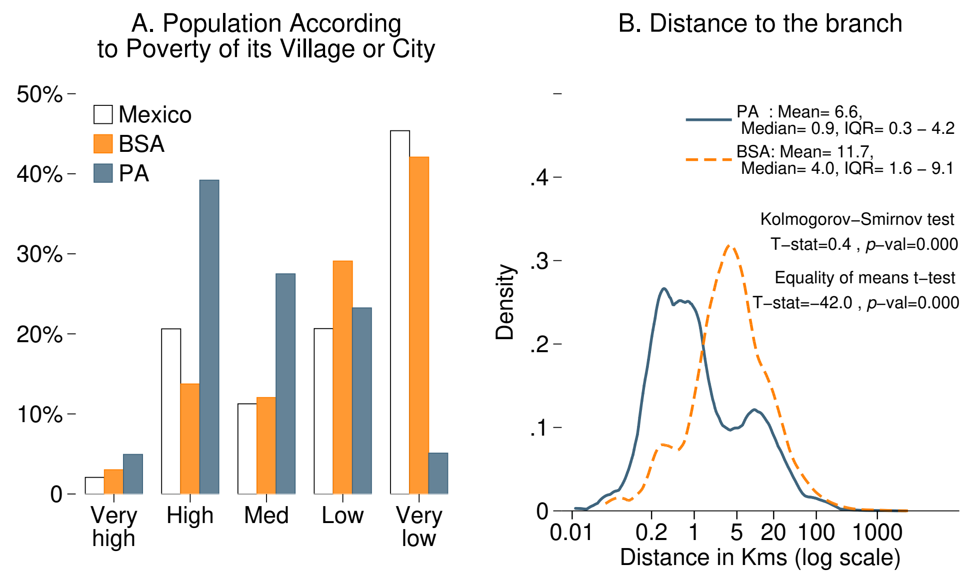
Self-selecting into Premiahorro, participants live in areas both poorer and closer to branches of the bank. Around 34 percent of the 111 million people in Mexico live in areas with a medium poverty level or higher. In areas with the same poverty level live more participants (72 percent) and fewer typical clients (29 percent). Participants live much closer to their branches. The median distance between a participant and her branch is 0.9 kilometers, a quarter of the distance between a typical client and her branch (4 kms).[27] Panel b provides test statistics and p-values for tests of equality of means and of distributions. Both tests reject the null of equality. Because participants and typical clients of the bank differ widely, a difference-in-differences that uses the whole sample can lead to biased results.[28] Before

[26] Around 184,000 people participated in Premiahorro. The number lowers to around 179,000 when the sample is the 395 branches I examine. Recall that I exclude branches that received or sent reallocated accounts.

[27] While 15 percent of typical clients live in the same postal code in which their branch is placed, far more participants do (74 percent).

[28] Consider distance to the branch. It is related to potential outcomes because the distance someone has to travel to access her account affects her savings behavior. Ashraf et al. (2006) find that a random offer of a deposit collector service, which reduces the distance to zero, increases by 40 percent the amount saved. Bachas et al. (2021) find that debit cards, which reduce the distance to the branch to the distance to the closest ATM, led to higher savings. In a companion paper (Bachas et al., 2018), they document strong correlations between savings behavior and travel distance: each 1 km reduction in the distance correlates with 0.06 more withdrawals and 28 more pesos in net savings.

Figure 7: Poverty and Distance to the Branch
Participants (PA) vs Typical Clients of Bansefi (BSA)



Sample of bank clients in 395 branches. BSA: Basic savings accounts holders in branches above the cutoff (n=428,000). PA: Premiahorro participants (n=179,000). IQR: Inter-quartile range. Poverty categories come from official estimations of the National Population Council (CONAPO). Poverty corresponds to 2010. Distance to the branch is the Euclidean distance, measured from the branch (GPS location) to the centroid of the postal code of the address of the client. For participants and for non-participants, estimation of the densities use the Epanechnikov kernel and a bandwidth of 0.30 log(kms).

participation, trends of outcomes might not be on parallel trends. Even if trends are parallel, after participation exposure to additional shocks might not be equal.

Only a third of participants had an account with Bansefi before or opened one after participating

The identification strategy only uses a sub-sample of all participants in Premiahorro. Around 65 percent of Premiahorro participants never had or never opened another account with Bansefi.[29] Table 3 details how many participants had an account before participating or opened one after it. The information only refers to the three main savings products. I lack, for example, information to know if participants had or received Oportunidades accounts. Only 35 percent or participants had an account before or opened one after participating. Of

[29] To open a Premiahorro account, a person needed not to be a current client. And because Premiahorro itself was a savings account, opening another account likely was unnecessary.

them, 63 percent had an account before, while 37 percent opened one when they opened the Premiahorro account or some time after. Only 31,000 of the 139,000 participants in the final but not in the preliminary design of the program had an account before participating (panel b, row 1, column 3). Of them, 28,000 had the basic savings account not linked to a credit card (panel b, row 2, column 3). The identification strategy uses a subset of these 28,000 participants.

Table 3: Premiahorro Participants
Other Accounts at the Bank

	Ever opened		Opening date relative to participation	
Type of account	No	Yes	Before	After
	A. 2009–2015 (n=183,972)			
Any of the three	119,229 (64.8%)	64,743 (35.2%)	40,940 (63.2%)	23,803 (36.8%)
- Basic Savings Account (BSA)	123,203 (67.0%)	60,769 (33.0%)	37,124 (61.1%)	23,645 (38.9%)
- BSA with Debit Card	178,649 (97.1%)	5,323 (2.9%)	3,106 (58.4%)	2,217 (41.6%)
- Commitment Savings	180,582 (98.2%)	3,390 (1.8%)	3,356 (99.0%)	34 (1.0%)
	B. 2010–2015 (n=138,962)			
Any of the three	90,031 (64.8%)	48,931 (35.2%)	30,974 (63.3%)	17,957 (36.7%)
- Basic Savings Account (BSA)	93,184 (67.1%)	45,778 (32.9%)	27,988 (61.1%)	17,790 (38.9%)
- BSA with Debit Card	134,943 (97.1%)	4,019 (2.9%)	2,393 (59.5%)	1,626 (40.5%)
- Commitment Savings	136,323 (98.1%)	2,639 (1.9%)	2,613 (99.0%)	26 (1.0%)

The table considers all who opened a Premiahorro account. It uses the year and month in which bank clients opened their accounts. Opening accounts after participation includes participants who opened the account the same month they opened the Premiahorro account.

Sample of individuals

The identification strategy uses a sample of 17,000 participants and an equal number of matched bank clients. Two sets of restrictions decrease the sample of participants. The first allows identification of the effect of Premiahorro on outcomes. Savings balances and transactions for Premiahorro and for the three main products of Bansefi are the only outcomes available. Lacking information on outcomes before participating, participants who never opened an account, or who opened one after participating, drop-out of the sample. Because I focus on the final design of the program (2010–15), I exclude anyone who opened an account in the preliminary design. Finally, I only keep participants who had the basic savings account not linked to a debit card, restricting the sample to 28,000 participants.

The first set of restrictions has two benefits. First, it excludes beneficiaries of Oportunidades. Some of its beneficiaries took part in Premiahorro. (Twenty-three percent of financial literacy trainees were beneficiaries of Oportunidades). But I lack information for Oportunidades accounts, which prevents analysis of the effect Premiahorro had on them. Restricting the sample to holders of basic savings accounts not linked to a debit card sorts beneficiaries of Oportunidades out. They already have the Oportunidades savings account. They have no need for another savings account (see Bachas et al. (2021) for descriptive statistics of use by beneficiaries of other accounts in any bank). Second, it removes the potentially confounding effect of having a debit card. Bachas et al. (2021) study the effect of providing debit cards to Oportunidades beneficiaries, finding that the cards allowed beneficiaries to save more.[30] Because the benefits of having a debit card can extend to anyone, I only keep in the sample holders of basic savings accounts not linked to a debit card.

The second set of restrictions decreases the sample size by 40 percent but aids identification. First, I only consider people with complete information on age, gender, and postal code. Second, I drop late entrants, participants who entered after the first trimester of each calenday-year cycle. Excluding late entrants eases matching participants with potential counterfactuals (see the identification strategy). And most participants entered early (see panel b of figure 3). The effect on sample size of the first two restrictions is low. Third, I only consider participants who opened the basic savings account at least 18 months before participating in Premiahorro. The final restriction causes most of the decrease in sample size (80 percent of the total decrease). But the restriction allows matching on a long trend of outcomes each participant with a counterfactual.

To select counterfactuals, I consider bank clients not from all branches but from branches above the cutoff. The pool of potential counterfactuals comprises 190,000 bank clients in branches above the cutoff. Publicity for Premiahorro was local; bank clients in branches above the cutoff were unaware of Premiahorro. Using as counterfactuals people unaware of Premiahorro addresses a common critique of matching: if matched counterfactuals are so alike to participants, why did they choose not participate? (see Lara Ibarra et al., 2017). Neither a decision-making progress nor unobservable characteristics mediated such

[30]The cards reduced transaction and monitoring costs. Beneficiaries no longer needed to travel to the branch to withdraw the conditional cash transfer. And by having an easy way to check their balance frequently, beneficiaries trusted the bank more. Increased trust encouraged them to leave in their accounts more money, to save more.

choice. Unmindful of Premiahorro, bank clients in branches above the cutoff made no choice. Further, branches above the cutoff are far away from branches that offered Premiahorro, limiting spillover effects and preventing people from knowing about the program. The median Euclidean distance between a bank client and her branch is 4 kilometers. Seven times higher is the distance from each branch that did not offer the program to the closest branch that did.[31]

The size of the pool decreases from all 428,000 basic savings account holders (not linked to a debit card) in branches above the cutoff to 190,000 owing to three restrictions. First, just as for participants, I only consider people with complete information on age, gender, and postal code. Second, explaining most of the decrease, I keep clients with at least 18 months with their basic savings accounts. Third, I discard anyone from branches in cities with more than one million people. People who live in large cities likely are exposed to different shocks. A gap in the distribution of branches suggests that branches in large cities are different from most branches of Bansefi (see figure 6).[32] Of 145 branches above the cutoff drop, 15 drop.[33]

3.3 Additional Indicators for Villages or Cities

I complement the financial records with indicators from three data sources at the village or city level. The first source is national population censuses: the population count for 2005 and the national census for 2010. I merge to these datasets an official poverty index calculated by the National Population Council (CONAPO). Using the first source, I create means of poverty, demographic, and education indicators. The second source is the registry of beneficiaries of Oportunidades. Each year from 1998 to 2014, the registry contains the number of families in the program. Using a simple interpolation, I complete for each village or city the series up to 2017.[34] Using the second source, I create the proportion of families in Oportunidades.[35]

The third source is the Municipalities Savings and Intermediation dataset (MSI). The MSI dataset contains indicators created from administrative records reported by financial institutions to the banking and securities regulator (CNBV). From 2000 to 2011, it provides

[31] Descriptive statistics of distance for the 145 branches above the cutoff: Median=30, Mean=98, SD=176, Min=2.5, Max=940, Interquartile range=18.5-77.3; Three of the 145 branches have a distance below 5kms; for twelve the distance is below 10kms.

[32] In large cities Bansefi caters to government employees.

[33] Four branches also drop from the sample because they offered no basic savings accounts, leaving a final sample of 376 branches ($= 395 - 15 - 4$).

[34] The program changed in 2015. After the change, administrative records are neither comparable nor available online.

[35] The denominator is the number of households in the 2010 population census.

detailed information on the presence of banks and on the number of savings accounts and of credit contracts. The information is available not at the village or city level but at the municipality level. Using the third source, I create indicators that approximate access to credit and savings in villages or cities within a municipality.

Merging village- or city-level indicators to branch-level records

Merging village- or city-level indicators to branch-level records is straightforward. The branch-level panel contains detailed information on the location of the branch.[36] Location of the branch corresponds to late 2009.

Merging village- or city-level indicators to individual-level records

Merging village- or city-level indicators to individual-level records requires matching. To match each person to a village or city, I use the postal code of her address. I also use the postal codes to calculate a crucial indicator, the distance between each person and her branch. The distance merely is the Euclidean distance from the centroid of the postal code to the GPS location of the branch. Boundaries of cities and postal codes lack a one-to-many correspondence because one postal code can cover two or more cities. The lack of a one-to-many correspondence forces a process of matching. Matching relies on geographic information system software and on layers for postal codes, cities, and villages.[37] For postal codes and for cities, layers are boundaries well-defined. For villages, layers are points defined by latitude and longitude.

Branch per branch the matching algorithm proceeds as follows. First, I list for each branch all postal codes from its clients. Matching starts with cities. Postal code by postal code, a postal code either matches no city or matches the city that contains the centroid of the postal code. After matching a city, the postal code drops from the list. Once the algorithm works through each postal code in the list, matching proceeds to villages. I calculate all pair-wise Euclidean distances between each village in Mexico and each centroid of the remaining postal codes. After matching its closest village, a postal code drops from the list. Once the list is empty, the algorithm for the branch ends and proceeds to the next branch. The algorithm entails matching error.[38] Even when matching error is present, a postal code matches a

[36] GPS coordinates, postal code, and a unique code used by Mexico's National Statistics Institute to identify each village or city.

[37] The three layers (shapefiles) correspond to 2010, the year in which the final design of Premiahorro started.

[38] For example, for cities a postal code that overlaps two or more cities is assigned only to one of them; for

village or city close to the true location of the client. Figure A.5 in the appendix takes a branch and illustrates the process.

4 Identification Strategy

A difference-in-differences identification strategy at the individual level provides the main results. Identification hinges on the choice of counterfactuals. For each of the 17,000 participants in the sample, I find within the pool of counterfactuals the closest match according to a propensity score. To estimate it, I use a post-double lasso strategy. To ensure parallel trends, I match on trends of outcomes before participation. The section first describes the estimating equation. Next it describes the outcomes as well as the variables the post-double lasso strategy uses. The section ends by detailing the selection of counterfactuals.

4.1 Estimating Equation

I estimate a standard difference-in-differences with two groups of individuals and multiple time periods:

$$y_{ibk} = \beta PA_{ibk} + \gamma OPORT_{bk} + \lambda_i + \lambda_k + \varepsilon_{ibk} \tag{1}$$

where y_{ibk} is the outcome for individual i in branch b at time relative to adoption k, measured in months; λ_i and λ_k are individual and time fixed effects. The time-varying control, $OPORT_{bk}$, is the proportion of Oportunidades of the families in the village or city in which the branch is placed.[39] The next section clarifies why the estimating equation includes this time-varying control. The parameter of interest is β. For participants, PA_{ibk} equals one from adoption to the end of the time-series; otherwise, it equals zero.[40] For matched counterfactuals, PA_{ibk} always equals zero. On inference, standard errors are clustered at the branch level b.[41]

The sample comprises 34,000 individuals (17,000 participants and 17,000 matched counterfactuals). The time-series covers 18 months before and up to 86 months since

villages a postal code might correspond not to the village with the closest latitude and longitude but to one close and extense.

[39] The information is at the year level. For a participant who entered 2014 and for her match, the corresponding value of $OPORT_{bk}$ equals the proportion in 2014 for periods $k \in [0, +11]$, the one in 2013 for periods $k \in [-12, -1]$, the one in 2012 for periods $k \in [-24, -13]$, and so on.

[40] A participant could have dropped after the first trimester of her first year, or finished the year and never participated again; or participated again a few more years; or participated in all years. I make no distinction. Treatment equals all months since adoption.

[41] Standard errors likely are biased downward because they lack adjustments to account for the uncertainty of estimating the propensity score, which allows to set the sample of counterfactuals. The towering size of the sample (NT=2.8 million) allays this concern. As the results show, standard errors would need to be five to seven times larger to render results non-statistically significant at conventional levels. (Refer to each first column of tables 5 and 6).

adoption ($k \in [-18, +86]$). For participants, I set as adoption time ($k = 0$) January of the year in which she entered the program for the first time.[42] For the matched counterfactuals (non-participants), adoption time is well-defined. It corresponds to January of the year in which the matched participant entered the program.[43] Individual time-series are as short as 45 months for those who entered the last year-cycle (2015) and as long as 105 for those who entered the first (2010).[44]

4.2 Outcomes; Variables the Post-double Lasso Strategy Uses

Outcomes

The two outcomes are the number of deposits in the month and the savings balance at the end of the month. For each outcome and person (17,000 participants and 190,000 potential counterfactuals), I create one variable. The variable sums each month the outcome from to the Premiahorro account (including the match) and from all basic savings accounts not linked to a debit card the person has. The sum of the number of deposits excludes the match and other payments made by the bank (e.g. interest payments).

Variables the post-double lasso strategy uses

To estimate the propensity score, regularized regressions in the post-double lasso strategy wade through a large list of controls, which ought to be selected using expertise and economic theory (Mullainathan and Spiess, 2017). Karlan et al. (2014) propose a taxonomy of savings constraints, grouping demand and supply constraints in five categories. Below I map to each category financial records and indicators for villages or cities.

