

# Estimating Treatment Effects with Big Data When Take-up is Low: An Application to Financial Education

**Gabriel Lara Ibarra, David McKenzie, and Claudia Ruiz-Ortega**

## Abstract

Low take-up of interventions is a common problem faced by evaluations of development programs. A leading case is financial education programs, which are increasingly offered by governments, nonprofits, and financial institutions, but which often have very low voluntary participation rates. This poses a severe challenge for randomized experiments attempting to measure their impact. This study uses a large experiment on more than 100,000 credit card clients in Mexico. The study shows how the richness of financial data allows combining matching and difference-in-difference methods with the experiment to yield credible measures of impact, even with take-up rates below 1 percent. The findings show that a financial education workshop and personalized coaching result in a higher likelihood of paying credit cards on time, and of making more than the minimum payment, but do not reduce spending, resulting in higher profitability for the bank.

**JEL classification:** D14, G21, G28, O12

**Keywords:** financial literacy, credit-card behavior, low take-up

## 1. Introduction

There has been rapid growth in the use of randomized experiments to measure the impact of different development interventions. In their simplest form, a treatment group is offered the program and is compared to a control group that is not offered the program. However, a common issue facing many of these experiments is incomplete compliance—not all of those offered the program take it up (and sometimes those who are not offered the program manage to receive it anyway). The standard solution to this problem

Claudia Ruiz-Ortega (corresponding author) is an economist in the Development Research Group at the World Bank, Washington, DC; her email address is [cruizortega@worldbank.org](mailto:cruizortega@worldbank.org). Gabriel Lara Ibarra is senior economist in the Poverty and Equity Global Practice at the World Bank; his email address is [glaraibarra@worldbank.org](mailto:glaraibarra@worldbank.org); David McKenzie is lead economist in the Development Research Group at the World Bank; his email address is [dmckenzie@worldbank.org](mailto:dmckenzie@worldbank.org). The authors thank Adolfo Albo, Juan Luis Ordaz, David Cervantes, and the BBVA Bancomer Financial Education Team for their collaboration on the experiment and for providing the anonymized customer data used in this research. The authors would also like to thank the editor, two anonymous referees, and participants at the EduFin Summit 2017 in Mexico City for comments and suggestions. No funding was received for this work from BBVA Bancomer, and independence was maintained in the analysis and reporting of the results. The findings, interpretations, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the World Bank, its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. A supplementary online appendix for this article can be found at *The World Bank Economic Review* website.

has been to either focus on estimating the effect of being randomly offered the program (the intention-to-treat effect), or to use random assignment of the program offer as an instrumental variable for program receipt, estimating the local average treatment effect. However, the statistical power to detect these treatment effects typically falls as the rate of noncompliance increases, and at very low take-up rates even large sample experiments may be uninformative about treatment impacts.<sup>1</sup> Take-up rates of 5 percent or lower have indeed been common among several types of interventions.<sup>2</sup> Examples include rainfall index insurance (Giné, Townsend, and Vickrey 2008), health insurance (Chemin 2018), firm formalization assistance (Bruhn and McKenzie 2014), wage subsidies (Levinsohn et al. 2013), and the case that is studied here, of financial education.

Financial education is of policy interest due to fact that low levels of financial literacy are pervasive in both developed and developing countries. Higher financial literacy is associated with a wide range of better financial decisions, including more retirement planning, greater stock market participation, and higher savings (Lusardi and Mitchell 2014). Credit card users with low levels of financial literacy are more likely to carry balances, pay only the minimum payment, or incur late fees (Mottola 2013; Lusardi, and Tufano 2015). As a result, a large number of governments, international organizations, nonprofits, and financial institutions have launched efforts to provide financial education in many countries around the world (Fernandes, Lynch, and Netemeyer 2014; Miller et al. 2015).

However, voluntary participation in many financial education efforts is often very low. Willis (2011, p. 230) notes that “voluntary financial education is widely available today, yet seldom used.” For example, Brown and Gartner (2007) report on experimental efforts by different credit card providers in the United States that aimed to provide online financial literacy training to delinquent and at-risk credit card holders: Target Financial Services made calls to 80,982 cardholders, reached only 6,417 of them, offered half of them the program, and had only 28 log in, and only 2 people completed the course; U.S. Bank had only 384 cardholders out of the 42,000 it attempted to reach complete its online program. In Mexico Bruhn, Lara Ibarra, and McKenzie (2014) sent 40,000 letters to bank clients to get them to enroll in a financial education course, of which only 42 responded with interest in the course; they also displayed 16 million Facebook ads, but received only 119 responses. In Peru, Chong, Karlan, and Valdivia (2010) abandoned a randomized experiment after only 7 percent of their treatment group listened regularly to a radio program with financial education messages, despite being given financial incentives to do so. Low take-up rates are also common among many offers of financial products and services by financial institutions. For example, in the United States the response rate to direct mail credit card solicitations fell from 2.2 percent to 0.6 percent between 1991 and 2012 (Grodzicki 2015).<sup>3</sup>

Such low take-up rates present a severe challenge to randomized experiments attempting to measure the impact of financial education on those who do participate. Due to the inverse-square rule, an experiment with 1 percent take-up requires 10,000 times the sample as an experiment with 100 percent take-up in order to have the same statistical power.<sup>4</sup> This can be one reason why many experimental evaluations

- 1 The extent at which power falls depends on the degree of heterogeneity in treatment effects, and how they are correlated with the take-up decision. Section 3 and the supplementary online appendix S1, available with this article at *The World Bank Economic Review* website, provide details.
- 2 Even with higher take-up rates in the treatment group, power is also low when the treatment-control difference in take-up rates is low, as occurs in many randomized encouragement designs (e.g., Goldstein 2011); and in cases where close substitutes to treatment are available, as in the microfinance evaluation of Banerjee et al. (2015) where the treatment-control difference in take-up of microcredit is only 8.4 percentage points.
- 3 Karlan, Mordoch, Mullainathan (2010) also note that low take-up is an issue for many evaluations attempting to measure impacts of savings, loan, and insurance products offered by microfinance institutions.
- 4 This is true for a homogeneous treatment effect, or if treatment heterogeneity is uncorrelated with take-up. Section 3 and online appendix S1 show that even when treatment heterogeneity is positively correlated with take-up, at take-up rates of 5 percent or lower, power will still be low and similar to that of homogeneous treatment effects in many realistic cases.