Transaction costs compose the first category. They divide into pecuniary (e.g. fees) and non-pecuniary costs (e.g. time to travel). Clients of the bank face the same fees but differ on how far they live from the branch. I use two variables that measure distance. Regularized regressions pit them against each other. The first is the Euclidean distance from the branch

[42] Because I exclude from the sample late entrants, all participants adopted Premiahorro in the first trimester of a calendar-year cycle. Although some opened the account in February or in March, most open it in January. To simplify matching, I set January as the adoption month for everyone.

[43] For a participant who entered 2014 and for her match, $k = 0$ corresponds to January 2014, $k = -12$ to January 2013, $k = +6$ to July 2014, etc.

[44] For those who entered the first year-cycle (2010), and for their matched counterfactuals, $k = -18$ corresponds to July 2008 and $k = +86$ to the last month available, March 2017. For those who entered the last year-cycle (2015), $k = -18$ corresponds to July 2013 and $k = +26$ to the last month available. From $k = +27$ to $k = +86$, they drop from the sample.

to the centroid of the postal code of the client. The second is whether the postal code of the branch and the client are the same.

Lack of trust and regulatory barriers compose the second category. All clients overcame regulatory barriers, the ‘know your client’ requirements. But clients might differ on how much they trust the bank, which mediates how much they use the accounts. Time being a client serves as a proxy for trust. I assume that the longer a person has been a client, the more she trusts the bank. To measure time being a client, I use the number of months the person has had the basic savings account. For the few people with more than one account, I use the oldest account.

Information and knowledge gaps compose the third category. On the one hand, financial literacy training decreased the gaps for participants, plausibly leading to lower gaps relative to non-participants. On the other, participants come from less populous villages or cities in which these gaps can be high, preventing training from eliminating, relative to non-participants, the gaps. To account for differences in information and knowledge gaps, I match on past savings behavior. I also create a large array of poverty and educational indicators for the village or city in which the person lives.

Social constraints (intra- and inter-household bargaining) compose the fourth category. How much a person in a couple influences decision-making compared to her spouse defines intra-household bargaining. Females typically have a lower influence, a lower intra-household bargaining power. I account for differences in intra-household bargaining between participants and non-participants in two ways. First, I create variables that measure intra-household bargaining power at the village or city level. Second, I stratify the sample and match within each stratum (see the next sub-section). One stratum is gender, which allows to match the variable exactly.

Social and family networks define inter-household bargaining. Their relation with savings is ambiguous. They can foster savings when they help the least wealthy members of the network or can hinder savings when they tax the most wealthy (Karlan et al., 2014). I use information from Oportunidades to account for differences in inter-household bargaining. Beneficiaries of Oportunidades are not in the sample. But they received cash periodically, making them valuable members of social or family networks. They also interacted with Bansefi frequently, increasing the chance they belong to the social networks of other clients of Bansefi. Premiahorro participants live in less populous villages or cities, places in which

Oportunidades as proportion of families is larger. Relative to non-participants, participants then more likely can be family members, or be friends, or be spouses, of beneficiaries. In the estimating equation, I directly account for differences in inter-household bargaining. The time-varying control is the yearly series of share of families in Oportunidades for the village or city in which the branch is placed. For regularized regressions, I use the share of families in Oportunidades in 2010 in the village or city in which the person lives.

Biases in preferences, expectations, price perceptions, and problem solving—all are biases that compose the fifth category. Because I lack information about any of them, I account for them indirectly. They should influence patterns of savings. And the longer a time-series of savings is, the more likely it will reflect them, at least partially. To account for the biases, just as for the third category, the identification strategy uses a long time-series of savings before participation.

Once the information latches onto each category, the next step is to set the specific variables. The appendix details the dictionary of controls (figure A.2). All of them precede either participation or the year 2010. For individuals, variables correspond to January of the year the person entered the program. For the places where they live, they correspond to 2009 (or to 2010 when earlier information is not available.)[45]I feed 276 variables to regularized regressions: 17 at the individual level, 179 at the village or city level, and 80 at the municipality level. At the individual level, I use levels of four variables: age, distance to the branch, living in the same postal code of the branch, and months being a client of the bank. Squares and interactions complete the seventeen variables. At the village or city level, I use levels of twenty-seven variables. The two main variables are years of education and poverty level. The remaining are: whether the location is a village or a city; the proportion of families in Oportunidades; measures of intra-household bargaining power; and variables that capture education, migration, health, and dwelling characteristics.[46] Interactions complete the list of 179 variables.[47]

[45]Information from population census corresponds to 2010, the first year cycle of the program. Information from the MSI dataset (savings and credit information of banks in the municipality) corresponds to 2009. When information for 2009 is not available, it corresponds to the first trimester of 2010.

[46]The list contains six variables that measure intra-household bargaining power: (1) Sex ratio: female to male population; (2) male employment rate; (3) female employment rate; (4) live births per females 12+ (fertility rate); (5) share of population 12+ married or living in couple; (6) and share of population living in female headed households. While all village or cities are inhabited by males and females, not all have both male and females employed. For this reason, instead of a ratio, I use employment rates separately.

[47]I interact with each of the secondary twenty-five variables with six crucial indicators: poverty, years of education, and the four variables at the individual level.

At the municipality level, I use levels of sixteen variables. The list includes variables that proxy for access to credit, which eases savings constraints; and for savings in the population, which measures desire an ability to save. Ten variables capture access to credit and four variables capture savings in the population.[48] Two variables, government expenditure per person and length of roads per squared kilometer, capture general features of the municipality. Interactions complete the list of 80 variables.[49]

4.3 Selection of Counterfactuals

(1) Stratification of the sample

To select 17,000 persons among the 190,000 potential counterfactuals, I stratify the sample and match within each stratum. I create 60 strata by interacting three variables with no measurement error: gender (n=2), year of first entry to Premiahorro (n=6), and population in the village or city of the branch (n=5). By matching within stratum, I perform exact matching on the three variables. I stratify by gender because females face savings constraints that males do not. Chief among them is intra-household bargaining. I stratify by year of first entry because persons who entered the program as soon as it was available likely differ from those who entered later. I stratify by population in which the branch is placed; otherwise, matches might pile up in a few branches. Across both participants (n=17,000) and non-participants (n=190,000), I create quintiles of the population in which the branch is placed. The first quintile spans from 543 to 4,000 inhabitants for participants and from 50,000 to 94,209 inhabitants for non-participants (see figure A.6 in the appendix). Table 4 details the sample of 17,000 participants across the 60 strata.

(2) Estimation of the propensity score using post-double lasso

Within each stratum, I estimate the propensity score—the probability an individual i from village or city v would enter Premiahorro. I use the linear probability model:[50]

$$Premiahorro_{iv} = x'_{s\cup d\cup p,iv}\theta_{ps} + u_{iv}$$

[48] Savings per person and three variables of banking infrastructure measure savings in the population. Amount and share in default of credit card and credit contracts measure access to credit.

[49] Instead of interacting with six variables, I interact with four. I exclude distance from the branch and living at same postal code of the branch. The two variables correspond to the village or city in which the person lives. At the municipality level, the measures are uninformative.

[50] Using a linear probability model instead of a logit or a probit model leads to propensity scores outside the range [0-1]. In the identification strategy, negative propensity scores are no concern. The Euclidean distance between two scores, regardless of their sign, is well-defined.

Table 4: Sample of Premiahorro Participants
per stratum (60 Strata)

		Population Quintiles					
Year	Gender	I	II	III	IV	V	Total
2010	Males	73	79	44	109	56	361
	Females	1204	1016	915	994	1037	5166
2011	Males	86	103	71	107	75	442
	Females	530	550	454	502	571	2607
2012	Males	79	73	48	103	76	379
	Females	355	401	434	343	333	1866
2013	Males	126	81	98	95	103	503
	Females	558	417	730	388	447	2540
2014	Males	65	58	58	66	52	299
	Females	212	302	272	210	336	1332
2015	Males	67	52	20	55	43	237
	Females	262	226	191	206	296	1181
						Males	2221
						Females	14692

Strata: Six cycles of Premiahorro (2010-2015), two genders (females and males), and five population quintiles (for the quintiles, see figure A.6 in the appendix).

where $x_{s \cup d \cup p, iv}$ are regressors selected using post-double lasso.

Proposed by Belloni et al. (2014), post-double lasso selection is model selection using the least absolute shrinkage and selection operator (lasso) in at least two reduced forms: the first for the outcome and the second for the treatment indicator.[51] Because I analyze two outcomes, I use three reduced forms. With a notation similar to the one Belloni et al. (2014)

[51] Post-double lasso selection reduces the probability of excluding relevant variables by mistake. For example, had I used the third reduced form only, a model for participation in Premiahorro, I could have excluded variables with mid to low correlation with participation but with high correlation with outcomes, variables that the first two reduced forms would have included. Had I used the third reduced form only, I would have excluded relevant variables by mistake, inducing omitted variable bias. Besides being orthogonal to participation or its anticipation, controls selected using post-double lasso affect both the outcome and the decision to participate, a critical property that controls used for matching should satisfy (Caliendo and Kopeinig, 2008).

use, the three reduced forms are:

$$
\begin{aligned}
Savings_{iv} &= x'_{1,iv}\pi_{1,s} + x'_{2,iv}\pi_{2,s} + x'_{3,iv}\pi_{3,s} + r^s_{iv} + \varepsilon_{iv} \\
Deposits_{iv} &= x'_{1,iv}\pi_{1,d} + x'_{2,iv}\pi_{2,d} + x'_{3,iv}\pi_{3,d} + r^d_{iv} + \epsilon_{iv} \\
Premiahorro_{iv} &= x'_{1,iv}\theta_1 \quad + x'_{2,iv}\theta_2 \quad + x'_{3,iv}\theta_3 \quad + r^p_{iv} + v_{iv}
\end{aligned}
$$

where for individual i from village or city v, $x_{1,iv}$, $x_{2,iv}$, and $x_{3,iv}$ are the variables at the individual, village or city, and municipality level in the dictionary of controls. The equations include approximation errors (r_{iv}) that satisfy the approximate sparsity condition.[52] $Savings_{iv}$ and $Deposits_{iv}$ are means over fifteen months before participation, $k \in [-18, -4]$.[53] I exclude the trimester preceding participation to avoid Ashenfelter dips. Premiahorro started with a preliminary design in 2009. Participants in 2010, the first year I analyze, as well as participants in later cycles, could have known about the program. Knowing about the program and anticipating entering it, a person could have changed her normal savings behavior. But I assume her behavior mostly changes the trimester before entering the program.

Post-double lasso selection proceeds as follows. First, I use the full dictionary of controls to estimate a regularized regression (lasso) for each reduced form. To choose the tuning parameters, I use 'rigorous lasso' (Belloni et al., 2012). The method uses theory instead of cross-validation or information criteria (AIC, BIC, or their extended versions). Then I identify the variables rigorous lasso selects. The appendix (figure A.3) reports for each stratum the three lists, x^s, x^d, and x^p. Post-double lasso merely is post-lasso (Belloni and Chernozhukov, 2013), an OLS that has $Premiahorro_{iv}$ as the dependent variable and the union of variables from the three lists, $x_{s \cup d \cup p, iv}$, as regressors.

Matching on trends

Post-double lasso selects variables that explain variation of means of outcomes preceding participantion ($k \in [-18, -4]$). Matching using the propensity score reduces differences in means between participants and non-participants. However, similar means can mask different trends, which ought to be matched explicitly. To match the trend, I devise an algorithm that requires creating two samples, one per outcome. Consider the first outcome, savings. For each individual i (n=207,000), I estimate a simple linear regression: $y_i = \alpha + \mu t + \epsilon_i$,

[52] Approximate sparsity: within all plausible controls, which might exceed in number the sample size, just a few approximate well the true data generation process (see Belloni et al. (2012).

[53] For example, for participants who entered 2012, $Savings_{iv}$ and $Deposits_{iv}$ are means over the period July 2010 to September 2011 (15 months).

where y_i are savings balances per month and t indexes time. Each regression has fifteen observations, one per month preceding participation. Time t spans from 1 to 15, covering the period $k = -18$ to $k = -4$. The parameter of interest, $\hat{\mu}$, estimates the trend preceding participation.[54] Using the estimates, I split the sample of participants into deciles, identifying the range of $\hat{\mu}$ for each. For each participant, decile by decile, I seek her nearest neighbor within the same range of $\hat{\mu}$. Each participant matches her nearest neighbor from a pool of non-participants, both in the same stratum and in the same range of trends.

Piecing together the three elements of selection of counterfactuals

The algorithm to select counterfactuals starts with the first outcome, savings balances, and the first stratum. The first stratum contains 1,204 participants, females who entered the program in 2010 and whose branch is in the first population quintile below the threshold (see table 4). Of the 190,000 people in the pool of counterfactuals, the same stratum contains 16,855 non-participants: females who in January 2010 have had their basic savings accounts for at least 18 months and whose branch is in the first population quintile above the threshold. Using post-double lasso selection, I calculate the propensity score in the sample of 18,059 females ($= 1204 + 16855$). I then exploit estimates of trends of savings. The mean of $\hat{\mu}$ for the 1,204 participants is +80 MXN. Before participation, on average, their savings balances were increasing. In the same period, July 2008 to September 2009 ($k \in [-18, -4]$), savings balances of their potential counterfactuals were decreasing. The mean of $\hat{\mu}$ for the 16,855 non-participants is -130 MXN. Selection of counterfactuals proceeds decile by decile. Spanning from -3900 MXN to -230 MXN, the first decile contains 121 participants. Using the propensity score, I match each participant with the nearest neighbor among non-participants with trends within the same range (3,500 non-participants).[55] Spanning from +300 MXN to +7900 MXN, the tenth and last decile contains 120 participants for whom there are 1,000 non-participants within the same range. After matching the 1,204 participants, the algorithm for the first stratum ends.

Once the selection of counterfactuals for the first stratum ends, matched non-participants drop from the pool of potential counterfactuals. Selection then proceeds stratum by stratum.

[54] Note that the regression uses error-free information from banking records. The estimated parameter is not an approximation but the actual trend of the outcome preceding participation.

[55] Participants can drop from the matched sample when the decile for participants lacks enough potential matches. For example, the first decile has 121 participants, but there could be no counterfactuals in the same range. If this happens, the 121 participants drop from the sample. Or the decile could have few potential matches, for example 10. If this happens, ten participants find a match and 110 drop-out. Having no or fewer potential matches seldom happened.

With the last stratum, matching for the first outcome ends. For the second outcome, the number of deposits, selection starts anew, creating a new sample of matched non-participants. Difference-in-differences regressions use two different samples, one per outcome. Both samples contain the same participants. The group of non-participants, however, differs because the trends of savings balances and of the number of deposits differ.

How the process to select counterfactuals performs

Matching substantially decreases differences in potential confounders between participants and non-participants. The decrease is higher for females, who make up 87 percent of the sample. Figures A.7 and A.8 in the appendix present standardized differences before and after matching. Before matching (figure A.7), and relative to non-participants, participants have been with the bank longer. They also live in areas with fewer banks and with more people that save and use credit less and that have completed fewer years of education. After matching (figure A.8), differences decrease substantially. For females, and for individual and village or city characteristics, all differences decrease to around 0.20 standardized differences or below. The largest decreases are for months with the account, poverty score, and years of education. For males, differences decrease too, but they remain above 0.40 standardized differences for the poverty score and years of education.[56] At the municipality level, differences for males and females decrease but hover around 0.40 standardized differences. Municipality-level characteristics proxy for village- or city-level characteristics. Because matching decreases differences in individual and in village- or city-level characteristics successfully, failure to decrease differences in municipality-level characteristics is less of a concern.

Models selected by post-double lasso outperform an ad-hoc model. I pit predictive performance of the models selected by post-double lasso against an ad-hoc model. The adjusted R^2 of the linear probability model that estimates the propensity score measures performance. Informed by the same taxonomy of savings constraints, the ad-hoc model uses the four variables at the individual level and their squares. It adds the levels and squares of years of education and poverty, two critical variables at the village or city level. There are no guidelines on how to choose among all other variables; so I select the six I deem most relevant.[57] Nor there are guidelines on how to adapt the model for each of the 60 strata; so I

[56] Post-double selection finds fewer controls for models for males, plausibly explaining the lower performance of the matching process for them (see figure A.2 in the appendix).

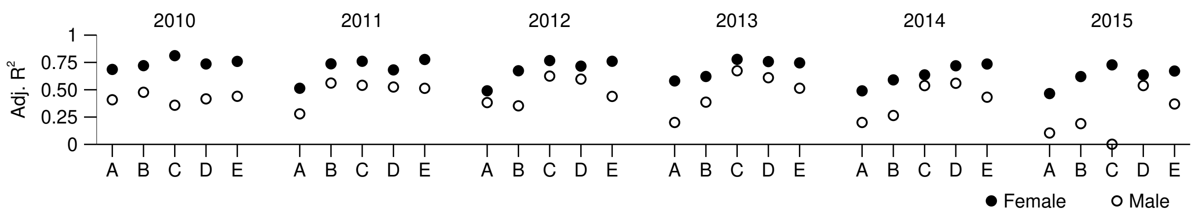
[57] (1) Sex ratio (proxy for bargaining power); (2) whether the location is a village or a city; (3) expenditures per person of the government of the municipality; (4) length of roads per km^2 of surface area; (5) savings per person, and (6) credit per person.

use the same model.