of financial literacy training struggle to find a significant effect (Fernandes, Lynch, and Netemeyer 2014; Miller et al. 2015). Yet there is often still interest on the part of researchers and financial institutions in learning whether the program had impacts for those individuals who do choose to participate. Moreover, while take-up rates may be low in response to specific offers, widespread availability can still mean that the total number of program participants can be large—just as many people have credit cards despite few people responding to any particular credit card offer.

The key question this paper attempts to address is whether it is still possible to obtain reliable measures of the impact of financial education when take-up rates are too low to enable estimation by experimental methods alone. The context is an experiment in Mexico, whereby the bank BBVA Bancomer worked with over 100,000 of its credit card clients, inviting the treatment group to attend its financial education program *Adelante con tu futuro* (Go ahead with your future). The program has had over 1.2 million participants between 2008 and 2016, yet only 0.8 percent of the clients in the treatment group attended the workshop. A second experiment, which tested personalized financial coaching also had low take-up, with 6.8 percent of the treatment group actually receiving coaching. Standard experimental estimation then finds no significant impact of either treatment on the key program outcomes, but with very wide confidence intervals for the estimated treatment effect on the treated (the impact of actually receiving training or coaching).

Nevertheless, this paper argues that the richness of financial data allows combining non-experimental methods with the random assignment from the experiment in a way that yields credible estimates of the treatment impact for those who do take up financial education. It shows that those who participate in financial education tend to pay more than the minimum payment on their credit cards, which is in line with the claim of Willis (2011) that those who participate voluntarily tend to already have more financial knowledge and better financial practices than those who do not. This means that simple comparisons of those who participate to those who do not will be biased. Instead, this study uses the rich time-series data that are available on credit card clients to match those who take up the workshops or coaching in the treatment group to clients in the control group who display similar levels and trends in key outcomes month after month for 16 months pretreatment. Matched difference-in-differences are then used to estimate the treatment impacts. This approach helps overcome several common concerns about the use of matching and difference-in-differences: the assumption of common trends becomes more plausible when it can be shown the two groups that are selected displayed similar behavior for 16 time periods beforehand; while the question of why matched individuals in the control group did not take up treatment if they are so similar to the treated has a ready answer in that they were randomly assigned to not be invited.

It is found that both attending the financial education workshops and receiving personalized coaching do lead to changes in financial outcomes for participants. Attending the workshop results in a 10.7 (s.e. 3.4) percentage point increase in the likelihood of paying more than the minimum payment; a 3.4 (s.e. 0.8) percentage point lower likelihood of not paying by the payment due date; and a 2.7 (s.e. 1.3) percentage point increase in the likelihood of also having a deposit account at the bank. The economic magnitude of the effects of coaching is similar and statistically significant, but less precisely estimated due to smaller samples. While paying more responsibly, clients do not reduce the amount of credit card spending, and actually spend more. The result is that financial education of either type leads to an increase in the likelihood that the client is considered profitable by the bank. This demonstrates an additional motive for financial institutions to offer financial education to their clients, beyond the social responsibility motives typically given.

Since the financial education programs were designed by the bank itself, a natural concern is the existence of conflicts of interest, as participants of the programs are also clients of the bank, using its products and services. This study takes an agnostic view on how aligned the content of the financial education interventions is with the needs of the participants, and focuses only on identifying changes in financial behavior as a result of the programs, without making any claims on welfare. This study's ability to

conduct any welfare analysis is hindered by the fact that while rich time-series data are available from BBVA Bancomer, the full financial portfolio of clients with the rest of the banks or other indicators associated with their wellbeing are not observed.

This paper is related to a broader empirical literature in economics that relies on big data for causal inference and policy evaluation (Einav and Levin 2014). As large-scale data sets become more available, economists working in fields such as labor, public finance, health, banking, or education are increasingly partnering with governments and firms to use these types of data in rigorous evaluation studies. Examples include Chetty, Friedman, and Rockoff (2014), who study the impact of teacher quality on students' outcomes in adulthood; the Tennessee STAR experiment (Chetty et al. 2011); evaluation of the Oregon Medicaid expansion (Taubman et al. 2014); and Seira, Elizondo, and Laguna-Müggenburg (2017), who use administrative data of a commercial bank as well as data from the Mexican credit bureau to evaluate different information disclosures in the credit card market. As access to this big data becomes more prevalent, the approach of combining experimental and non-experimental methods to overcome low compliance concerns is likely to be applicable in a wider range of settings.

The remainder of the paper is structured as follows: section 2 discusses the context of the experiment, details of the two interventions, and the experimental assignment; section 3 discusses the challenge of low take-up for statistical power; section 4 provides the treatment impacts using pure experimental methods, showing these to be uninformative; section 5 then discusses the study's methodology for combining experimental with non-experimental methods and the resulting estimates of treatment impact; and section 6 concludes.

## 2. The Intervention, Samples and Take-up

### Context

Mexico has experienced a remarkable increase in the penetration of financial services in the population in recent years. For instance, the number of credit cards owned was estimated at 17.3 million in June of 2016—a number 6.4 percent higher than that registered at June of 2015 (Banco de Mexico 2016). In a country with 76 million adults, this implies that there is one credit card for every four adults.<sup>5</sup> As of June 2016, the balance in credit card debts represented 39.5 percent of the total credit to personal consumption issued in the country (Banco de Mexico 2017).

This rapid growth in credit card usage in Mexico may clash with the amount of time that people have had to learn what an appropriate use of the cards may be. In Mexico, only 32 percent of adults were found to be financially literate, positioning it around the 85th place in a worldwide survey of 142 countries (Klapper, Lusardi, and van Oudheusden 2015). Equally worrying, Ponce, Seira, and Zamarripa (2017) illustrate how Mexican cardholders with more than one card typically leave money on the table by using the higher-rate card, and experience 31 percent higher cost than the minimum to finance their existing debt.

The experiment of this study was rolled out in several cities across Mexico by BBVA Bancomer, the study's partner bank. BBVA Bancomer is the largest player in the Mexican credit card market, having issued over 4.6 million credit cards of the 17.3 million that were in circulation as of June 2016. In terms of volume lent, BBVA Bancomer also has a dominant position. For instance, by June 2016, it concentrated 27 percent of the total lending in credit cards in Mexico (Banco de Mexico 2017).

Clients of the partner bank appear to be good candidates for improvement in their financial literacy skills. More than half are classified as either medium- or high-risk, and practically one in five is considered by BBVA Bancomer to be high-risk. BBVA Bancomer credit card clients also appear to use their cards

5 Based on INEGI's population Count 2015 and including those 20 years and older.

regularly and heavily: they spend on their credit cards about 6,600 Mexican pesos (MXP) per month, which is not far from the 7,365 MXP monthly salary in 2014.<sup>6</sup>

### The Interventions

This paper tackles the question of how effective financial literacy is in modifying financial behavior through two distinct treatments. One treatment arm is based on providing the typical approach to financial training: a classroom setting. Since this is the most common tool used by private and public institutions alike to promote financial education, understanding the potential effects and benefits remains an important question. The second treatment arm is based on personalized coaching sessions. This arm responds to the concern in Willis (2011) that due to the heterogeneity of households' circumstances and needs, effective financial education needs to be structured "in a one-on-one setting, with content personalized for each consumer." Due to the more personal approach, individuals may better internalize the information provided and receive more actionable recommendations based on their own situation and thus be more likely to change their financial behavior than in a situation where only generic advice was provided. However, such an approach can be costlier to provide and scale. The paper now describes each of the treatment arms in detail.

#### Financial Literacy Workshop

The first treatment arm is a financial education program named "Adelante con tu Futuro" (Go Ahead with Your Future), which was launched in Mexico by BBVA Bancomer in 2008. The program is part of a global social initiative funded by the BBVA group to promote financial knowledge and healthy financial management practices among the adult population. It has now expanded to other Latin-American countries where BBVA is present, such as Chile, Colombia, Uruguay, Paraguay, and Peru, and won an award in 2010 from the Inter-American Development Bank for innovativeness in fostering financial education.

"Adelante con tu Futuro" consists of workshops on a variety of modules including savings, retirement savings, credit card use, mortgages, and life insurance, as well as a series of courses for small and medium enterprises. Workshops are free and open to the public, and are offered both online and throughout selected BBVA Bancomer branches across the country. As of November 2016, over 4.8 million workshop sessions have provided training to about 1.2 million participants (as participants may take more than one module). Mexican regulators have been supportive of the workshops and help disseminate them among the population.<sup>7</sup> In 2017, BBVA Bancomer partnered with the national cash transfer program Prospera (formerly Progresa) to offer a version of the workshops to the program recipients.

The workshops are offered in person and online, and generally follow the same structure: two-hour interactive sessions, with material being provided in multimedia. A facilitator presents the material, some videos are shown, and each participant receives a notebook to conduct personal evaluations of their financial knowledge and behaviors, as well as a personal computer to work on. Participants get to take home the notebook, which also contains all the information reviewed, and a CD with exercises. Participants are evaluated at the end of the workshop and receive a certificate of completion. Importantly, none of the material, videos, or discussions that take place have any explicit or implicit references to BBVA products. Instead the material refers to generic financial products such as "credit cards," "your bank," "savings accounts," and so on. As such, the material does not encourage participants to prioritize use or payment of BBVA Bancomer products.

The workshop that is examined under this treatment arm is Credit Card Use and Financial Health, which consists of two parts. The first part of this two-hour course delves into the use of credit cards,

6 [Informe Anual del Observatorio de Salarios \(2016\)](#).

7 Mexico's consumer protection agency for financial related services (Condusef) disseminates them on a regular basis in its website and monthly magazine.

associated fees, and how to decipher a credit card statement. Hands-on exercises make participants go through the explanation of what the different elements of a credit card are (digits, expiration date, the security code in the back of the card), understand and differentiate between the payment period and the closing-date of purchases, and read all the elements of a bank credit card statement. Participants work on fictional case studies, where for instance a person has two credit cards (referred to as “credit card 1” and “credit card 2”) and is having trouble identifying which one is costlier and should be paid first. At the end of each case study, there is a group discussion where the facilitator explains the solutions and answers questions from participants.

The second part focuses on recommended credit card debt management practices. In this section, individuals learn what the credit score and credit history are, and their determinants, such as failure to pay old debts or frequently keeping high balances. Participants take a self-evaluation of their financial health, followed by a discussion of the steps that individuals can take to preserve and improve their credit management.<sup>8</sup> The steps are based on general management of financial products, such as “only use the credit cards that you can pay” or “if your financial health is critical, pay first the costliest credits or consolidate your debt.” Throughout the workshop, participants are reminded of rules of thumb labeled “golden rules” for good credit card behavior as a way to make the messages and advice concrete and easy to remember. These rules, presented in table S2.2 in the supplementary online appendix, emphasize the importance of paying on time and paying more than the minimum payment.