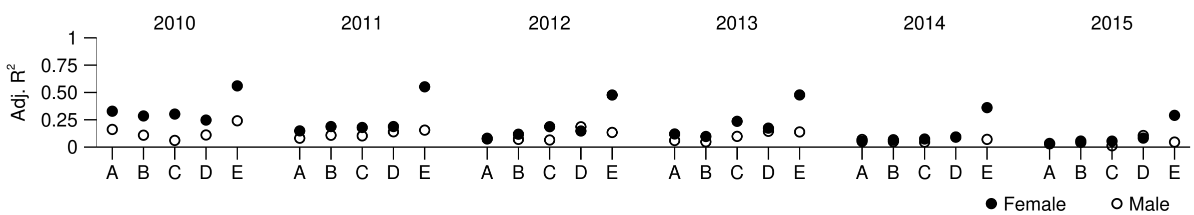
Figure 8 presents the comparison. Each dot denotes the adjusted R^2 for each of the 60 strata. Panel a presents results for post-double lasso selection; panel b for the ad-hoc model. While promising a high out-of-sample prediction, post-double lasso selection delivers a high within-sample prediction, particularly for females. For males, the adjusted R^2 hovers around 0.40-0.60; for females, around 0.70-0.90. The ad-hoc model performs poorly. Except for 2010, when it delivers adjusted R^2 hovering around 0.30 for females and 0.14 for males, most of the adjusted R^2 are below 0.16 for females and 0.11 for males. Guided by the same taxonomy of savings constraints, and in a data-driven way, post-double lasso selection delivers models that vastly outperform it.

Figure 8: Fit of Linear Probability Models that Estimate the Propensity Score (60 Strata)

A. Post-double lasso selection



B. Ad-hoc model



5 Main Results: Difference-in-Differences

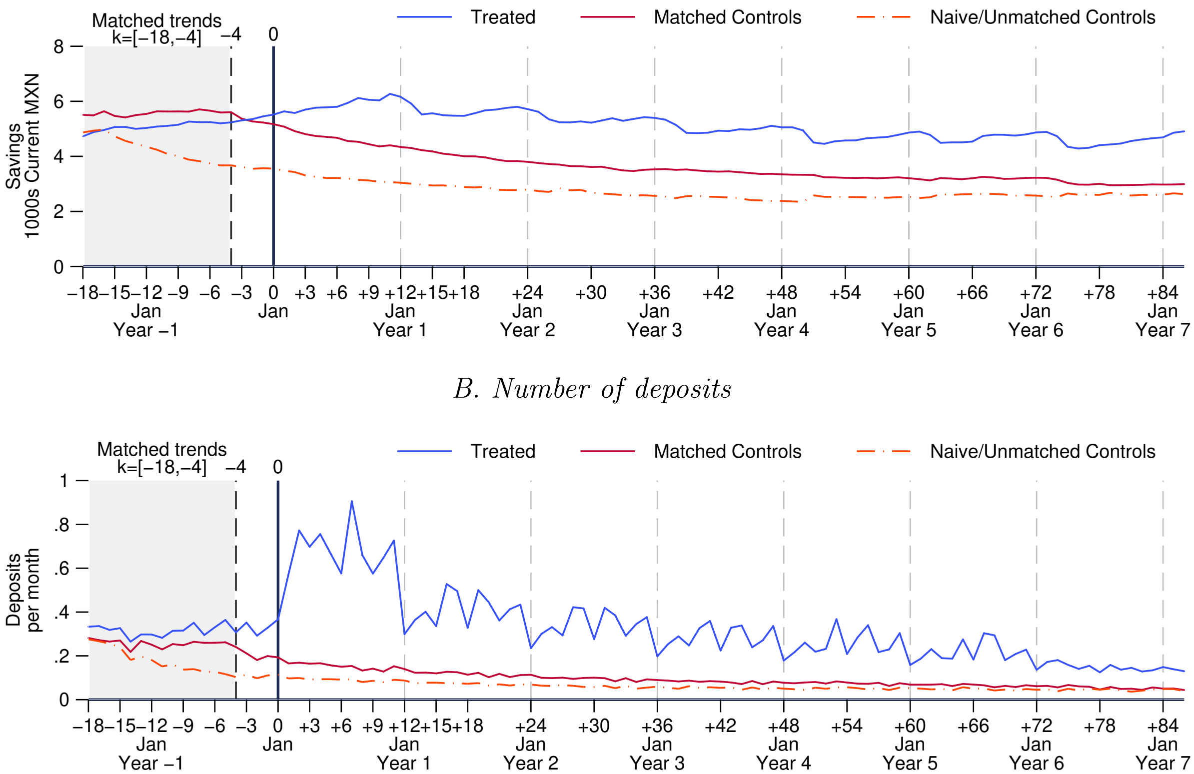
5.1 Trends of Outcomes in the Matched Sample

Figure 9 presents trends of outcomes in the matched sample. Panel a refers to savings balances and panel b to the number of deposits. On each panel, a highlighted region encompasses the fifteen months matched, (k=-18 to k=-4). The figure adds in dashed lines trends for a naïve group of counterfactuals. Instead of being carefully selected, the group consists of 17,000 people selected randomly from the 190,000 pool of counterfactuals.[58] Not only trends for the naïve group of counterfactuals are not parallel before participation, they also suggest a higher effect. The estimating equation using the naïve group would have retrieved estimates biased upwards.

Figure 9: Difference-in-Differences: Trends

A. Savings balances

B. Number of deposits



Sum of end of month balances and number of deposits of the basic savings account (not linked to a debit card) and the Premiahorro account. Savings balances but not the number of deposits include the match.

Sample for savings balances: Treated=16,854 Controls=16,854.

Sample for the number of deposits: Treated=16,427, Controls=16,427.

The trends suggest that difference-in-differences will retrieve large effects. In a long period before participation, trends of outcomes are on parallel paths. In the fifteen months matched, levels are similar and paths are parallel, making identification possible. By assuming equal exposure to additional shocks—or by controlling for them—the identification strategy can retrieve unbiased estimates of the effect of Premiahorro. In the trimester preceding

[58] Random selection takes place within strata, performing exact matching on the variables used for stratification. But neither trends of outcomes nor other potential confounder are matched.

participation, trends start to differ. Participants saved and made deposits at a similar rate as before. Matched non-participants, however, began to make fewer deposits and to decrease savings balances.[59] Because participants likely were aware of the program, I ascribe the difference in behavior to anticipation. Once participation starts, trends suggest that the effect of Premiahorro is large. For both outcomes, a large gap between participants and matched counterfactuals emerges. After seven years since adoption, the large gap in savings balances persists. For the number of deposits, a flurry of activity emerges for participants in the first year since adoption. After the first year, the gap sharply narrows. By the seventh year, the gap is small.

5.2 Effects on Savings Balances and on the Number of Deposits

Table 5 presents the main results for savings balances. Column (1) excludes from the equation the proportion of families in Oportunidades. Column (2) adds the proportion back. Column (3) uses the sample of females and column (4) the sample of males. The bottom rows present summary statistics and confidence intervals for the parameter of interest. Summary statistics (mean and standard deviation over all time periods) are for the matched counterfactuals; a row presents the effect size.[60]

Premiahorro increases savings balances by 2120 MXN (column 2), an increase equal to 52 percent of the mean of matched counterfactuals. The effect size (0.16) suggests a relatively small effect. Excluding the variable for Oportunidades has, reassuringly, a negligible effect on the point estimate (column 1). Both males and females benefit from Premiahorro, but females benefit more. Although confidence intervals of the effects largely overlap, the mean of the outcome for female counterfactuals is lower. For males, the effect equals 35 percent of the mean of matched counterfactuals (effect size=0.13); for females, it equals 56 percent, a higher effect (effect size=0.17).

Premiahorro more than doubles the number of deposits. Following the same structure, table 6 presents the results. Premiahorro increases the number of deposits by 0.21 deposits per month (column 2), an increase of 162 percent over the mean of matched counterfactuals. The effect size (0.44) suggests a large effect. Excluding the variable for Oportunidades has, once more, a negligible effect. Females again benefit more. For males, the effect equals 95

[59] The trimester before the participation is from October to December, a period of high spending owing to holidays.

[60] The effect size equals the estimated parameter of interest, β, as proportion of the standard deviation of the outcome for counterfactuals, $\hat{\sigma}^C$.

Table 5: Effect of Premiahorro on Savings Balances

	All		Females	Males
	(1)	(2)	(3)	(4)
$PA_{ibk} = 1$	2155.99***	2119.29***	2082.91***	2369.40***
	(151.09)	(150.91)	(147.25)	(426.27)
$OPORT_{bk} = 1$		1480.16*	1195.98	3770.82*
		(758.21)	(733.27)	(1935.09)
Observations	2,817,492	2,817,492	2,476,836	340,656
Branches (Clusters)	376	376	376	361
Individual (i) FE	Yes	Yes	Yes	Yes
Time (k) FE	Yes	Yes	Yes	Yes
Couterfactuals				
Mean ($\hat{\mu}^C$)	4125.18	4125.18	3751.6	6841.45
Std. Dev. ($\hat{\sigma}^C$)	12855.56	12855.56	12011.35	17592.9
Effect Size ($\hat{\beta}/\hat{\sigma}^C$)	.17	.16	.17	.13
95% Confidence interval				
Lower	1858.89	1822.55	1793.37	1531.1
Upper	2453.08	2416.03	2372.45	3207.69

Robust standard errors cluster at the branch level in brackets.
**p*-value<0.10, ***p*-value<0.05, ****p*-value<0.01.
Sum of end of month balances of the Premiahorro account and of all basic savings account not linked to a debit card. Balances include Premiahorro's match.
Sample: 33,708 individuals, 14,658 female and 2,196 male participants matched to an equal number of counterfactuals.

percent of the mean of matched counterfactuals (effect size=0.32); for females, it equals 183 percent, a much higher effect (effect size=0.46).

A higher proportion of families in Oportunidades helps people in the sample to reach their savings goals faster. It increases savings balances (table 5, columns 2 to 4) while reducing the number of deposits (table 6, columns 2 to 4). The higher the proportion is where the branch is placed, the more likely a person in the sample is a friend, or a relative, or a spouse of a beneficiary of Oportunidades. A higher proportion then expands social and family networks, the expansion having a positive effect. Males benefit more than females. Females in the sample can only be friends or be relatives of beneficiaries. But males can also be spouses, expanding for males the probability of having beneficiaries in their networks.

Table 6: Effect of Premiahorro on The Number of Deposits

	All		Females	Males
	(1)	(2)	(3)	(4)
$PA_{ibk} = 1$	0.20***	0.21***	0.22***	0.18***
	(0.01)	(0.01)	(0.01)	(0.02)
$OPORT_{bk} = 1$		-0.34***	-0.34***	-0.26**
		(0.09)	(0.09)	(0.12)
Observations	2,748,846	2,748,846	2,440,344	308,502
Branches (Clusters)	376	376	376	359
Individual (i) FE	Yes	Yes	Yes	Yes
Time (k) FE	Yes	Yes	Yes	Yes
Couterfactuals				
Mean ($\hat{\mu}^C$)	.13	.13	.12	.19
Std. Dev. ($\hat{\sigma}^C$)	.48	.48	.47	.57
Effect Size ($\hat{\beta}/\hat{\sigma}^C$)	.42	.44	.46	.32
95% Confidence interval				
Lower	.18	.19	.19	.14
Upper	.22	.23	.24	.22

Robust standard errors cluster at the branch level in brackets.
*p-value<0.10, **p-value<0.05, ***p-value<0.01.
Sum of the number of deposits to the Premiahorro account and to all basic savings account not linked to a debit card. The sum excludes the match from Premiahorro and any other payment from Bansefi.
Sample: 32,854 individuals, 32,854 individuals, 14,444 female and 1,983 male participants matched to an equal number of counterfactuals.

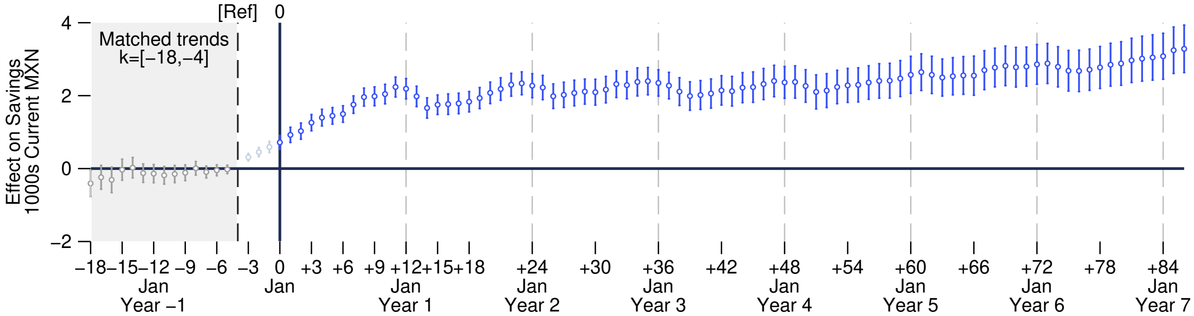
5.3 Effects During and After Premiahorro Ended

As a first approach, figure 10 depicts results from a leads-and-lags regression. For each month relative to adoption, the figure depicts point estimates and confidence intervals. Because I allow for anticipation effects, I set $k = -4$ as the reference category. Unlike trends in figure 9, these results partial-out time and individual fixed effects and the effect of Oportunidades. Results formally show parallel trends. In the period $k \in [-18, -4]$, estimates and intervals hover around zero. Results show a large effect, in particular for savings balances. By the seventh year since adoption, effects of Premiahorro on both outcomes persist.

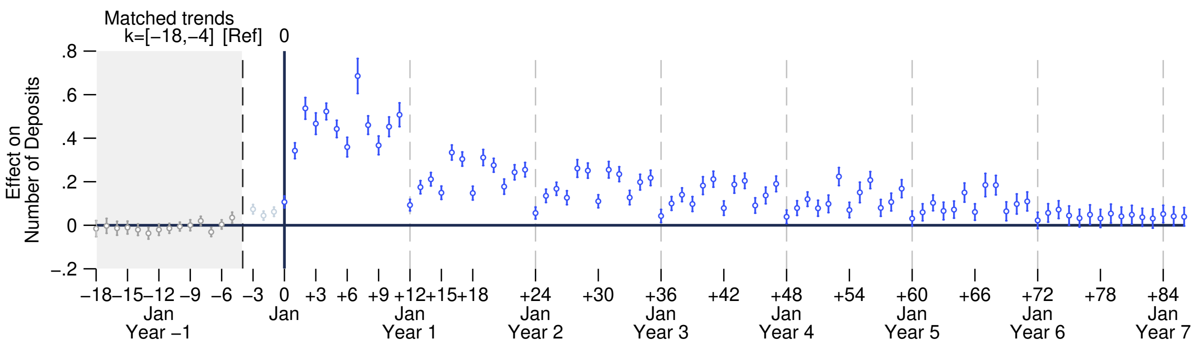
As a second approach, I formally test whether after Premiahorro ends the effect persists. Because Premiahorro ended in 2015, I create a dummy variable equal to one for all periods

Figure 10: Effect By Time Since Adoption
Leads-and-Lags Regression

A. Savings balances



B. Number of deposits



Sum of savings balances or of the number of deposits to the Premiahorro account and to all basic savings account not linked to a debit card. For the number of deposits, the sum excludes the match from Premiahorro and any other payment from Bansefi. For savings balances, it includes the match.
Sample for savings balances: Treated=16,812, Controls=16,812.
Sample for the number of deposits: Treated=16,791, Controls=16,791.

k corresponding to January 2016 onwards; otherwise it equal zero.[61] I interact the dummy variable with the variable PA_{ibk}, creating a new one, PA_{ibk}^{After}. The estimated parameter for the new variable measures the effect of Premiahorro after it ended. I add the two new variables to the estimating equation. Adding them transforms the original variable, PA_{ibk}, to PA_{ibk}^{During}. Its estimated parameter now measures the effect Premiahorro had during its run.

Table 7 presents the results for savings balances. Serving as reference, column (1) corresponds to the original estimating equation. Column (2) corresponds to the new equation,

[61] The months the dummy variable uses vary by year of entry. For example, for participants who entered 2010 (and for their matches), it equals one when $k \in [+72, +86]$. For those who entered the last year, 2015, it equals one when $k \in [+12, +26]$. The upper limits, 86 and 26, correspond to March 2017, the last available information.

which divides the effect Premiahorro had during and after its run. Columns (3) and (4) repeat the pattern for the sample of females; columns (5) and (6), for the sample of males. The table adds a row that presents the p-value from a test of equality of coefficients for PA_{ibk}^{During} and PA_{ibk}^{After}.