Invitations and attendance to the workshops took place from July through December 2016. Invitations were distributed through email and the BBVA Bancomer call center. Clients who were contacted were invited to participate in the face-to-face training in their city of residency. Clients from Mexico City, Puebla, Guadalajara, Morelia, Cuernavaca, Mérida, and Tijuana were included in this treatment arm. They were offered BBVA Bancomer rewards points for completing the course, which could be redeemed towards small rewards like a meal or merchandise.<sup>9</sup>

### Personalized Coaching

The second treatment arm consists of a series of personalized coaching sessions. The content was developed by BBVA Bancomer with the objective of “bringing information and tools to participants so that they have the capacity to make an adequate use of their credit card and to keep an excellent credit health.” This coaching entailed calling and scheduling a series of conversations with the participant to discuss her credit history, health, and behavior, and to help solve any issues or doubts she may have. The coaching was provided by highly trained *asesores* (financial advisers). Each of these *asesores* was assigned a group of clients whom they would call and invite to engage in these coaching sessions. If a client agreed to participate, the *asesores* would ask her about doubts and questions that she might have about her credit card and credit card use. *Asesores* were equipped to provide suggestions on how to improve the individual’s credit and help them pursue healthy financial behavior. The recommendations closely followed the contents and advice that the workshops had, such as paying more than the minimum to lessen the total interest payment overall, remembering the payment date, and budgeting in order to avoid unnecessary use of credit: “credit is not an extension of your salary.”

After the initial call and introduction, the *asesores* would discuss and agree on a follow-up call with the participant at a time that fit her schedule. The calls were aimed to be roughly two weeks apart, with a total of four calls with each participant. The calls were intended to be thematic and follow a specific progression: diagnostic, budget, credit, and credit health. Each call was planned to last about 10 minutes. The list of topics and structure of the discussion are presented in table S2.3 in the supplementary online appendix.

8 The detailed topics covered by the training are included in table S2.1 of the supplementary online appendix.

9 They were given up to 4,000 points, which are valued at approximately 300 pesos (US\$16).



The timeframe was the same as the workshops, with coaching sessions taking place between June and December 2016. While the call center was situated in Mexico City, participants of the coaching sessions could be residing in any of the nine cities that were part of the study. As with the workshops, participants were offered rewards points for completing the program.

It is worth noting that the delivery methods studied in the treatment arms are easily scalable and of relatively low cost from the point of view of a financial institution. Each participant in the workshop costs BBVA Bancomer 86 MXP or around 5 USD, while the cost per person coached was estimated at 131 MXP (7 USD). To put this cost in perspective, the annual fee for a set of BBVA Bancomer's credit cards is between 631 and 5,275 MXP. The average yearly profitability of BBVA Bancomer credit card clients is 1,056 MXP. Thus, expanding the provision of such interventions is not a high cost to pay. While credit card holders are already profitable to the bank, the expected effects of financial education through better (and increased) use of their credit cards may further strengthen the business case to continue the BBVA Bancomer financial literacy program.

### Outcomes and Measurement

The study obtained administrative data from BBVA Bancomer on 136,104 credit card clients from December 2014 to February 2017.<sup>10</sup> The data set contains rich information summarizing the monthly evolution of each client's credit card balance, payments, purchases, delays, profitability for the bank, and ownership of basic deposit accounts with the bank. The data set also includes the seniority of clients with the partner bank and their background characteristics such as age and gender.

The analysis is centered on a set of outcomes that are believed are more likely to be affected by the interventions, based on the material covered both in the workshops and coaching. More concretely, there are three rules that the workshops and coaching emphasize to help participants achieve a more responsible use of their credit cards and avoid extra fees and future over-indebtedness. The first rule is to cover at least the minimum payment required by the bank. The second one is to identify from the credit card statements the payment due date and make sure to pay before that date. A third piece of advice is to limit the use of credit to an amount that the client can comfortably pay later. Along these three rules, the importance of saving and better managing expenses is frequently discussed.

The study first tests whether participants of the interventions are more likely to follow these rules by analyzing key outcomes directly related to them. These outcomes are 1) the likelihood of paying more than the minimum payment; 2) the likelihood of delinquency, defined as paying after the payment due date; and 3) total credit card purchases. As the interventions highlight the importance of saving, the study also analyzes whether participants are more likely to own a basic deposit account after the interventions. Finally, the paper investigates how profitable these financial education interventions are for the partner bank. This is done by analyzing an indicator variable that equals 1 if a client at a given month is profitable for the bank and 0 otherwise.<sup>11</sup> Supplementary online appendix S3 also shows that the results are similar

10 To obtain this sample, BBVA Bancomer first restricted its universe of credit card clients to the sample of 502,283 clients that reside in areas with BBVA Bancomer branches that have Adelante con tu futuro classrooms. A second filter dropped 33,007 clients with no valid contact information and 28,935 clients that were younger than 18 and older than 70 years old, resulting in a sample size of 440,341 clients. This sample was then divided in seven arms. Three of these arms correspond to the intervention groups of the study (Control, Workshops, and Coaching groups), while the remaining four arms were interventions designed by BBVA Bancomer to test financial education tips sent via SMS messages. However, BBVA Bancomer decided to discontinue the planned SMS interventions, and these groups were then dropped from the sample.

11 The profitability variable corresponds to the difference between the revenue obtained from a client and the expenditures he or she generated for the bank. The revenue is measured as the interest income plus paid commissions and fees. The expenditures include operational costs, cost of capital, loan losses, and reserves. On one hand, profits to the bank may increase if clients spend more on their credit cards and thus the revenues from commissions and interest income rise.

when the analysis considers several other related outcomes concerning balances, number of products, payment delays, and bank income suggested by a referee.

The data make it possible to identify if the workshops and coaching affect the way participants manage their credit cards with BBVA Bancomer. However, the study cannot track whether management of other financial products with other financial institutions also changes as a response of the programs. One concern might be that the partner bank tailors the financial education programs to its own benefit: for instance, by encouraging participants to prioritize payment of BBVA Bancomer credit cards. The concern about the bank using the interventions to push its products is reduced by the fact that the material of the workshops has no reference to BBVA Bancomer or its products, and instead offers generic financial management tips. Moreover, according to the 2015 National Survey on Financial Inclusion, 74 percent of credit card holders in Mexico have only one credit card. This suggests that most individuals in the sample only manage the credit cards that are observed in the data.