After Premiahorro ends, its effect on savings balances persists. During Premiahorro, the effect is an increase of 48 percent over counterfactuals (effect size=0.15). After Premiahorro, it is 66 percent (effect size=0.21). For females, during Premiahorro the effect is an increase of 51 percent over counterfactuals (effect size=0.16); after Premiahorro, it is 71 percent (effect size=0.22). For males the corresponding increases are 31 percent (effect size=0.12) and 47 percent (effect size=0.18).

After Premiahorro ends, the effect on number of deposits substantially wanes but persists. Following the same structure, table 8 presents results for the number of deposits. The effect Premiahorro had on the number of deposits during its run was 0.26 deposits per month. The effect equals an staggering increase of 200 percent over counterfactuals (effect size=0.54). Once the program ended, the effect substantially wanes to 0.05 but remains statistically different from zero. The effect equals a much lower increase of 38 percent (effect size=0.11). For females, during Premiahorro the effect is an increase of 217 percent over counterfactuals (effect size=0.55); after Premiahorro, it is 50 percent (effect size=0.13). For males the corresponding increases are 121 percent (effect size=0.40) and a neither statistically nor substantially significant 11 percent (effect size=0.04).

5.4 Effects on Active Use of the Basic Savings Accounts

Premiahorro increased how often people deposited savings in their accounts. The effect, however, could only be on the Premiahorro account. For this reason, I test whether active account use of the basic savings accounts increased. I use two measures of active account use. Gertler et al. (2021) review the literature and highlight three measures: making at least one deposit in the last six months, at least two, and at least five in the last two years. I settle for the first two measures.[62] Figure 11 presents the results (panel a for making at least one deposit in the last six months and panel b for making at least two). Each panel presents estimates and confidence intervals for the effect of Premiahorro (PA_{ibk}, solid dot), and for its effect during (PA_{ibk}^{During}, hollow diamond) and after it (PA_{ibk}^{During}, hollow square).

[62] For each month, I sum the number of deposits in the month and in the preceding five. Active account use equals one when the individual made the required number of deposits; otherwise, it equals zero. Because five lags are not complete when $k \in [-18, -14]$, for active account use the time-series restricts to $k \in [-13, 86]$.

Table 7: Effects During Premiahorro and After it Ended
Outcome: Savings Balances

	All		Females		Males	
	(1)	(2)	(3)	(4)	(5)	(6)
$PA_{ibk} = 1$	2119.29***		2082.91***		2369.40***	
	(150.91)		(147.25)		(426.27)	
$PA_{ibk}^{During} = 1$		1944.82***		1922.04***		2099.45***
		(136.96)		(134.68)		(409.37)
$PA_{ibk}^{After} = 1$		2738.04***		2667.96***		3211.31***
		(220.62)		(212.15)		(560.63)
$1(k \geq$ Jan-2016)		-168.07**		-144.85*		67.90
		(83.63)		(83.76)		(249.55)
$OPORT_{bk} = 1$	1480.16*	502.76	1195.98	277.26	3770.82*	2049.70
	(758.21)	(737.39)	(733.27)	(723.62)	(1935.09)	(1795.64)
Observations	2,817,492	2,817,492	2,476,836	2,476,836	340,656	340,656
Branches (Clusters)	376	376	376	376	361	361
Individual (i) FE	Yes	Yes	Yes	Yes	Yes	Yes
Time (k) FE	Yes	Yes	Yes	Yes	Yes	Yes
Couterfactuals						
Mean (μ^C)	4125.18	4125.18	3751.6	3751.6	6841.45	6841.45
Std. Dev. (σ^C)	12855.56	12855.56	12011.35	12011.35	17592.9	17592.9
Effect Size ($\hat{\beta}/\sigma^C$)†	.16	.21	.17	.22	.13	.18
95% Confidence interval						
Lower	1822.55	2304.23	1793.37	2250.8	1531.1	2108.78
Upper	2416.03	3171.84	2372.45	3085.11	3207.69	4313.84
Equality Test						
$H_o : \beta^{During} = \beta^{After}$ $H_a \neq$		0.00		0.00		0.00

Robust standard errors cluster at the branch level in brackets.
*p-value<0.10, **p-value<0.05, ***p-value<0.01.
Sum of end of month balances of the Premiahorro account and of all basic savings account not linked to a debit card. Balances include Premiahorro's match.
† In columns (2), (4), and (6), effect size and confidence intervals refer to the effect of Premiahorro after it ended (PA_{ibk}^{After}).
Sample: 33,708 individuals, 14,658 female and 2,196 male participants matched to an equal number of counterfactuals.

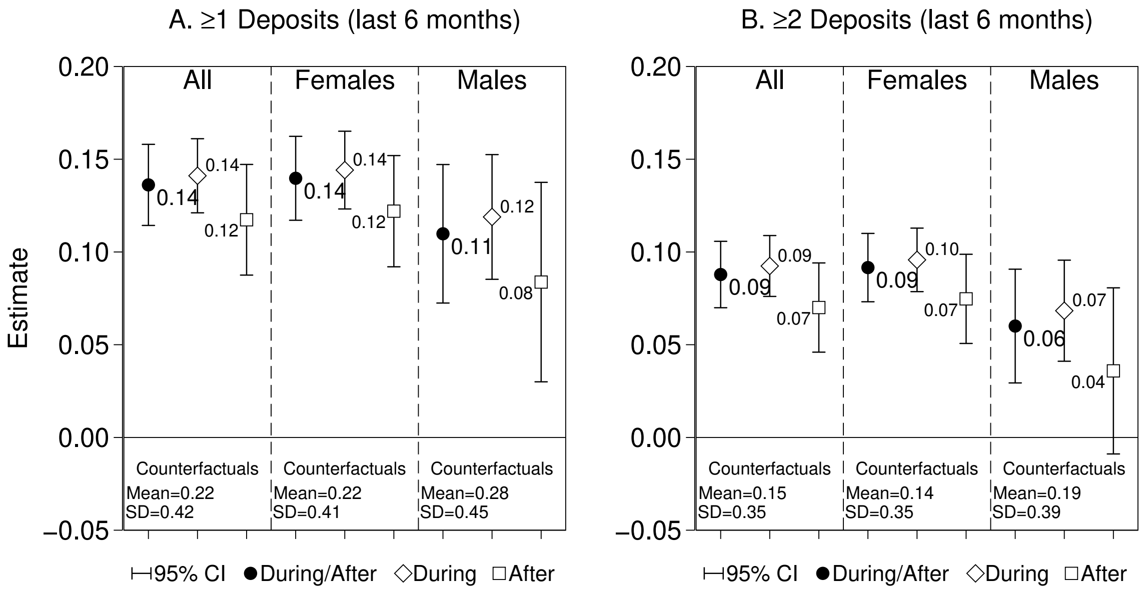
Premiahorro increased active use of the basic savings account, both during and after its

Table 8: Effects During Premiahorro and After it Ended
Outcome: Number of Deposits in the Month

	All		Females		Males	
	(1)	(2)	(3)	(4)	(5)	(6)
$PA_{ibk} = 1$	0.21***		0.22***		0.18***	
	(0.01)		(0.01)		(0.02)	
$PA_{ibk}^{During} = 1$		0.26***		0.26***		0.23***
		(0.01)		(0.01)		(0.02)
$PA_{ibk}^{After} = 1$		0.05***		0.06***		0.02
		(0.01)		(0.01)		(0.03)
$1(k \geq$ Jan-2016)		0.03***		0.03***		0.05***
		(0.01)		(0.01)		(0.01)
$OPORT_{bk} = 1$	-0.34***	-0.08	-0.34***	-0.09	-0.26**	0.00
	(0.09)	(0.07)	(0.09)	(0.07)	(0.12)	(0.10)
Observations	2,748,846	2,748,846	2,440,344	2,440,344	308,502	308,502
Branches (Clusters)	376	376	376	376	359	359
Individual (i) FE	Yes	Yes	Yes	Yes	Yes	Yes
Time (k) FE	Yes	Yes	Yes	Yes	Yes	Yes
Couterfactuals						
Mean (μ^C)	.13	.13	.12	.12	.19	.19
Std. Dev. (σ^C)	.48	.48	.47	.47	.57	.57
Effect Size ($\hat{\beta}/\sigma^C$)†	.44	.11	.46	.13	.32	.04
95% Confidence interval						
Lower	.19	.03	.19	.03	.14	-.03
Upper	.23	.08	.24	.08	.22	.08
Equality Test						
$H_o: \beta^{During} = \beta^{After}$ $H_a \neq$		0.00		0.00		0.00

Robust standard errors cluster at the branch level in brackets.
*p-value<0.10, **p-value<0.05, ***p-value<0.01.
Sum of the number of deposits to the Premiahorro account and to all basic savings account not linked to a debit card. The sum excludes the match from Premiahorro and any other payment from Bansefi.
† In columns (2), (4), and (6), effect size and confidence intervals refer to the effect of Premiahorro after it ended (PA_{ibk}^{After}).
Sample: 332,854 individuals, 14,444 female and 1,983 male participants matched to an equal number of counterfactuals.

Figure 11: Outcome: Active Use of the Account
Basic Savings Accounts Only



Numbers of deposits in basic savings account only.
Sample: 32,854 individuals, 14,444 female and 1,983 male participants matched to an equal number of counterfactuals.

run. In the first definition (≥ 1 deposits in the last six months), Premiahorro increases active use by 14 percentage points. The increase equals 64 percent over counterfactuals (effect size=0.33). During the program, the point estimate for the effect is the same at 14 percentage points; after it, the point estimate is lower at 12 percentage points. But confidence intervals for the effect during and after Premiahorro largely overlap. In the second definition (≥ 2 deposits in the last six months), usage increases by 9 percentage points. The increase equals 60 percent over counterfactuals (effect size=0.26). But confidence intervals for the effect during (9 percentage points) and after (7 percentage points) Premiahorro also overlap. Point estimates are higher for females, but the proportional increase over counterfactuals between males and females is the same. Premiahorro increased how often people deposited savings not only in the Premiahorro account but also in the basic savings account. Premiahorro developed in participants the habit of saving, a habit that persisted after Premiahorro ended.

6 Supporting Results: Regression Discontinuity Design at the Branch Level

The regression discontinuity design at hand lacks power, but its results are consistent with the main results. Section A.2 in the appendix describes outcomes and controls and discusses the internal validity of the design. This section describes the estimating equation, provides the results, and discusses the power calculations for the design.

6.1 Estimating Equation

Estimation relies on the continuity approach proposed by Hahn et al. (2001). They formalized conditions for identification and proposed local linear regressions for estimation.[63] Estimation of the effect of Premiahorro at the cutoff uses two local linear regressions, each weighted by a triangular kernel, and both defined by:

$$y_b = \tau_{SRD} PA_b + \beta_1 Forcing_b + \beta_2 (PA_b \times Forcing_b) + \beta_3 y_b^{Bef} + \beta_4 X_b^{Pred} + \varepsilon_b \qquad (2)$$
$$\forall b: \ Forcing_b \in (h_{MSE}^{Below}, h_{MSE}^{Above})$$
$$Kernel = Triangular$$

where y_b is the outcome for branch b and $Forcing_b$ is the population where the branch is placed. The two outcomes are number of accounts and of transactions. The forcing variable is centered at the cutoff (50,000) and expressed in 10,000 people. Premiahorro, PA_b, equals 1 for branches that offered it since 2010; otherwise, it equals 0. Because the design is sharp, PA_b also equals 1 when the population is below the cutoff and 0 when it is above. I use the MSE-optimal bandwidth proposed by Calonico et al. (2014).[64] I allow the bandwidth to differ below and above the cutoff (h_{MSE}^{Below} and h_{MSE}^{Above}) because the branches that offered Premiahorro are squeezed in a small range (0–50,000) while those that did not spread in much larger range (50,000–1.8 million). The parameter of interest, τ_{SRD}, is the difference

[63] They show that the average causal effect of treatment, τ_{SRD}, in this case of a Premiahorro, equals:

$$\begin{aligned}\tau_{SRD} =& \mathbb{E}[y_b(1) - y_b(0) | Forcing_m = 0] \\ =& \lim_{x \downarrow 0} \mathbb{E}[y_m | Forcing_m = x] - \lim_{x \uparrow 0} \mathbb{E}[y_m | Forcing_m = x] \\ =& \mu_b(x) - \mu_a(x)\end{aligned}$$

where $\mu_b(x)$ and $\mu_a(x)$ are interpolations of the outcome at the cutoff from below and from above.

[64] Their method to estimate MSE-optimal bandwiths improves the seminal method by Imbens and Kalyanaraman (2012) in two ways. First, their bandwidth avoids estimating some parameters required for estimation of the variance. Second, their pilot bandwidths are mean-squared error optimal. To implement their method, I use the Stata commands `rdbwselect` and `rdrobust` by Calonico et al. (2017).

between two non-parametric estimations of the outcome at the same point, at the cutoff. The estimating equation includes the outcome before Premiahorro, y_b^{Bef}, as well as other predetermined variables, X_b^{Pred} (see table A.1 in the appendix). Predetermined variables enter the estimation equation additively, as Calonico et al. (2019) suggest.[65]

For inference, I use the robust bias-corrected confidence intervals by Calonico et al. (2014). Their method is robust because it accounts for the estimation of the bias when estimating the variance that in turn is used to calculate the confidence intervals. I also calculate and report the minimal coverage error (CE) confidence intervals by Calonico et al. (2020). While keeping MSE-Optimal bandwidths for identification, they propose different, narrower bandwidths for inference. Rather than balancing a bias-variance trade-off, these bandwidths minimize coverage error.[66]

6.2 Results

Table 9 presents the regression discontinuity results at the branch level. Owing to lack of power, results largely are uninformative. The table suggests large increases in the number of accounts and of transactions, but no estimate is statistically different from zero. Table A.2 in the appendix presents results for predetermined variables. A valid design must show for them continuity at the threshold. No estimate is statistically different from zero, suggesting continuity.

[65] Calonico et al. (2019) formally define covariate adjustment in regression discontinuity designs. Their covariate-adjusted estimator requires separability between a covariate and the forcing variable (e.g. the variables cannot be interacted). It also requires that treatment has no effect on covariates (e.g. the variables must be continuous at the cutoff). When covariates truly are predetermined, they argue, including covariates could increase precision.

[66] For example, if researchers intend a nominal target of 95 percent, they might unwittingly report 90 percent confidence intervals if the coverage error is large (see Monte Carlo simulations in Calonico et al. (2020).)

Table 9: Results: Regression Discontinuity Design
Triangular Kernel, Polynomial of Degree 1

	# of Accounts				# of Transactions			
	During and after		Before		During and after		Before	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Estimate ($\hat{\tau}_{SRD}$)	982.4	889.4	-314.3	-377.2	266.2	163.6	154.8	-2.1
Control mean	3841.0	3696.5	5709.6	5812.6	879.1	829.1	928.5	971.1
Effect size	0.42	0.41	-0.06	-0.07	0.52	0.33	0.23	-0.00
h_{MSE}^{Below}	-3.005	-2.706	-2.602	-2.423	-2.981	-1.954	-2.240	-2.172
h_{MSE}^{Above}	25.133	10.813	16.332	12.764	25.363	15.047	19.637	22.631
Observations	132	93	101	90	131	89	103	108
<50,000	51	37	35	32	50	24	32	31
>50,000	81	56	66	58	81	65	71	77
95% CI (MSE-Opt)								
Lower	-1171.8	-350.3	-3325.3	-3356.1	-272.1	-388.8	-472.5	-541.2
Upper	3517.1	2228.5	3016.9	3131.7	877.9	697.9	870.7	670.4
p-value	[0.327]	[0.153]	[0.924]	[0.946]	[0.302]	[0.577]	[0.561]	[0.834]
95% CI (CE)								
Lower	-1378.6	-295.7	-3562.8	-4366.5	-319.9	-577.8	-502.8	-533.8
Upper	3620.7	2216.9	3350.1	2610.8	905.3	642.9	1026.3	823.7
p-value	[0.379]	[0.134]	[0.952]	[0.622]	[0.349]	[0.917]	[0.502]	[0.676]
Covariates	No	Yes	No	Yes	No	Yes	No	Yes

p-value in brackets, *p<0.10, **p<0.05, ***p<0.01.
Mean of controls equals the mean of the outcome over the right-side of the cutoff. Effect size is the point estimate divided by the standard deviation of the outcome over the right-side of the cutoff.
Covariates: (a) Outcome before Premiahorro (except for columns 4 and 8). (b) Village or city (2010): years of education and % of families in Oportunidades. (c) Municipality (2009): Savings in all banks, number of branches per km^2 per 100,000 people, and length of roads per km^2. No municipality has branches both offering and not offering Premiahorro. The 250 villages or cities that offered Premiahorro are within 249 municipalities; the 145 not offering are within 121.