### Samples and Random Assignment

Given that the study expects financial education interventions to have a greater impact among participants with riskier credit card management practices, the sample was stratified into six groups based on two risk measures of the clients. The first measure corresponds to the risk classification of each client. Every client is classified by the partner bank as low-risk, medium-risk, and high-risk. Since clients who struggle to cover the minimum payment are at a higher risk of facing credit card management issues in the future, the study produces a complementary risk measure that classifies clients according to how often their payment exceeds the minimum required by the bank. Clients are defined as “with frequent low payments” if more than one-third of the time they pay the required minimum or less.<sup>12</sup>

While the partner bank had no capacity constraints to deliver the financial education workshops, only 300 coaching interventions could be given.<sup>13</sup> Therefore, it was decided to restrict the coaching group to clients belonging to the stratum with the highest-risk clients (i.e., clients with frequent low payments and classified by the bank as high-risk). The highest-risk clients were randomly assigned into three groups: workshops, coaching, and the control group. For all other strata, clients were randomly assigned into either the workshop or the control group.

Clients in the coaching group were also randomly divided into three lists: the main list and two wait lists. Clients in the first wait list would only be contacted if there were still sessions available after coaching was offered to clients in the first list.<sup>14</sup> After contacting clients from the first two lists, all coaching sessions were exhausted. Therefore, 1,354 clients that were assigned to the third list were never contacted to participate in the intervention and were dropped from the sample.

To have at least one year of pre-intervention data to get accurate counterfactuals, the study also dropped from the sample 20,524 clients. These clients were new to the bank and only had six months of data before the interventions. Therefore, the final sample consists of 114,226 clients.<sup>15</sup>

In addition, if clients are more likely to pay on time, the costs of monitoring and recovering loans drop. On the other hand, profits may decline if better credit card management translates in lower interest payments and fees.

- 12 On average, 27 percent of individuals in the sample paid the minimum required payment or less in the pre-intervention period.
- 13 From the institution’s point of view, the intervention was thought as a pilot that could be scaled up based on the results and lessons learned.
- 14 Likewise, clients from the second list would only be contacted if there were still coaching sessions available after having invited all clients from the first wait list.
- 15 The partner bank has a list of all individuals that have participated in the workshops before. Participants in the intervention were cross-checked with this list (by first and last name) throughout the intervention period. None of the participants in the intervention, either in the control or treatment groups, were found in the lists of workshop participants. As the coaching sessions were a pilot established for this study, no client had previously taken them.



Table 1 presents the summary statistics of the sample, divided by the group of clients assigned to the workshops (panel A) and the set of clients assigned to coaching (panel B). Each panel is divided into four columns. The first column presents the characteristics of clients in the control group. The second and third columns of each panel show the characteristics of clients assigned to, and that effectively attended, each intervention. The fourth column presents the mean differential of the characteristics of clients assigned to the control group with clients taking up each intervention.

Clients in the sample are on average 46 years old, and about half of them are women. Most clients live in Mexico City and have been clients with the partner bank for about 12 years. In terms of their risk profile, 19 percent of clients are classified by the bank as being high-risk, 37 percent as medium-risk, and 44 percent as low-risk. Per the definition, 26 percent of clients struggle to pay more than the minimum required. That is, their payments do not exceed the minimum payment required in at least 4 of the 12 months that the study observes them before the intervention. Each month, clients tend to spend about 7,000 Mexican pesos on their credit cards. On average, 86 percent of clients pay more than the minimum payment required by the bank, and only 1 percent make their payments past the due date; 70 percent of clients own a deposit account with the partner bank. In terms of profitability, each month the bank obtains approximately 1,000 Mexican pesos for each client in the sample.

Currently, BBVA Bancomer offers 11 different credit cards in Mexico (online table S2.4). After the closing date (the last day of the monthly billing cycle), all clients have 20 days to pay at least the minimum payment (typically 20 percent of the balance) before incurring late fees. The terms vary across credit cards: APRs on BBVA Bancomer cards range from 18.6 to 115.6 percent, with those for the most common cards

**Table 1.** Baseline Characteristics by Treatment Assignment

	Panel A. Workshop sample				Panel B. Coaching sample			
	Assigned to control	Assigned to workshop	Attended workshop	Test of differential take-up	Assigned to control	Assigned to coaching	Attended coaching	Test of differential take-up
Time unvarying characteristics of clients								
Female	0.5	0.5	0.51	−0.01	0.48	0.45	0.35	0.12***
Age	46	46	46	0	46	46	46	0
From Mexico City	0.63	0.63	0.74	−0.11***	0.63	0.62	0.74	−0.12***
Years with the partner bank	12	12	13	−1*	12	11	12	0
Variables used for stratification								
High-risk client	0.19	0.19	0.16	0.03	1	1	1	0
Medium-risk client	0.37	0.38	0.38	0.00	0	0	0	0
Low-risk client	0.44	0.44	0.46	−0.02	0	0	0	0
With frequent low payments	0.26	0.26	0.20	0.06***	1	1	1	0
Time varying characteristics of clients (Average over pre-intervention period)								
Payment above minimum required	0.86	0.86	0.91	−0.06***	0.56	0.57	0.74	−0.18***
Pays past due date	0.01	0.01	0.00	0.0***	0.02	0.03	0.01	0.02***
Monthly credit card purchases (Mx \$)	6,594	6,729	9,428	−2,833***	5,258	5,491	9,901	−4,643***
Owens deposit account	0.69	0.69	0.75	−0.06***	0.88	0.88	0.86	0.02
Profitability to the bank (Mx \$)	1,056	1,073	1,142	−86	1,881	1,974	2,399	−518**

Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.