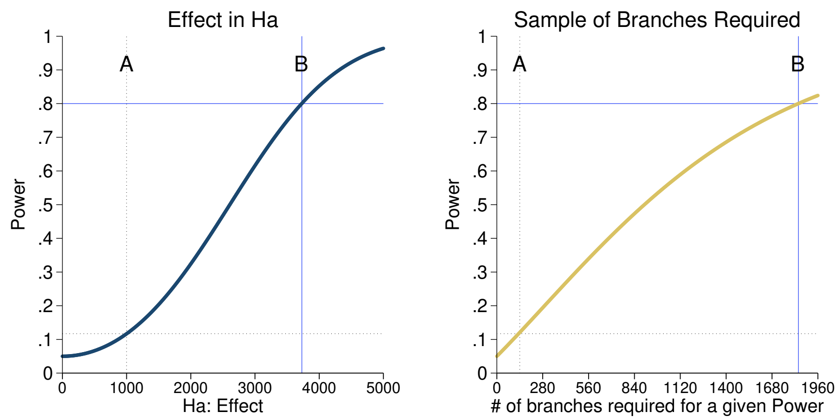
6.3 The Discontinuity Design Lacks Power

Explaining why results largely are uninformative, the regression discontinuity design at hand lacks power. I present power calculations, which need an alternative hypothesis and a bandwidth. Point estimates hint that Premiahorro each month increased the number of accounts by 1,000 and of transactions by 300. Both numbers serve as the alternative hypotheses, the effect size. I set an ad-hoc bandwidth of 30,000 below and 300,000 above the cutoff. Albeit inadequate for point estimation and for inference, the ad-hoc bandwidth helps to approximate the effective sample size. Within the bandwidth, the sample of branches decreases from 395 to 140, to 51 below and 89 above the cutoff. With alternative hypotheses and bandwidth defined, I estimate two functions. The first, given sample size, is power as a function of effect size. The second, given effect size, is power as a function of sample size.

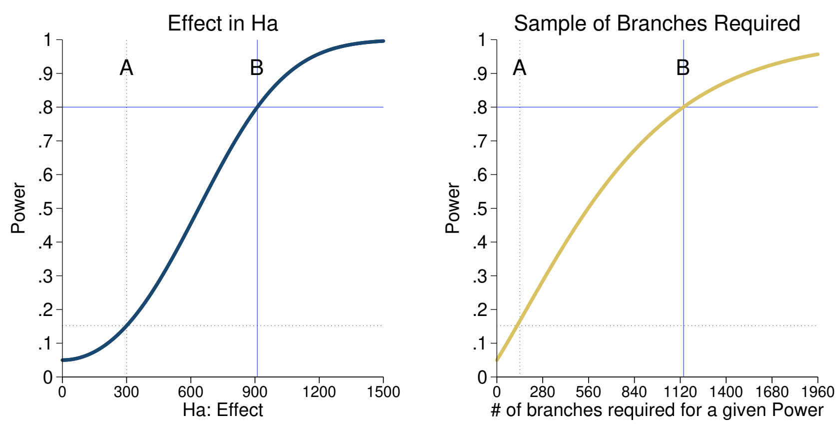
Figure 12 depicts the functions. For the number of accounts, panel a shows that the power to detect an increase of 1000 accounts by the current sample is 0.12. The sample required to identify an increase of 1000 with 0.80 power is, at 1,841 branches, far greater than the total sample of 395 branches. For the number of transactions, panel b shows similar results: with the current sample, low power (0.15); with the plausible effect, high sample (1,140 branches). The design is under-powered.

Figure 12: Power Calculations for the Regression Discontinuity Design
A: Actual Sample of Branches. B: Sample Required

A. Number of Accounts



B. Number of Transactions



Of the 395 branches, power calculations consider 140 branches only. Below the cutoff the 51 within the range 20,000–50,000 (bandwidth=30,000). Above the cutoff the 89 within a range 50,000–350,000 (bandwidth=300,000). Calculations use the Stata packages `rdpow` and `rdsampsi` by Cattaneo et al. (2019)

7 Concluding Remarks

I provide empirical evidence on the positive effects of a matched savings program that provided financial literacy training. Besides being a matched savings program, Premiahorro was a flexible commitment savings strategy. But its prominent feature was the match, one equal to a 62 percent annual interest rate. By combining the match, the training, and the flexible commitment strategy, Premiahorro helped people to save. Identification relies on high-frequency and high-volume financial records. For identification, I use a differences-in-differences on a sample of participants matched with an equal number of counterfactuals. Counterfactuals come from a large pool of non-participants from branches that did not offer the program. To select each counterfactual, I use a propensity score estimated using post-double lasso selection and use trends of outcomes over the 18 months preceding participation. Among these bank clients, I find that Premiahorro increased savings balances by 48 percent (effect size=0.15) and number of deposits made by 162 percent (effect size=0.44). Premiahorro ran from 2010 to 2015, year when it was phased out. After it was phased out, its effects persist. Up to seven years since participants adopted the program and from the moment it was phased out, Premiahorro increased savings balances by 66 percent (effect size=0.21). The effect on number of deposits made waned but persisted (38 percent; effect size=0.11).

I contribute to the literature by providing evidence that combining matched subsidies, financial training, and commitment savings strategies can be effective, particularly to overcome low take-up rates when each is offered on its own. Another contribution is evidence that effect of providing a match can persist long after the match is phased-out. However, an open question is which combination of the three features is more cost-effective while still helping people to save. And the results are limited to a specific sub-sample of bank clients and to formal savings. Premiahorro likely improved a downstream outcome. After the program was phased-out, participants kept at the bank two-thirds of what they saved and received in matches. As they were taught, they saved most of the money, increasing their resilience against shocks.

References

Ashraf, N., Karlan, D., and Yin, W. (2006). Deposit Collectors. *The B.E. Journal of Economic Analysis & Policy*, 6(2):1–24.

Bachas, P., Gertler, P., Higgins, S., and Seira, E. (2018). Digital Financial Services Go a Long Way: Transaction Costs and Financial Inclusion. *AEA Papers and Proceedings*, 108:444–448.

Bachas, P., Gertler, P., Higgins, S., and Seira, E. (2021). How Debit Cards Enable the Poor to Save More. *The Journal of Finance*, 76(4):1913–1957.

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse Models and Methods for Optimal Instruments With an Application to Eminent Domain. *Econometrica*, 80(6):2369–2429.

Belloni, A. and Chernozhukov, V. (2013). Least Squares After Model Selection in High-dimensional Sparse Models. *Bernoulli*, 19(2):521–47.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, 81(2):608–650.

Bryan, G., Karlan, D., and Nelson, S. (2010). Commitment devices. *Annual Review of Economics*, 2(1):671–698.

Caliendo, M. and Kopeinig, S. (2008). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys*, 22(1):31–72.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal Bandwidth Choice for Robust Bias-corrected Inference in Regression Discontinuity Designs. *The Econometrics Journal*, 23:192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2017). Rdrobust: Software for regression-discontinuity designs. *Stata Journal*, 17(2):372–404.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019). Regression Discontinuity Designs Using Covariates. *The Review of Economics and Statistics*, 101(3):442–451.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs: Robust Nonparametric Confidence Intervals. *Econometrica*, 82(6):2295–2326.

Cattaneo, M. D., Jansson, M., and Ma, X. (2020). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 115(531):1449–1455.

Cattaneo, M. D., Titiunik, R., and Vazquez-Bare, G. (2019). Power Calculations for Regression-Discontinuity Designs. *The Stata Journal*, 19(1):210–245.

Duflo, E., Gale, W., Liebman, J., Orszag, P., and Saez, E. (2006). Saving Incentives for Low- and Middle-Income Families: Evidence from a Field Experiment with H&R Block. *The Quarterly Journal of Economics*, 121(4):1311–1346.

Gertler, P., Higgins, S., Scott, A., and Seira, E. (2021). Increasing Financial Inclusion and Attracting Deposits through Prize-Linked Savings.

Hahn, J., Todd, P., and Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica*, 69(1):201–209.

Imbens, G. and Kalyanaraman, K. (2012). Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies*, 79(3):933–959.

Imbens, G. W. and Lemieux, T. (2008). Regression Discontinuity Designs: A Guide to

Practice. *Journal of Econometrics*, 142(2):615–635.

John, A. (2020). When Commitment Fails: Evidence from a Field Experiment. *Management Science*, 66(2):503–529.

Kaiser, T. and Menkhoff, L. (2017). Does Financial Education Impact Financial Literacy and Financial Behavior, and If So, When? *The World Bank Economic Review*, 31(3):611–630.

Karlan, D., Ratan, A. L., and Zinman, J. (2014). Savings by and for the Poor: A Research Review and Agenda. *Review of Income and Wealth*, 60(1):36–78.

Lara Ibarra, G., McKenzie, D., and Ruiz Ortega, C. (2017). Learning the Impact of Financial Education When Take-Up is Low. Technical Report 8238, World Bank.

Lee, D. S. and Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2):281–355.

McCrary, J. (2008). Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test. *Journal of Econometrics*, 142(2):698–714.

Mullainathan, S. and Spiess, J. (2017). Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives*, 31(2):87–106.

Schaner, S. (2018). The Persistent Power of Behavioral Change: Long-Run Impacts of Temporary Savings Subsidies for the Poor. *American Economic Journal: Applied Economics*, 10(3):67–100.

A.1 Appendix

A.1.1 Additional Information: Main Savings Products

The three main savings products of Bansefi

(1) Basic Savings Account (BSA,'Cuentahorro'): To open an account, clients submit a proof of identity (usually the national voting identification) and any proof of residence. Deposits and withdrawals are done at the branch. The deposit to open the account should be 50 MXN (around 4USD). The minimum deposit amount is 30 MXN (around 2USD) and the minimum withdraw is 50 MXN. To receive interest payments, clients must keep a minimum monthly balance of 50 MXN. The interest rate paid varies per balance; it amounts to between 0.12 and 1.15 percent. Clients receive no penalty when they fail to keep the minimum balance. The bank charges an administration fee of 20 MXN (1.5 USD) per month but only after a year of inactivity. There are no other fees.

(2) BSA-DC ('Debicuenta'): Similar to the Basic Savings Account but it offers a debit card that clients can use at any ATM. Using the card entails paying commissions. Whereas Bansefi has been offering the BSA savings account since its inception, it has offered, at selected branches, the debit card since 2007.

(3) Maturity-based commitment savings product ('Tandahorro'): The client selects a maturity term from 1 to 36 months. She can only withdraw savings until her selected term lapses. The minimum deposit is 50 MXN. The client is expected to deposit monthly but failing to deposit entails no penalty. Clients need to open or to have a savings account at the branch (either BSA or BSA-DC). Once the selected term lapses, the bank transfer the money to the savings account. The interest rate paid varies per balance and amounts to between 0.39 and 1.16 percent. The product guarantees to cover inflation. If the interest paid is below inflation, Bansefi covers the difference. To protect against inflation, the selected term must be at least 6 months and the ending balance must be at least 1,800 MXN.

A.1.2 Additional Information: Financial Literacy Workshop

Figure A.1: Financial Literacy Workshop
General Characteristics and Protocol

GENERAL CHARACTERISTICS

Duration: 90 minutes
Spatial coverage: 227 of the 258 that offered the program between 2010-2015
Temporal coverage: Around 1800 per year workshops between 2009 and 2013
Participants: Groups between 20-50 persons. Anyone could attend but preference was given to Premiahorro account holders

PROTOCOL

Greetings and introduction — Time: 5 min

Introduce yourself, then ask who already has the Premiahorro account. Pick a person at random. Ask the person to explain how the program works. Briefly explain. Tell the audience that the workshop will be about the importance of savings

Identify the expenses people face — Time: 10 min

Objective: Teach the four type of expenses
- Daily expenses: Regular and required for subsistence. They are predictable.
- Future expenses: Large expenses, like weddings, that require savings. One can plan for them.
- Unexpected expenses: They are unpredictable. It is not possible to know when they will happen or how much will they cost.
- Investment opportunities: Savings will lead to income in the future.

How to reduce expenses — Time: 10 min

Objective: Emphasize the importance of reducing expenses
Teach: Resources come from two sources, separating some of the income and reducing expenses.
Activity: Ask who buys sodas, then ask how much each costs. Calculate with participants the expense per month in sodas. Ask them about other expenses they could avoid. Use the didactic material and teach to: spend less on non-essentials, parties, festivals; buy items in bulk and not piece by piece; and use credit less, specially for superfluous expenses.
Activity: Ask people who are willing to reduce expenses to stand up. Invite those sitting to do small efforts to reduce expenses.

Explain Preamiahorro — Time: 10 min

Objective: Explain in detail the program
Use the didactic material to explain the program. Clarify any doubts.

Savings Goals — Time: 10 min

Objective: Explain that they should be realistic, precise, measurable, and within a time limit.
Activity: Use the didactic material of the ladder towards a savings goal. Explain that each of the four rungs corresponds to each of the four characteristics of savings goals.

Emergency Funds — Time: 10 min

Objective: Emergency funds should be three times the monthly expenses.
Activity: Pick two or three persons at random. Ask each not to tell you how much do they spend per month but to tell you whether they know how much they spend. Ask in general what would be the monthly expenditure of a family in this area. Use the amount and explain that three times the amount should be an emergency fund for a typical family.
Activity: Ask to raise the hand to anyone who knows someone who had a serious disease or accident during the last year. Ask them to think about how much peace of mind having the amount calculated in the previous activity will give.

Financial needs and products — Time: 10 min

Objective: Associate savings, credits, and insurance to the four types of expenses.
Activity: Show the image of the four types of expesnes. Ask them how would they get the money to cover each. Make them aware that monthly income might not be enough.
Teach: Explain each of three products offered by financial institutions—savings, credit, and insurance.
Activity: Ask participants which of the products would they use for each of the four expenses.
Teach: You must cover daily expenses with your income and not with credit. You should save for future events. You should also save for unexpected expenses but insurance will help. Investment opportunities can be fulfilled with a combination of savings and credit.

Life Insurance — Time: 25min

Objective: Explain how useful insurance is.
Get two volunteers, one will be the insurer and the other will be the funeral house. The remaining participants will form five groups. Each group receives 10 stones, equivalent each to 100 pesos (a total of 1000 pesos). Four families will buy insurance by giving one stone to the insurer. Each of the four families now has 9 stones while the fifth keeps the 10 stones. The insurer has 4 stones from families in the area but it also has contributions from many other families—give the insurer many stones.
Ask for a volunteer from each of the five families. Use the five envelopes prepared for the activity and ensure that you know which of the five contains the only red card. Give each volunteer an envelope and ensure that the family without insurance receives the envelope with the red card. Tell the families that if they receive a red card they must pay 10,000 to the funeral house. Ask the volunteers to open the
Ask the participants. Who won in this exercise? Answer: Those who bought insurance.

A.1.3 Additional Information: Post-double Lasso Selection

Figure A.2: Dictionary of Controls

Individual	
A	Age in January 1, 2010. Decimals are days remaining expressed as fractions of 365
D	Euclidean distance: From centroid of the postal code of the person to the branch
T	Months with the oldest basic savings account at January of each year cycle
SP	The person and the branch share postal code
Total: 17	Each (4), each squared (3, SP is a dummy variable), bi-variate interactions (6), tri-variate interactions (3: AxDxT AxTxSP DxTxSP), and AxDxTxSP

Village or city in which the person lives (2010)	
Y	Years of education
P	Poverty score
L1	Sex ratio: Female / Male population
L2	Fertility: Number of live births per female age 12+
L3	% of pop. who was borned in the state
L4	% of pop. who speaks an indigenous language (age 3+)
L5	% of pop. 6-11 not attending school
L6	% of pop. 12-14 not attending school
L7	% of pop. 12+ Economically Active
L8	Male employment rate
L9	Female employment rate
L10	% of pop. without access to health services
L11	% of pop. with formal employment health services
L12	% of pop. 12+ married or living in couple
L13	% of pop. in female headed households
L14	% of dwellings unhabitated or temporally habitated
L15	Persons per room
L16	% of dwellings with: Electricity
L17	% of dwellings with: Dirt floor
L18	% of dwellings with: Only one room
L19	% of dwellings with: Piped water
L20	% of dwellings with: Sewage
L21	% of dwellings with: Car
L22	% of dwellings with: Computer
L23	% of dwellings with: Internet
L24	% of families in Oportunidades
L25	Urban area (population in 2010>=2,500)
Total: 179	Levels (29: 25 + years of education, poverty score, and each squared). Interactions with: years of education (25), poverty score (25); and age (25), months (25), distance (25), same postal code (25)

Municipality in which the person lives (2009 and 2010)	
M1	Municipality expenditure per capita (2010)
M2	Length of Roads per km2 of surface area (2010)
	Savings and banking infrastructure (2009)
M3	Savings per person (all deposits, all banks)
M4	# ATMs per km2 per 100,000 person
M5	# Bank branches per km2 per 100,000 person
M6	# of Checking and savings Accounts per 100,000 people
	Consumer credit outstanding (June-Dec 2009) per person
M7	Credit, leasing of vehicles, equipment, of software
M8	Credit cards
M9	Vehicles
M10	Loans for durable goods
M11	Personal loans
M12	Payroll loans
	Credit contracts (June-Dec 2009) per person
M13	% of credit card contracts at default
M14	% of credit contracts at default
M15	% of credit card contracts of total loans
M16	% of loans for durable goods of total loans
Total: 80	Levels (16). Interactions with: years of education (16), poverty score (16); age (16), months (16). Not interacted with distance or same postal code.

Figure A.3: Models Selected by LASSO

Year	S	M	Female	N	Male	N
2010	A	*y1*	A2 L22xD L18xY M10	4	A AxT M10	3
		y2	T L21 L6xD L8xT L11xT L12xT L18xSP L2xY M2xD M2xT M10xT M15xY	12	L6xD M16xT	2
		T	DxSP L11 L13 L20 L3xA L16xA L18xA L4xD L5xD L24xD L13xT L20xT L3xSP L4xSP L6xSP L18xSP L23xSP L24xSP L25xSP L5xY L17xY L18xY L4xP M2 M1xA M10xA M13xA M15xA M9xD M3xT M11xT M1xSP M2xSP M3xSP M9xSP M10xSP M11xSP M16xSP M1xY M13xY M4xP M7xP M14xP	43	L3 L3xA L18xSP L23xSP L6xY L15xY L18xY L6xP M1xA M1xSP M10xSP M1xP M11xP	13
	B	*y1*	A2 L6 L4xA L7xT L4xY M7xD M6xT	7	A AxT L4 L4xY M2xSP	5
		y2	T L11xA L1xT L7xT L9xT L12xT L20xT L18xY M11xA M12xA M13xT M7xY M11xY	13	A L4xA L3xY L4xP M6xA	5
		T	L8 L23 L2xA L24xD L7xT L3xSP L24xSP L4xY L13xY L18xY L14xP L18xP M3 M6 M10 M11 M1xA M14xT M2xSP M3xSP M5xSP M6xSP M7xSP M9xSP M10xSP M11xSP M15xSP M9xY M2xP	29	L8 L13 L3xA L12xA L4xD L3xSP L4xY L13xY L17xY M3 M11 M1xA M10xA M2xD M2xSP M3xSP M7xSP M11xSP M2xP	19
	C	*y1*	A2 T2 L12 L20xT L25xT L15xY L25xY L25xP M1xT M2xT M8xT M12xT M15xT M3xP	14	A AxT L16xT	3
		y2	T L9xA L2xT L3xT L7xT L8xT L12xT L15xT L21xT L18xY L25xY L13xP M1xA M13xA M10xSP M4xY M9xY M10xY M9xP	19		0
		T	DxSP P2 L7 L1xA L3xA L7xA L4xD L17xD L7xT L10xT L2xSP L3xSP L10xSP L17xSP L24xSP L2xY L4xY L5xY L10xY L15xY L17xY L18xY L21xP L22xP L23xP M1xA M10xA M11xA M12xA M5xT M1xSP M5xSP M8xSP M10xSP M11xSP M12xSP M16xSP M13xY M14xY	39	DxTxSP P2 L2xSP L24xSP L5xY M3 M10 M11 M3xSP M11xSP M13xY	11
	D	*y1*	A2 T2 P2 L11xT L13xT L23xT L24xT L17xY M15xT M7xSP M1xY	11	A AxT L9xT L20xT	4
		y2	T L13xA L3xT L16xT L19xT L21xSP L2xY L18xY M7 M12 M6xA M15xA M2xSP M7xSP	14	M14xA	1
		T	Y2 L7 L12xA L5xD L6xD L2xSP L3xSP L17xSP L24xSP L2xY L7xY L10xY L15xY L18xY L23xY L2xP L10xP L14xP M3 M6 M7 M11 M13 M6xA M13xA M16xA M13xT M14xT M3xSP M6xSP M7xSP M9xSP M10xSP M16xSP M1xY M2xY M9xY M10xY M15xY	39	L23 L3xA L10xY L15xY L18xY L5xP M3 M13 M2xSP M3xSP M16xSP M9xY M15xY	13
	E	*y1*	A2 T2 L1xT L16xT L4xSP L18xSP M16xSP	7	A AxT L11xT L4xSP	4
		y2	L9xA L12xT L4xSP L2xY L3xY M13xT M14xT	7	T2 L20xT L4xSP	3
		T	P2 L7 L8 L3xA L8xA L21xA L3xT L3xSP L23xSP L24xSP L6xY L14xY L24xY M8 M14xA M15xA M3xT M10xT M4xSP M6xSP M7xSP M8xSP M9xSP M10xSP M5xP	25	L3 L8 L3xA L5xSP L17xSP L7xY L16xP M7 M12 M7xSP M3xY M13xP M15xP	13
2011	A	*y1*	A2 L21xA L12xT L18xY M10xT M16xT M15xP	7	A T2 L1xT	3
		y2	T L3xA L2xD L4xD L6xD L11xT L12xT L13xT M1xT M10xT	10	L6xD L7xY M16xT M15xP	4
		T	L18 L19 L20 L3xA L16xA L20xA L5xD L2xT L5xT L13xT L20xT L3xSP L18xSP L23xSP L24xSP L25xSP L5xY L17xY L18xY M2 M1xA M15xA M2xT M3xT M11xT M13xT M1xSP M2xSP M3xSP M5xSP M7xSP M8xSP M12xSP M16xSP M7xY M4xP M14xP	37	L3 L3xA L4xD L5xY L13xY L18xY L23xY M1 M3 M11 M10xA M11xT M1xSP M3xSP M8xSP M11xSP	16
	B	*y1*	A2 L4xT L14xY M6xA M12xA M16xD M6xT M6xY	8	A L9xA	2
		y2	T L1xT L11xT L12xT L18xY M6xA M12xA M16xD M7xSP	9	A2 L4xD L10xY M6xA M2xP	5
		T	L7 L8 L12xA L14xA L4xD L3xSP L6xSP L17xSP L24xSP L4xY L10xY L13xY L18xY L2xP L14xP L18xP L23xP M3 M6 M10 M14 M1xA M1xT M2xT M5xT M9xT M11xT M1xSP M2xSP M3xSP M4xSP M5xSP M6xSP M7xSP M9xSP M10xSP M11xSP M2xY M10xY M15xY M2xP	41	L19 L4xD L6xD L17xD L24xD L3xSP L4xSP L6xSP L24xSP L10xY L13xY L14xP L17xP M5 M6 M11 M1xSP M2xSP M3xSP M4xSP M7xSP M10xSP M11xSP M10xY M13xY M2xP	26
	C	*y1*	A A2 L12 L22xT L23xT L25xT L13xP L25xP M1xT M3xT M8xT M12xT	12	A L16xA	2
		y2	A L21 L2xT L3xT L7xT L12xT L2xSP L13xY L18xY L25xY M4xA M5xA M1xSP M10xY M9xP	15	L8xY L4xP M14xSP	3
		T	P2 L7 L12 L1xA L3xA L4xD L17xD L1xT L2xSP L3xSP L14xSP L24xSP L2xY L4xY L5xY L11xY L14xY L17xY L10xP L21xP L25xP M11 M1xA M10xA M12xA M1xSP M2xSP M5xSP M8xSP M10xSP M11xSP M16xSP M2xY M13xY	34	P2 L7 L15xSP L24xSP L10xY M11 M10xA M1xSP M3xSP M11xSP M13xY	11
	D	*y1*	A2 P2 L3 L23xT L17xY L22xP M13 M2xSP M7xSP M15xY	10	A T2 L13xT L19xT L22xT	5
		y2	T L8xA L3xT L11xT L12xT L14xT L16xT L19xT L18xY M7 M6xA M1xT M2xT M13xT M7xSP	15	L9xA	1
		T	L7 L2xA L23xA L5xD L17xD L5xT L2xSP L15xSP L17xSP L24xSP L18xY M3 M7 M14 M6xA M13xA M14xA M13xT M16xT M3xSP M5xSP M7xSP M9xSP M10xSP M16xSP M1xY M13xY M15xY M6xP M9xP M16xP	31	P2 L22 L3xA L2xT L17xSP L24xSP L12xY L15xY L18xY M3 M13 M2xA M2xSP M3xSP M5xSP M9xSP M16xSP M1xY M9xY M15xY M1xP	21
	E	*y1*	A2 T2 L13xT L4xSP M7xD M15xT	6	A T2 L15 L18 L6xD L11xT	6
		y2	L12 L2xY L3xY M14 M4xA M15xA M14xT	7	DxSP L3 L8xA M1xP	4
		T	P2 L7 L8 L21 L3xA L8xA L3xT L3xSP L5xSP L23xSP L24xSP L14xY M14 M15 M2xT M3xT M8xT M10xT M5xSP M7xSP M9xSP M15xP	22	P2 L3xA L8xA L3xT L3xSP L24xSP L18xY L21xP M3 M7 M9 M14 M4xSP M7xSP	14
			A L21xA L3xT L12xT L15xT L17xT M10xT M16xT	8	A M16xT	2

2012	A	*y1*	T L3xD L4xD L5xD L24xD L2xT L11xT L12xT M6xA M2xT M10xT M13xT M10xY	13	L7xA L17xD L2xT L12xT M6xA M16xT	6
		y2	DxSP L3 L18 L20 L3xA L12xA L4xD L5xD L2xT L5xT L20xT L25xT L3xSP L18xSP L23xSP L24xSP L25xSP L5xY L17xY L18xY L5xP L6xP L25xP M1 M2 M3xT M10xT M11xT M13xT M15xT M1xSP M2xSP M10xSP M11xSP M15xSP M4xP M13xP	37	DxSP DxTxSP L3 L9 L4xD L23xSP L24xSP L25xSP L15xY L18xY L25xY M1 M3 M10xA M3xT M1xSP M10xSP M11xP	18
		T	A2 DxT L4xD L4xT L12xT L15xT M6xA M9xA M7xD M13xT M14xT	11	A	1
	B	*y1*	T L4xD L3xT L9xT L11xT L12xT L15xT L21xT L18xY L18xP M6xA M12xA M1xT	13	L4xD L15xT M1 M6xA M16xD	5
		y2	L4 L12xA L19xA L4xD L5xD L6xD L4xT L3xSP L15xSP L17xSP L24xSP L2xY L10xY L13xY L14xY L2xP L5xP L14xP L18xP M6 M10 M13 M15 M9xA M1xT M5xT M1xSP M2xSP M3xSP M4xSP M5xSP M6xSP M7xSP M9xSP M10xSP M11xSP M16xSP M10xY M13xY M14xY M15xY	41	L3 L7 L2xD L4xD L24xD L4xT L3xSP L4xSP L2xY L5xY L7xY L13xY L4xP M3 M5 M11 M10xA M3xSP M4xSP M7xSP M10xY M13xY	22
		T	A A2 L12 L17xT L25xY L4xP M1xA M2xA M4xT M8xT M3xP	11	A L16xA L6xT	3
	C	*y1*	A2 L8xA L14xD L2xT L3xT L7xT L10xT L15xY L14xP M1xSP	10	L12xSP M2xD	2
		y2	DxSP TxSP AxDxTxSP P2 L3 L7xA L12xA L14xA L4xD L17xD L2xSP L3xSP L24xSP L2xY L14xY L17xY L18xY M1 M9 M10 M11 M1xT M1xSP M2xSP M5xSP M6xSP M10xSP M12xSP M13xSP M16xSP M2xY M13xY	32	DxSP DxTxSP L4xD L24xSP L10xY L13xY M10xA M11xT M3xSP M5xSP M6xSP M13xSP M13xY	13
		T	A2 L3 L3xT L14xT L15xT L17xT L24xT L22xP L23xP M13 M2xT M7xSP M15xY	13	A L12 L22xA	3
	D	*y1*	T L8xA L3xT L11xT L12xT L14xT L16xT L18xY M7 M6xA M2xD M1xT M2xT M13xT M7xSP	15	L9xA L6xD L25xD	3
		y2	L7 L9 L12 L23 L1xA L5xA L12xSP L17xSP L24xSP L10xY L18xY M3 M7 M13 M2xT M6xT M2xSP M3xSP M5xSP M7xSP M8xSP M10xSP M16xSP M1xY M9xY M15xY M16xP	27	P2 L6xSP L24xSP L18xY M3 M13xA M3xSP M5xSP M9xSP M16xSP M9xY M13xY M15xY M1xP	14
		T	A2 L12 L21 L6xD M7 M4xA M15xD M15xY	8	A L25xP M1	3
	E	*y1*	L3 L8xA L21xT L16xSP L3xY M14xT M13xSP M16xSP	8	T2 L3	2
		y2	P2 L7 L21 L3xA L8xA L17xD L3xT L4xT L14xT L21xT L3xSP L5xSP L23xSP L24xSP L4xY L14xY L19xP M14 M14xA M15xA M3xT M10xT M3xSP M7xSP M9xSP M15xP	26	DxSP P2 L3xA L21xA L3xT L4xSP L24xSP M7 M3xA M9xA M7xT M7xSP M2xP M14xP	14
		T	A2 L21xA L12xT L15xT L17xT L18xT L21xP M5xA M10xT M16xT	10	A L17xT M16xT	3
2013	A	*y1*	T L4xD L17xD L6xT L11xT L12xT L13xT L18xT L14xSP M6xA M15xD M16xD M10xT M10xY	14	L7 L2xT L12xT L19xP M16xT	5
		y2	DxTxSP L3 L12 L14 L21 L3xA L12xA L19xA L4xD L24xD L3xSP L10xSP L14xSP L18xSP L21xSP L23xSP L25xSP L5xY L6xY L14xY L15xY L18xY L6xP L25xP M1 M2 M2xT M10xT M13xT M1xSP M2xSP M10xSP M15xSP M16xSP M13xY M14xY M16xY M4xP M7xP	39	L3 L4xD L4xSP L18xSP L24xSP L6xY L18xY L25xP M1xA M2xA M10xA M1xSP M10xSP M6xP	14
		T	A2 L4xD L11xT L12xT L15xT M3xA M6xA M9xA M16xD M14xT	10	A M2xY	2
	B	*y1*	T L3 L8xT L9xT L11xT L12xT L14xY L18xY L18xP M6xA M8xA M9xA M14xT	13	L21 L4xD L15xT M6 M1xA	5
		y2	AxTxSP P2 L4 L4xD L19xT L3xSP L6xSP L18xSP L19xSP L24xSP L2xY L13xY L5xP L14xP L18xP L23xP M6 M13 M14 M5xA M1xD M1xT M2xSP M3xSP M4xSP M5xSP M6xSP M7xSP M9xSP M10xSP M11xSP M9xY M10xY M9xP	34	L3 L13xA L4xD L24xD L3xSP L4xSP L24xSP L2xY L13xY L14xP M3 M10 M2xSP M3xSP M4xSP M7xSP M11xSP	17
		T	A2 L12 L15xT L17xT L25xY L4xP L25xP M1xA M2xA M8xT M1xY	11	A L16xA	2
	C	*y1*	A2 L1xA L14xD L24xD L2xT L3xT L7xT L19xT L25xT L2xSP L24xSP L14xP M16xD	13	L2xA L25xD L24xT L24xP M16xD M1xSP	6
		y2	DxSP AxTxSP AxDxTxSP P2 L1 L3 L7 L1xA L3xSP L15xSP L17xSP L24xSP L2xY L4xY L10xY L12xY L15xY L17xY L18xY L25xP M2 M1xA M11xA M12xA M1xT M7xT M1xSP M3xSP M5xSP M7xSP M8xSP M10xSP M11xSP M12xSP M16xSP M2xY M10xY M13xP	38	P2 L3 L4xD L3xT L3xSP L17xSP L24xSP L5xY L13xY L23xP M3xA M10xA M11xA M1xSP M3xSP M7xSP M11xSP M13xY M14xY	19
		T	A2 L3 L14xT L15xT L6xY L22xY L23xY M13 M13xT M2xSP M7xSP	11	A L8xA L9xA L22xA	4
	D	*y1*	T L8xA L13xA L2xT L3xT L11xT L14xT L16xT L18xY M7 M6xA M2xD M9xD M1xT M2xT M13xT M7xSP	17	A L6xD L10xD L12xT M15 M6xA M8xA M13xT M14xT M15xT M6xY M10xY M13xY M15xY M6xP M16xP	16
		y2	TxSP L8 L13 L23 L5xA L23xA L5xD L2xT L2xSP L12xSP L15xSP L17xSP L23xSP L24xSP L4xY L10xY L13xY L18xY M3 M7 M13 M13xA M2xD M7xT M3xSP M5xSP M8xSP M9xSP M16xSP M1xY M9xY M10xY M15xY	33	P2 L5xD L2xSP L23xSP L24xSP L10xY L13xY L18xY L5xP M3 M10xT M2xSP M3xSP M5xSP M7xSP M9xSP M16xSP M1xY M9xY M13xY M14xY M15xY	22
		T	A2 L21 L16xA L6xD M4xT M15xY	6	A L6xD L17xD L17xY L25xP M1 M1xD	7
	E	*y1*	L3 L8xA L11xT L14xT L21xT L2xSP L3xY M14xT M16xSP	9	T2 L3	2
		y2	P2 L7 L21 L3xA L3xT L3xSP L4xSP L6xSP L17xSP L23xSP L24xSP L6xY L19xP M3 M14 M15 M1xD M3xT M10xT M16xT M3xSP M4xSP M7xSP M9xSP M10xSP	25	L21 L3xA L6xD L3xSP L4xSP L24xSP L19xP M14 M3xT M4xSP M7xSP M8xSP M3xP M15xP	14
		T	A2 L21xA L17xT L18xT L21xP M5xA M16xT	7	A L9xP M6xP	3
			A T L4xD L17xD L18xD L24xD L11xT L12xT L13xT L15xT L14xSP M2xA M6xA M10xT M13xT	15	L6xD L2xT L9xP	3