Note: Panels A and B present the characteristics of the sample assigned to the workshops and coaching groups, respectively. The first three columns of each panel present the summary statistics of clients assigned to the control group and to the workshops (panel A) or coaching (panel B), as well as clients that effectively attended the workshops (panel A) or coaching (panel B). The fourth column of each panel presents the mean difference between clients assigned to the control group and clients in workshops (panel A) or coaching (panel B) that effectively attended the intervention. The time varying characteristics of clients correspond to the average over 12 months prior to the interventions, except for the variable "profitability to the bank," which is only available 5 months prior to the interventions. \*, \*\*, \*\*\* indicate significance at the 10 percent, 5 percent, and 1 percent levels, respectively.

ranging from 68.2 percent to 91.6 percent. The fixed penalty for not paying on time is approximately 377 MXP (or 24 dollars) without counting the added interest. The annual fee of BBVA Bancomer credit cards is about 631 MXP (33 dollars), though it can be much higher for certain types of cards.

Take-Up

The marketing strategy that the partner bank implemented to invite clients to the interventions was as follows. Clients in the workshops and coaching groups were first sent an email whenever an email was available in the bank’s database. The email introduced the intervention that each client was assigned to and invited him or her to register. Clients who were not reached or registered at this stage were then contacted by the bank’s call center. Several contact attempts and follow-up calls were made at different days and times to maximize the possibility that the invitation was appropriately delivered. From a total of 114,226 clients, 36,946 were assigned to the control group, 73,654 were assigned to the workshop treatment arm, and 3,626 were assigned to the coaching treatment arm. As in many other settings analyzing financial behavior, the implementation of the study faced a major challenge regarding take-up rates. At the end, only 0.8 percent of the workshop treatment arm clients actually received the treatment, and 6.8 percent of the clients in the coaching treatment arm participated in the sessions (table 2). There are several reasons explaining this low take-up rate.

In the case of the workshop treatment arm, the resources needed to reach out to such a large number of clients were underestimated. Thus, during the implementation phase that lasted about six months, of the original group assigned to the workshop treatment, contact was attempted only to about 47.3 percent (34,818 clients). No record is available of the process used to decide which of those assigned to treatment to contact first. Next, despite repeated efforts to contact the individuals in this group, only 8,900 clients were effectively contacted. This means that over 25,000 clients did not pick up the phone during the outreach or that they answered and asked to be called later (without success). Thus, only a little over 12 percent of the group assigned to the workshops could be contacted and actually invited to participate in the treatment. From this group, 2,672 clients agreed to participate in the workshop, and a mere 583 attended and completed the workshop.

Similar challenges were faced in the roll-out of the coaching intervention. Due to the relatively low number of clients assigned to the treatment arm, the vast majority had at least one attempted contact (88.5 percent). From these, only about a third picked up the phone, and less than a sixth agreed to participate in the coaching sessions (14 percent of the original treatment group). Finally, 246 clients completed at least one session with the coach, translating to a take-up rate of 6.8 percent.<sup>16</sup> While these take-up

Table 2. The Take-up Challenge

	Panel A. Workshops			Panel B. Coaching		
	Number of clients	%	% (subject to being contacted)	Number of clients	%	% (subject to being contacted)
Assigned to treatment	73,654	100	—	3,626	100	—
Contact attempted	34,818	47.3	—	3,209	88.5	—
Able to be contacted	8,900	12.1	—	1,164	32.1	—
Agreed to participate	2,672	3.6	30.0	509	14.0	43.7
Actually received treatment	583	0.8	6.6	246	6.8	21.1

Source: Authors’ analysis based on the study implementation data provided by BBVA Bancomer.  
Note: The table summarizes the take-up rates of participants at different stages of the Workshops (panel A) and Coaching interventions (panel B). Columns (1) and (4) present the number of clients, columns (2) and (5) list the take-up rates; and columns (3) and (6) present the take-up rates conditional on having contacted participants.

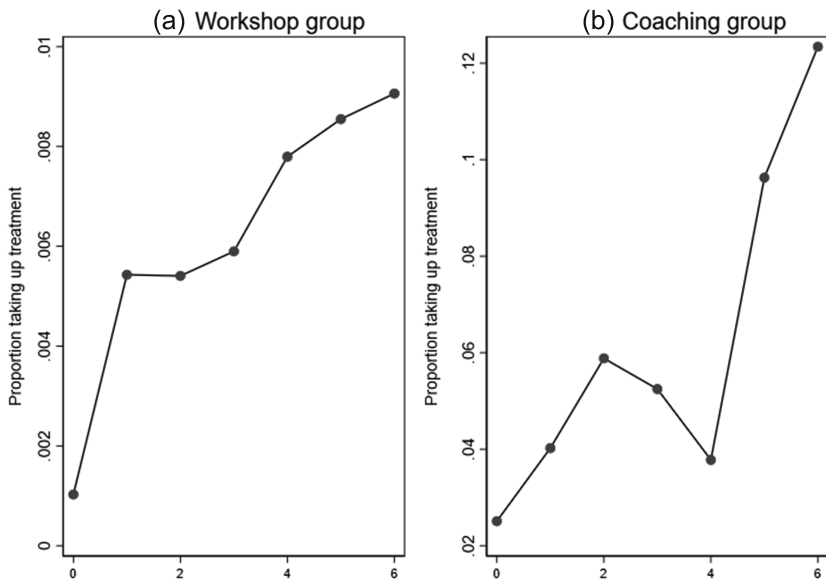
16 Conditional on having been contacted by our partner bank, 6.6 and 21.1 percent of clients assigned to the workshops and coaching group, ended up participating in the workshops and coaching, respectively (Table 2).

rates seem very low, as discussed in the introduction, they are unfortunately not unusual in the randomized controlled trial (RCT) universe, nor in the marketing reach-out campaigns of financial institutions. Anecdotal evidence from BBVA Bancomer deposits department puts the typical response rate of the bank’s marketing campaigns at 2 percent.