2014	A	*y1*	DxSP L12 L15 L18 L21 L3xA L4xD L5xD L23xD L18xSP L21xSP L23xSP L24xSP L25xSP L6xY L15xY L6xP L21xP L25xP M2 M3xA M1xT M3xT M1xSP M2xSP M3xSP M4xSP M11xSP M16xSP M14xY M13xP	31	DxSP L17 L4xD L18xSP L24xSP L15xY L18xY L18xP L25xP M1xA M10xA M3xT M10xSP	13
		y2	A2 L3 L4xD L15xT M3xA M6xA M13xT	7	A L9xA L17xD	3
		T	A T L3 L1xT L7xT L8xT L11xT L12xT L18xY M2 M6xA M8xA M10xA M1xT M14xT	15	L21 L4xD L15xT M11 M1xD M10xD M1xT	7
	B	*y1*	L4 L19xA L24xD L4xT L18xT L3xSP L6xSP L15xSP L18xSP L24xSP L5xY L7xY L10xY L13xY L2xP L14xP L18xP L23xP M3 M4 M7 M10 M11 M1xA M5xA M1xT M3xSP M4xSP M5xSP M6xSP M7xSP M10xSP M11xSP M9xY M10xY M16xP	36	P2 L3 L5 L4xD L3xSP L4xSP L7xY L15xY M3 M8 M11 M10xA M16xD M3xSP M7xSP M8xSP M10xY	17
		y2	A2 L3 L12 L20xA L25xY L25xP M1xA M2xA M4xT M8xT M1xY	11	A T2 L7xA	3
		T	A L1xA L14xD L2xT L3xT L7xT L9xT L11xT L16xT L25xT L24xSP L15xY L11xP M2 M7 M5xA M16xD M16xT M15xSP	19	L2xA M10xD M16xD M10xY	4
	C	*y1*	TxSP P2 L3 L7 L7xA L17xT L18xT L3xSP L14xSP L17xSP L24xSP L2xY L4xY L5xY L15xY L17xY L23xP L25xP M1 M11 M10xA M12xA M1xSP M2xSP M5xSP M10xSP M11xSP M16xSP M2xY M10xY M13xY	31	P2 L3 L4xD L17xSP L24xSP L5xY L13xY L18xY M11xA M1xSP M7xSP M11xSP M2xY M10xY M13xY M10xP	16
		y2	A2 L3 L14xT L15xT L24xT L6xY L22xY L23xY L6xP L22xP M13 M9xD M2xT M8xT M2xSP M7xSP M15xY	17	A2 L22xA L18xD L22xT	4
		T	L8xA L24xD L2xT L3xT L11xT L12xT L16xT L18xY M7 M11xA M2xD M9xD M1xT M13xT	14	A2 L3 L21 L3xD L14xD L2xY M6 M8 M13 M15 M8xA M13xT M14xT M15xT M13xY M15xY M16xY M2xP M6xP M10xP	20
	D	*y1*	L1 L5xA L13xA L23xA L2xSP L12xSP L17xSP L23xSP L24xSP L10xY L13xY L18xY L5xP M3 M7 M13 M14xA M2xD M3xSP M5xSP M7xSP M8xSP M9xSP M16xSP M1xY M13xY M9xP	27	A2 P2 L17xSP L24xSP L10xY L15xY M3xA M13xA M3xSP M5xSP M9xSP M16xSP M1xY M13xY M15xY	15
		y2	A2 L21 L1xA L14xD L18xD L10xY M3xA M4xA M3xT M15xY	10	A L6xD L21xD L1xY L1xP L25xP M1 M1xD M9xD	9
		T	L3 L11 L8xA L20xA L14xT L21xT L15xSP L24xSP L24xY M7xT M2xSP	11	L2 L3 L20xA L14xP M4xA	5
	E	*y1*	P2 L3 L21 L3xA L7xA L2xT L3xT L3xSP L4xSP L23xSP L24xSP L2xY L10xY L19xP M3 M14 M5xD M3xT M7xT M8xT M10xT M1xSP M3xSP M4xSP M7xSP M9xSP M12xSP M14xSP M15xP	29	P2 L3 L3xA L3xSP L24xSP M7 M8 M9 M14xA M3xT M7xT M7xSP M3xY	13
		y2	A2 L21xA L12xT L21xP M5xA M16xT M14xY	7	A L22xT	2
		T	A L4 L3xD L5xD L24xD L6xT L8xT L11xT L12xT L13xT L15xT L10xSP L10xY M1 M2xA M6xA M10xT M13xT M14xT M15xSP M6xY M2xP	22	L6xD L2xT L11xT L12xT M1xT M15xT M13xSP M13xY	8
2015	A	*y1*	DxSP L1 L3xA L9xA L21xA L4xD L10xSP L18xSP L23xSP L24xSP L25xSP L2xY L10xY L13xY L15xY L25xP M2 M3 M5 M1xA M10xT M15xT M1xSP M2xSP M3xSP M7xSP M8xSP M10xSP M11xSP M15xSP M16xSP M15xY M16xY	33	DxSP L3 L14xA L24xSP L13xY L15xY L18xY L23xY M1xA M10xT M10xSP M10xY M6xP M10xP	14
		y2	A2 L4xD L4xT L15xT L10xY M3xA M6xA M9xA M13xT M10xY	10	A T2 L4xD	3
		T	A T L3 L24xD L1xT L11xT L12xT L24xY M3xA M6xA M1xD M1xT M14xT M10xY	14	L4xD L4xT L15xT M1xD M1xT	5
	B	*y1*	L5 L19xA L2xD L4xD L6xD L17xD L24xD L4xT L3xSP L4xSP L6xSP L18xSP L24xSP L1xY L2xY L13xY L18xY L2xP L14xP L18xP M6 M1xD M1xT M5xT M3xSP M4xSP M5xSP M6xSP M7xSP M10xSP M11xSP M9xY M10xY M13xY M14xY	35	L3 L9 L4xD L4xT L5xT L4xSP L24xSP L13xY L15xY L18xY M3 M11xA M10xT M3xSP M10xY M13xY	16
		y2	A2 L3 L12 L25xP M4xA M5xA M4xT M8xT M1xY M2xY M13xY	11	A T2	2
		T	A L1xA L8xA L4xD L5xD L14xD L2xT L3xT L9xT L16xT L25xT L10xY L15xY L11xP M2 M10xD M7xT M13xT	18	L2xT L3xT L6xT M10xD M16xD	5
	C	*y1*	DxSP AxDxTxSP P2 L17xT L3xSP L14xSP L17xSP L24xSP L2xY L5xY L10xY L4xP L23xP L25xP M1 M10 M11 M12xA M5xT M1xSP M5xSP M7xSP M10xSP M11xSP M12xSP M16xSP M2xY M13xY M14xY	29	DxSP M4xSP M11xSP	3
		y2	A2 Y2 L3 L25xD L6xT L14xT L24xT L19xY L22xP M13 M9xD M2xT M13xT M2xSP M3xY M15xY	16	A2 L20xA L16xD L22xT M1xD	5
		T	A L9xA L2xT L3xT L11xT L12xT L16xT L18xY M2xD M9xD M1xT M2xT M13xT M16xSP	14	A2 L3 L21 L3xD L12xT L2xY M2 M6 M10 M13 M15 M14xT M13xY M15xY M16xY M2xP M6xP M16xP	18
	D	*y1*	L19 L5xSP L6xSP L17xSP L24xSP L3xY L13xY L18xY L5xP L18xP M2 M3 M14 M13xA M2xSP M3xSP M5xSP M8xSP M9xSP M16xSP M14xY M15xY	22	L9 L6xSP L24xSP L3xY M3 M9 M3xSP M5xSP M7xSP M9xSP M1xY M13xY M15xY	13
		y2	A2 L14 L21 L19xD L10xY L10xP M3xA M7xD M3xT M4xT M15xY	11	A T2 L6xD L21xD L25xT L1xY L1xP L25xP M9xD	9
		T	L20xA L2xT L3xT L11xT L14xT L21xT L10xSP L14xSP L18xSP L24xSP L24xP M16xD M14xT M2xSP M2xY M15xY	16	L3 L20xA L2xT L6xT L14xT L17xT L21xT L24xT M12xD	9
	E	*y1*	L3 L7 L21xA L11xD L3xSP L4xSP L5xSP L23xSP L24xSP L6xY L14xY L19xP M7 M8 M14 M3xA M3xT M7xT M2xSP M4xSP M7xSP M9xSP M12xSP M10xY M15xP	25	P2 L3 L24xSP M3xT M7xT M7xSP M2xY M14xY	8
		y2	A L20xA L4xT L11xT L14xT L15xT L21xT L4xP L10xP M13xT M15xY	11	L2 L3 L8xA L20xA L2xT L14xT L21xT	7
		T	L3 L8 L25 L7xA L3xD L11xD L24xD L9xT L21xT L4xP L19xP L23xP M3 M5 M7 M8 M14 M14xA M2xT M14xT	20	P2 L3 L7xY L22xP M7 M5xA M8xA M3xT M14xY	9

A: Age, T: Time being client of the bank, D: distance, SP: Same Postal code; and Y: Years of education (village or city), P: Poverty (village or city). The letter L followed by a number stands for one of 25 variables at the village or city level. For example, L1 stands for the sex ratio (see figure A.2). For municipality level variables, the letter M stands for one of 16 variables.

A.2 Additional Information: Regression Discontinuity Design

Outcomes and Controls

Table A.1 provides summary statistics for the regression discontinuity design, detailing outcomes and controls. Column (1) presents the means and standard deviations for the sample of 395 branches, column (2) for the 250 below the 50,000 cutoff, and column (3) for the 145 above it. Column (4) provides difference in means and the corresponding standard error. The table contains three sets of variables. Summarizing the forcing variable and the proportion of branches that offered Premiahorro, the first set describes the design. The second set provides two outcomes for the accounts that anyone can open: number per month of accounts and of transactions. Means of outcomes consider two periods. The first covers 2010 to 2017, from the start of the final design to the end of the available information.[67] The second covers 2008, a year before the preliminary design began. The third set consists of predetermined variables corresponding to the location of the branch. At the village or city level, the variables are poverty score, years of education, sex ratio, and the number of families in Oportunidades as proportion of all families. At the municipality level, the variables are savings per capita and number of ATMs per km^2 per 100,000 people.

Since Premiahorro started, and relative to branches above the cutoff, branches offering Premiahorro had substantially more accounts and transactions, suggesting plausible positive effects of the program. But the naïve comparison belies how different branches below the cutoff are from those above it. For branches above the cutoff the mean population at their location is 30 times greater. For them poverty, education, savings, and banking infrastructure are significantly and substantially different.

Internal validity of the regression discontinuity design

The sharp discontinuity design of Premiahorro is internally valid. In a seminal contribution, McCrary (2008) notes that when agents sort themselves around the cutoff point—when they manipulate their way into the winning side—the design is invalid. He proposes testing the continuity of the density of the forcing variable at the cutoff point.[68] To test continuity of

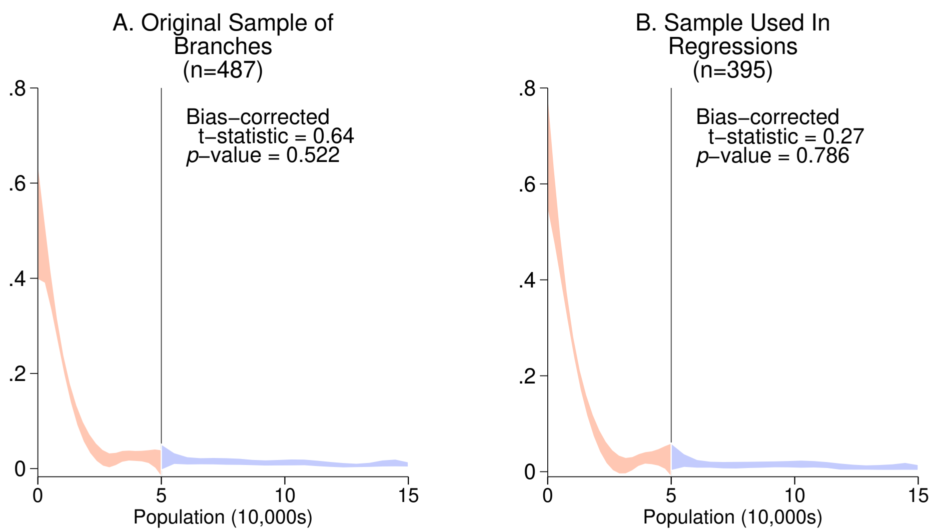
[67]The period excludes the first trimester of 2010. That trimester a glitch in the information system prevented to record information on Premiahorro. Premiahorro started in the second trimester of 2009, likely causing the glitch.

[68]The density for each agent is unobservable (only a single realization per agent is observed). But if the density

population at the 50,000 cutoff point, I use Cattaneo et al. (2020) non-parametric density estimator test. Because it is implemented in an automatic, data-driven way, their test improves upon the one proposed by McCrary (2008).[69]

Figure A.4 presents the results. Panel a for the original sample of 487 branches and panel b for the restricted sample of 395 branches. The shaded areas represent the 95 confidence interval of densities of the forcing variable estimated at each side of the cutoff point. Shaded areas around the cutoff point suggest continuity, and the formal test cannot reject the null hypothesis of continuity. Manipulation is not a concern. The design is internally valid.[70]

Figure A.4: Manipulation Test
Cattaneo et al. (2020b)



H_o : The density of the forcing variable is continuous at the cutoff point.

for each agent is continuous, then, McCrary argues, the marginal density of the forcing variable over the population should be continuous too. If it is continuous, the design should be valid. Because aggregates over agents can confound subgroup systematic behavior, early practical guides to researchers advised them that continuity of forcing variable at the cutoff point is not needed in principle. Its absence, however, does cast doubt on the validity of the design (Imbens and Lemieux, 2008; Lee and Lemieux, 2010).

[69] McCrary's test requires pre-binning the data (it requires selecting the width of the bars of the histogram of the forcing variable). Results are sensible to the bandwidth researchers select to pre-bin the data. No clear criterion informs how to select it.

[70] Manipulation in Premiahorro cannot be direct. Branches cannot sort themselves around the cutoff point. But Bansefi could have chosen the 50,000 population cutoff point in a non-arbitrary way. This is unlikely. The round number is half the official definition in Mexico of a large city. And the cutoff also splits the total number of Bansefi branches by half.