The challenge of low take-up rates can pose an even bigger problem if it is selective. For instance, it is easy to argue that bank clients will be less likely to answer a call from their bank if they are having trouble keeping their finances in order, are often late in paying their cards, or have typically large balances on their cards. Thus, when such clients get a call from the bank to be invited to take a training, a coaching session, or any other reason, they are less likely to answer the call in the first place. If good (i.e., more financially literate) clients self-select into participating in the treatment, while bad clients tend to self-select out of the treatment, a direct comparison of their outcomes with the control group will yield biased results. Financially literate clients are expected to have less to learn from more financial education, thus hinting that the workshops or coaching sessions may not affect individuals’ financial behavior significantly.

The study finds evidence of selective participation. Clients that end up taking the workshop or the coaching sessions appear to be in lower need of financial education than the average client. The more often an individual paid above the minimum payment in her card, the higher the likelihood she signed up for the workshop (fig. 1 panel a) or the coaching (fig. 1 panel b). For instance, an individual assigned to the coaching group who paid more than the minimum for six months is more than twice as likely to complete the coaching session than an individual who paid more than the minimum in three months only. Other characteristics also hint at positive selection among treatment takers. Clients who were contacted, accepted participating, and actually took the workshop (or coaching) are more likely to make payments above the minimum required, less likely to pay late, and more likely to also own a deposit account than

**Figure 1.** Take-Up Rates are Higher for Clients that Already have Better Repayment



Source: Authors’ analysis based on the study implementation data provided by BBVA Bancomer.

Note: Figures plot the take-up rate for the workshop intervention (panel a) and coaching intervention (panel b) according to the number of months out of six that the individual paid more than the minimum payment on their credit card in the six months prior to the intervention.

those who were assigned to the same treatment group but did not sign up and received the treatment (table 1). Within the workshop treatment group, the takers are also more likely to avoid making low payments on a regular basis.

The study now turns to the description of the methodological approaches used to estimate the effects of the treatment arms on financial behavior. The study first describes the purely experimental approach that is applied in most RCTs. Next, to overcome the low take-up rate problem, a combination of non-experimental methods is applied to get at a cleaner estimate of the impact of the workshops and coaching sessions. These approaches are described in detail below.

### 3. The Challenge of Low Take-Up for Statistical Power

Consider a simple comparison of treatment and control means in a randomized experiment of sample size  $N$ , which allocates a proportion  $P$  of subjects to the treatment, and  $1-P$  to the control, and for which unit  $i$  experiences a treatment effect  $\gamma_i$  if it receives the treatment. Let  $s_i$  be a dummy variable that takes value 1 when unit  $i$  takes up treatment, and 0 otherwise. The data generating process is then:

$$Y_i = a + \gamma_i s_i + u_i \quad (1)$$

The intent-to-treat effect can then be estimated in a regression of the form:

$$Y_i = a + b T_i + \varepsilon_i \quad (2)$$

Where  $T_i$  is a dummy variable denoting assignment to treatment, and the error  $\varepsilon$  is assumed i.i.d. with variance  $\sigma^2$ . Let  $\pi_T = E(s_i | T_i = 1)$  denote the take-up rate in the treatment group, and  $\pi_C = E(s_i | T_i = 0)$  denote the take-up rate in the control group.<sup>17</sup> Then the expected value of the ITT estimator  $\hat{b}_{ITT}$  is:

$$E(\hat{b}_{ITT}) = E(Y_i | T_i = 1) - E(Y_i | T_i = 0) \quad (3)$$

With standard error:

$$s.e.(\hat{b}_{ITT}) = \sqrt{\frac{\sigma^2}{N} \frac{1}{P(1-P)}} \quad (4)$$

Note that this standard error does not depend at all on the take-up rate.

#### Power with Homogeneous Treatment Effects for Those Treated

First consider the case when  $\gamma_i = \gamma$ , equation (3) simplifies to:

$$E(\hat{b}_{ITT}) = \gamma(\pi_T - \pi_C) \quad (5)$$

Since the take-up rate does not affect the standard error in equation (4), the only impact of a low take-up rate on statistical power will be through reducing the effect size in equation (5). The sample size  $N$  needed to detect effect size  $b = \gamma(\pi_T - \pi_C)$  at significance level  $\alpha$  and power  $\beta$  is then (e.g., Duflo, Glennerster, and Kremer 2008):<sup>18</sup>

$$N = \left[ \frac{(t_{1-\beta} + t_{\alpha/2})}{\gamma(\pi_T - \pi_C)} \right]^2 \frac{\sigma^2}{P(1-P)} \quad (6)$$

It is seen that the sample size required is proportional to the inverse of the difference in take-up rates squared  $(\pi_T - \pi_C)^{-2}$ . The consequence is that low take-up rates dramatically increase the sample size

17 Note that when analyzing an experiment with a given take-up rate, it is assumed that which units end up taking up the treatment is stochastic, but that the take-up rates are not. Thus  $\pi_T$  and  $\pi_C$  are treated as constants.

18 With more rounds of data and the use of difference-in-differences or Ancova estimation, the variance term becomes more complicated, but the influence of the take-up rates remains the same (McKenzie 2012).

required to detect the impact of training: if take-up is 10 percent, 100 times the sample is needed than with full take-up; if it is 5 percent, 400 times the sample is needed; and if it is the 0.5 percent that is common in responses to bank direct mail promotions, 40,000 times the sample is needed. This makes it extremely challenging for experimental methods to detect the impact of interventions when take-up rates are very low (or more generally, even if take-up rates are higher but the control group also takes up the intervention so that the treatment-control take-up gap is small).

Researchers will also typically instrument receipt of treatment with random assignment to the instrument, to get the IV instrument  $\hat{b}_{IV}$ . This is equivalent to the Wald estimator:

$$\hat{b}_{IV} = \frac{\hat{b}_{ITT}}{(\pi_T - \pi_C)} \quad (7)$$

And so  $E(\hat{b}_{IV}) = \gamma$ , which does not change with the take-up rate. However, the variance of the IV estimator will be proportional to  $(\pi_T - \pi_C)^{-2}$ , and the power of the IV estimator to detect an effect size of  $\gamma$  will be exactly the same as that of the ITT estimator to detect an effect size of  $b = \gamma(\pi_T - \pi_C)$ .