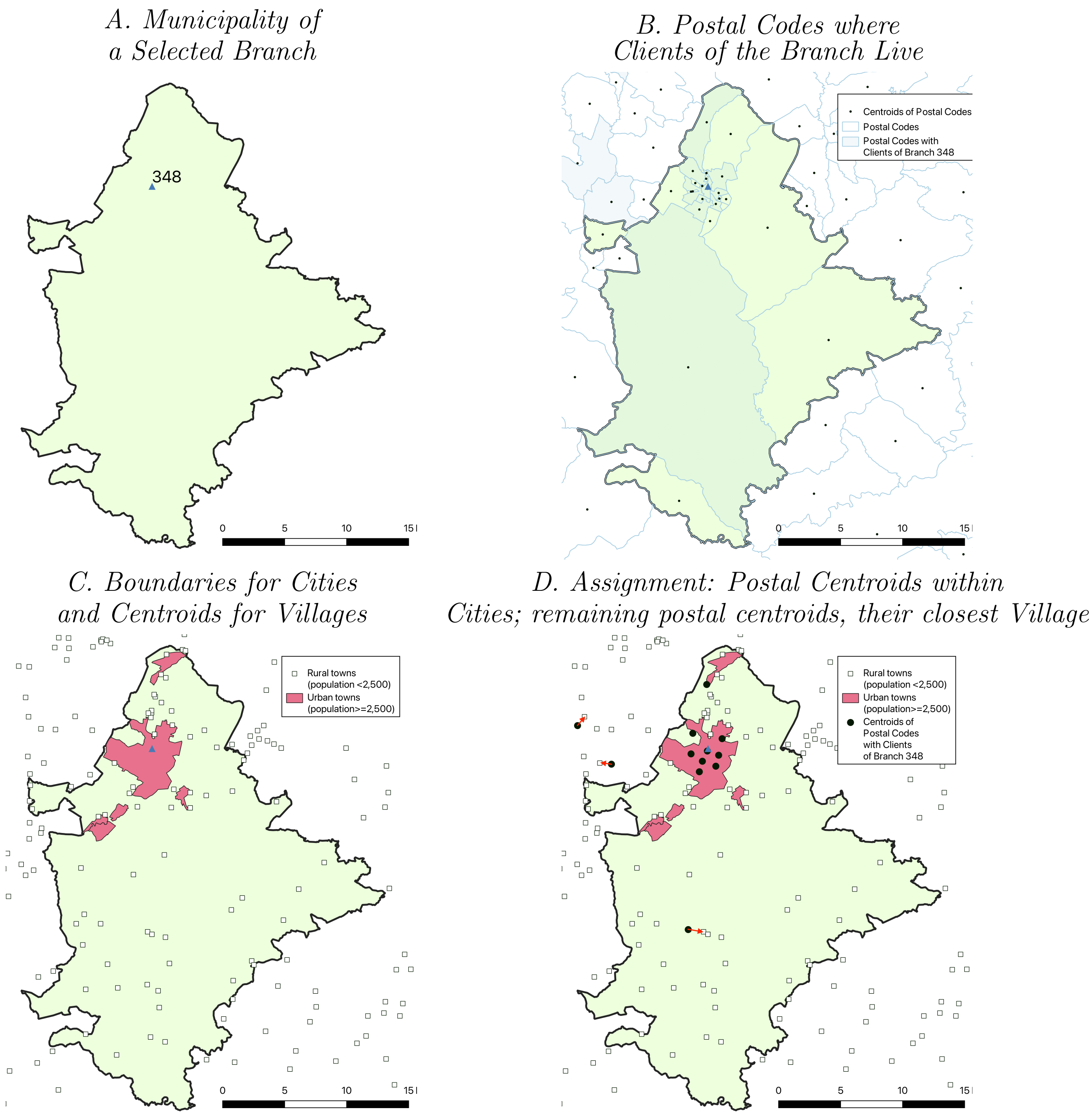
Table A.1: Summary Statistics
Regression Discontinuity Design

	Population (2005)			Difference
	All	< 50,000	≥ 50,000	Below - Above
	Mean [SD]	Mean [SD]	Mean [SD]	Est. (SE)
	(1)	(2)	(3)	(4)
Regression discontinuity design				
Forcing: Population (10,000s, 2005)	15.06	1.24	38.90	-37.67***
	[29.62]	[1.11]	[38.66]	(3.21)
(Population<50,000)=1	0.63	1.00	0.00	1.00***
	[0.48]	[0.00]	[0.00]	(0.00)
Outcomes				
Accounts (1000s)				
- During and after: 2010m4–2017m6	3.57	3.23	4.17	-0.94***
	[2.09]	[1.74]	[2.49]	(0.23)
- Before: 2008m1–2008m12	4.07	2.39	6.96	-4.57***
	[4.01]	[1.82]	[4.99]	(0.43)
Total transactions				
- During and after: 2010m4–2017m6	983.19	1028.28	905.45	122.84**
	[510.75]	[475.87]	[559.08]	(55.30)
- Before: 2008m1–2008m12	818.3	679.2	1058.1	-378.92***
	[546.6]	[383.9]	[686.5]	(61.92)
Predetermined (branch location)				
- Village/City: Poverty score (2005)	-1.25	-1.09	-1.54	0.46***
	[0.37]	[0.34]	[0.17]	(0.03)
- Village/City: Years of education (2010)	8.2	7.5	9.3	-1.82***
	[1.4]	[1.2]	[0.9]	(0.11)
- Village/City: % of Oport. families (2010)	0.19	0.25	0.08	0.17***
	[0.16]	[0.17]	[0.06]	(0.01)
- Municipality: Total savings per people,	14.33	4.82	30.72	-25.89***
all deposits, all banks (1000s, 2009-I)	[42.67]	[15.16]	[64.48]	(5.43)
- Municipality: # of ATMs,	0.09	0.04	0.17	-0.13***
per km2 per 100,000 people (2009-I)	[0.24]	[0.16]	[0.32]	(0.03)
- Municipality: Lenght of roads,	0.31	0.33	0.26	0.07**
per km2 (2009)	[0.28]	[0.30]	[0.22]	(0.03)
Observations	395	250	145	

*p<0.10, **p<0.05, ***p<0.01. Standard deviations in brackets and standard errors in parentheses. Account and transactions correspond to the following accounts: Basic savings accounts with or without debit card, commitment savings accounts, and Premiahorro. The final design of Premiahorro was implemented in 2010–2015. Information for Premiahorro for the first trimester of 2010 was not recorded on the financial records of the bank. Predetermined variables are means for the population of the village or city, or municipality, in which the branch is placed. No municipality has branches both offering and not offering Premiahorro. The 250 villages or cities that offered Premiahorro are within 249 municipalities; the 145 not offering are within 121.

A.3 Additional Tables and Figures

Figure A.5: Process to Match Postal Codes to Cities or Villages



The branch is ubicated in a municipality with 73000 people in 2010 (panel a). Forty four villages (mean population of 260), five small cities (mean population of 3150), and a medium-sized city of 43000 people compose the municipality. The branch is ubicated in the medium-sized city. Clients of the branch live in the postal code of the branch, the ones surrounding it, and in few close but outside the municipality (panel b). For each postal code, I calculate the centroid and overlay them to the shapefile of cities and villages (panel c). For cities the shapefiles provide boundaries, and for villages a single point. I match in two steps (panel b). In the first, postal codes whose centroid is within the boundaries of a city get assigned to the city and get removed from the matching process. In the second, the remaining postal codes are matched to the village closest to its centroid.

Figure A.6: Population Quintiles of the Sample Below and Above the Cutoff

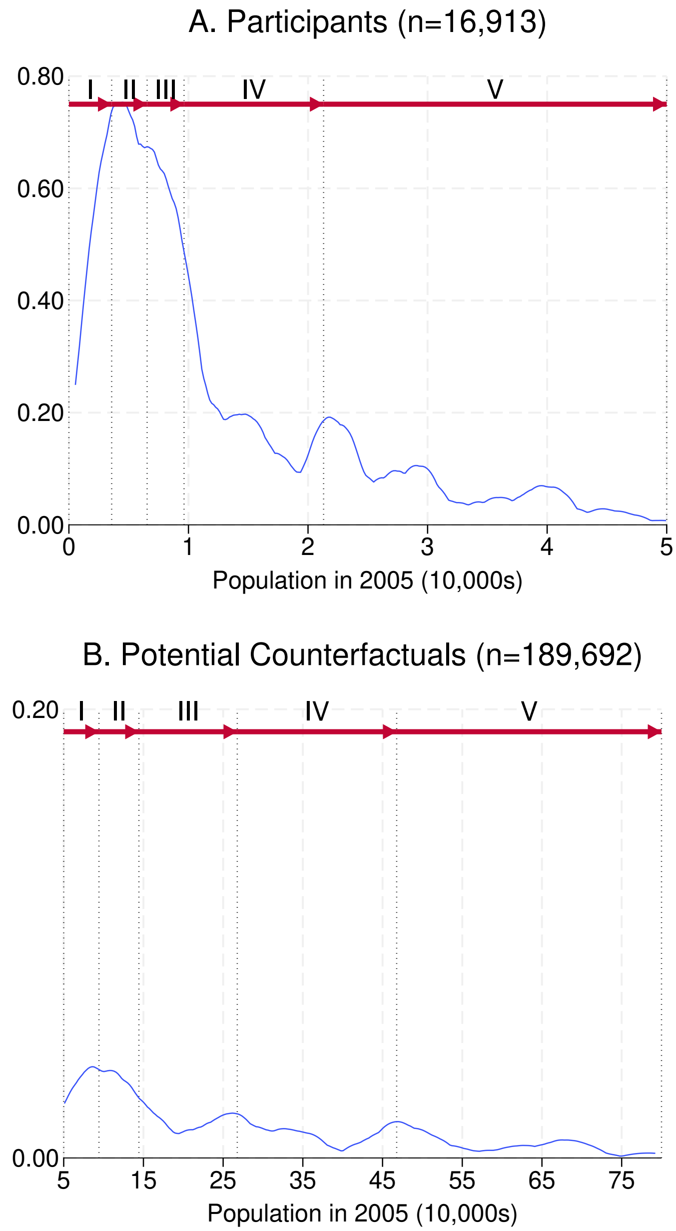
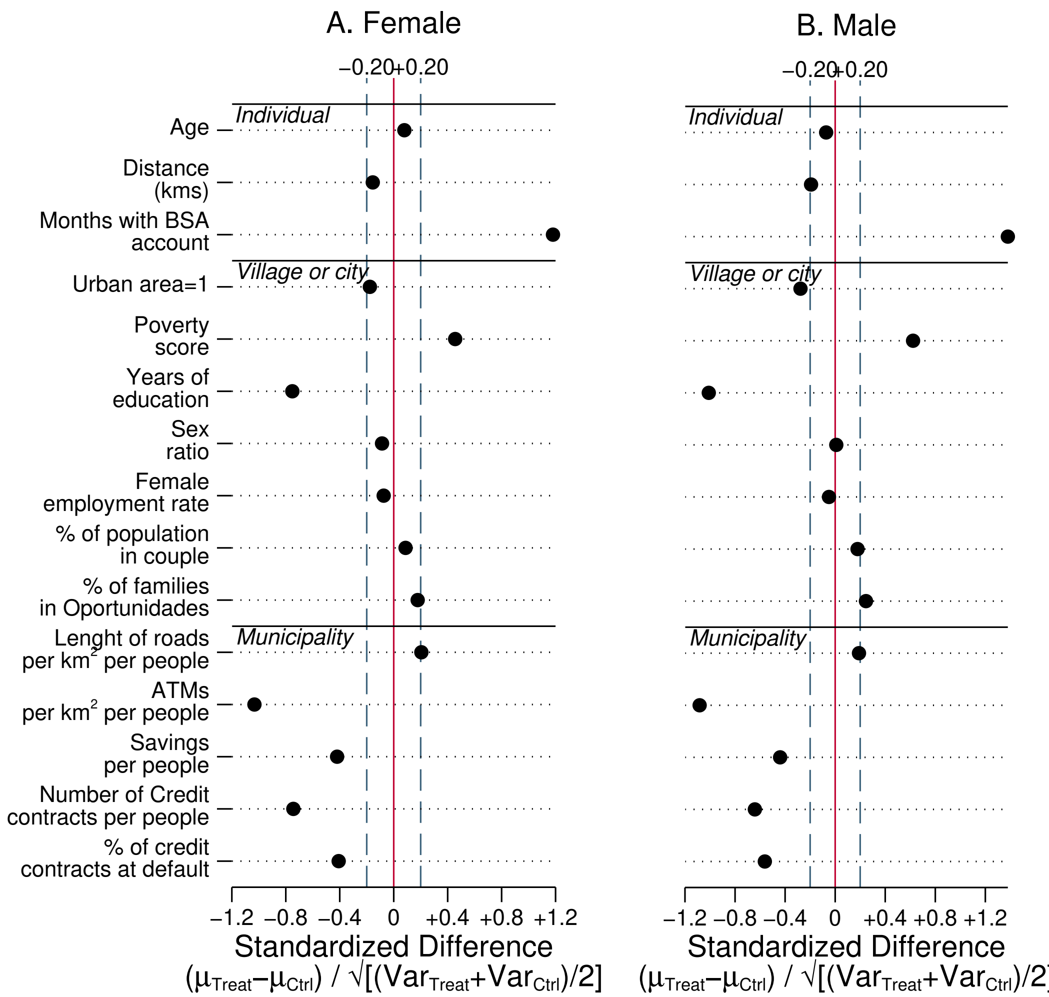


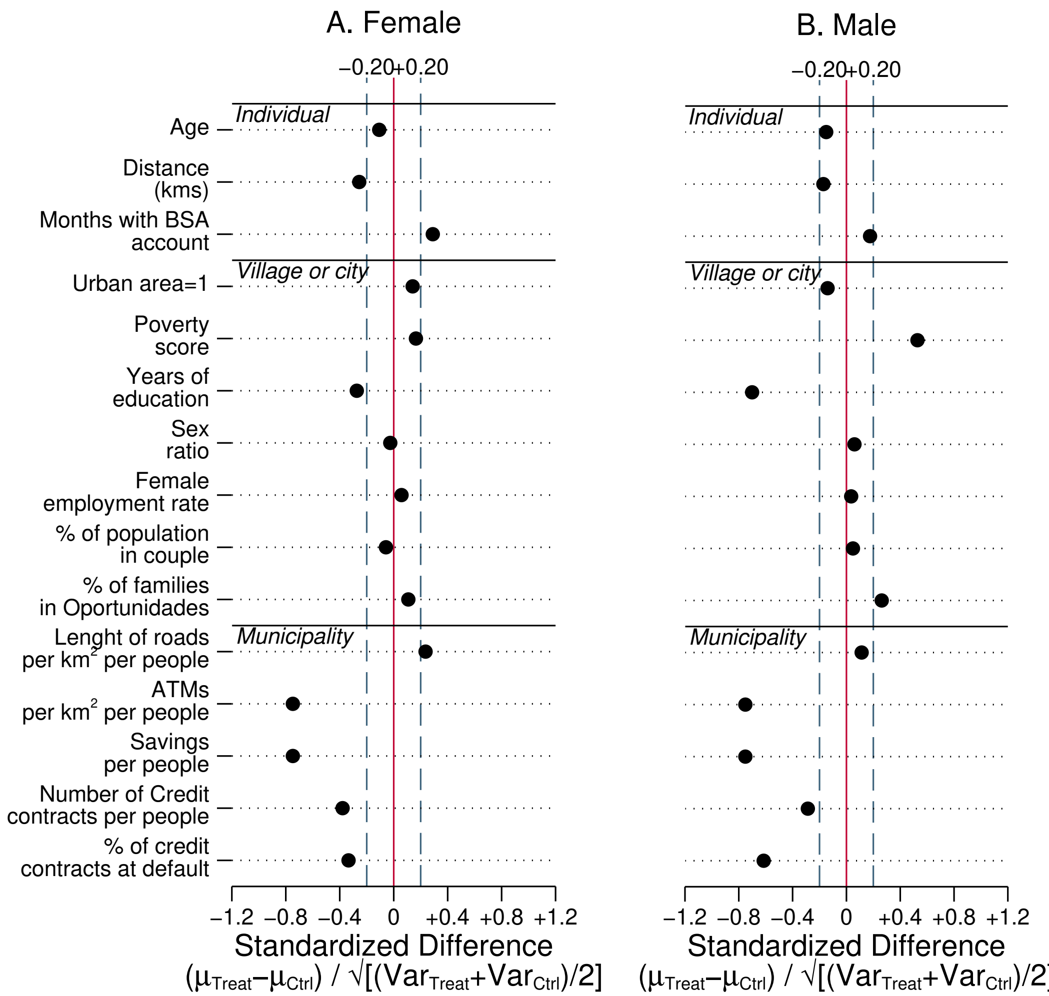
Figure A.7: Standardized Differences Before Matching



Individual characteristics and characteristics of the place where people live.

Sample. Female: Participants (n=14,692), Potential counterfactuals (n=142,969); Male: Participants (n=2,221), Potential counterfactuals (n=46,723)

Figure A.8: Standardized Differences After Matching



Individual characteristics and characteristics of the place where people live. Matched sample: Females: Participants (n=14,658 [99%]) and Non-participants (n=14,658 [10%]); Male: Participants (n=2,196 [99%]), Potential controls (n=2,196 [5%]). Proportion of the sample before matching in brackets. Sample regression analysis used when the outcome is savings balances.

Table A.2: Results: Regression Discontinuity Design
Triangular Kernel, Polynomial of Degree 1
Predetermined Variables

	Village or City			Municipalities		
	Poverty	Education	Oportunidades	Savings	Branches	Roads
Estimate ($\hat{\tau}_{SRD}$)	0.107	0.105	0.046	-2.167	0.013	-0.079
Control mean	-1.479	8.858	0.097	14.949	0.101	0.285
Effect size	0.61	0.12	0.66	-0.21	0.05	-0.41
h_{MSE}^{Below}	-2.273	-2.261	-2.589	-2.468	-3.043	-2.333
h_{MSE}^{Above}	26.330	16.310	41.163	11.932	15.033	16.253
Observations	115	98	132	91	117	98
<50,000	32	32	34	33	52	32
>50,000	83	66	98	58	65	66
95% CI (MSE-Opt)						
Lower	-0.035	-0.453	-0.044	-12.801	-0.064	-0.326
Upper	0.225	0.840	0.130	10.155	0.112	0.188
p-value	[0.152]	[0.557]	[0.336]	[0.821]	[0.586]	[0.598]
95% CI (CE)						
Lower	-0.074	-0.288	-0.026	-13.732	-0.048	-0.334
Upper	0.205	0.964	0.156	10.728	0.125	0.264
p-value	[0.356]	[0.290]	[0.162]	[0.810]	[0.381]	[0.817]

p-value in brackets, *p<0.10, **p<0.05, ***p<0.01.

Control mean equals the mean of the outcome over the right-side of the cutoff. Effect size is the estimated effect divided by standard deviation outcome over the right-side.

Village or city: Poverty score (2005), years of education (2010), and % of families in Oportunidades (2010).

Municipality (2009): Savings in all banks, number of branches per km^2 per 100,000 people, and length of roads per km^2. No municipality has branches both offering and not offering Premiahorro. The 250 villages or cities that offered Premiahorro are within 249 municipalities; the 145 not offering are within 121.