### Power with Heterogeneous Treatment Effects

Now consider the more general case, where  $\gamma_i$  is heterogeneous, and comes from a distribution with mean  $\mu_\gamma$  and variance  $\sigma_\gamma^2$ . Since the focus is on low take-up experiments, assume that none of the control group take-up the intervention ( $\pi_C = 0$ ). Then the expected value of the ITT estimator in equation (3) becomes:

$$E(\hat{b}_{ITT}) = E(s_i \gamma_i | T_i = 1) = \pi_T \mu_\gamma + \rho(s_i, \gamma_i) \sigma_\gamma \sqrt{\pi_T (1 - \pi_T)} \quad (8)$$

This consists of two terms. The first term shows how lowering the take-up rate reduces the ITT by reducing the average effect size. The second term is the selective take-up effect. This is the covariance between the take-up rate and treatment heterogeneity, which can be expressed as the product of the correlation between these two variables  $\rho(s_i, \gamma_i)$ , the standard deviation of the heterogeneity, and the standard deviation of the take-up variable  $s_i$ . The IV estimator  $\hat{b}_{IV}$  then has expected value:

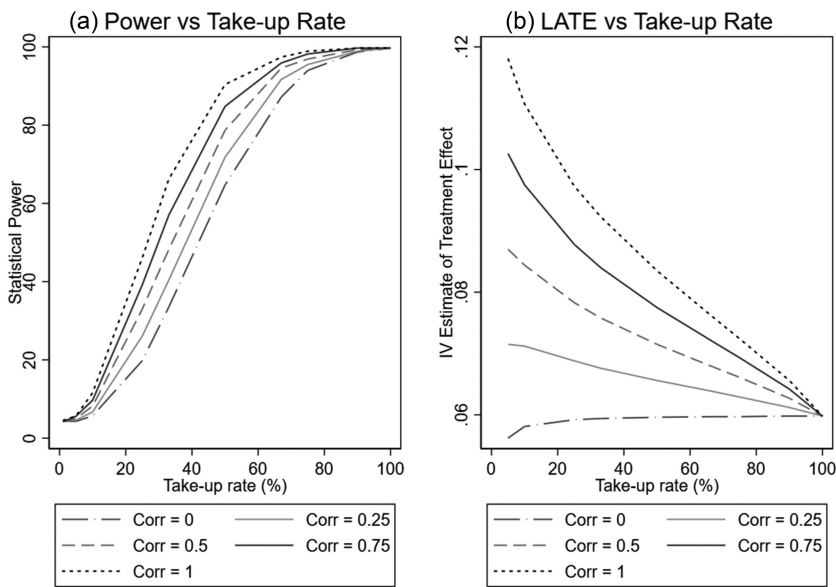
$$E(\hat{b}_{IV}) = E(\hat{b}_{ITT}) / \pi_T = \mu_\gamma + \rho(s_i, \gamma_i) \sigma_\gamma \frac{\sqrt{\pi_T (1 - \pi_T)}}{\pi_T} \quad (9)$$

A first implication of equations (8) and (9) is that if there is treatment effect heterogeneity, but this is uncorrelated with take-up, then the expected value of the IV estimator does not depend on the take-up rate, and power is the same as in the homogeneous treatment effect case (since the standard error does not vary with treatment effect heterogeneity).

In the more general case, individuals who expect to gain more from the treatment may be more likely to take up the intervention, which is referred to by [Heckman, Urzua, and Vytalil \(2006\)](#) as essential heterogeneity. In this case, the correlation  $\rho(s_i, \gamma_i)$  will be positive, and the stronger this correlation and the larger the treatment effect heterogeneity, the greater will this second term be. However, note that  $\sqrt{\pi_T(1 - \pi_T)}$  is maximized when  $\pi_T = 0.5$ . Holding  $\mu_\gamma$  constant, the effect of adding treatment effect heterogeneity (increasing  $\sigma_\gamma$ ), or making take-up more positively selective on this heterogeneity (increasing  $\rho(s_i, \gamma_i)$ ) on statistical power will then be greatest for take-up rates near 50 percent, since this is when the second term in equation (7) is largest. In contrast, when take-up rates get very low, then most people have no treatment effect, and while the IV estimator of the treatment effect for those who do take-up treatment will be largest at low take-up rates, the impact of this on power compared to a situation of no treatment heterogeneity will be low.

This is illustrated in [fig. 2](#), which uses simulations based on this study’s outcome of paying more than the minimum payment, and estimated effect size from the coaching treatment (see supplementary online appendix S1 for details). In this example, when treatment heterogeneity is uncorrelated with take-up, power drops from 99.7 percent with 100 percent take-up to 64.7 percent with 50 percent take-up, and only 4.3 percent with 5 percent take-up. If, the study were instead in the extreme case where individuals perfectly order themselves into taking up the intervention by what their treatment effect would be (correlation of 1), power would only fall to 90.4 percent with 50 percent take-up, a huge gain. However, it would still drop to 5.5 percent with 5 percent take-up. The right panel shows these power gains come from the IV estimate growing with selective take-up. Supplementary online appendix S1 shows that a similar story holds when increasing the amount of treatment heterogeneity, while holding the correlation constant.<sup>19</sup> Nevertheless, for the very low take-up rates common in financial interventions, power will still be low even with treatment heterogeneity that is highly correlated with take-up. Moreover, in practice, many individuals may not be able to identify their treatment effects even *after* experiencing a program ([McKenzie 2018](#)), and take-up is driven by many other factors such as which individuals the program officials are able to reach, transport distances, and many other factors, so that it is unclear that those who stand to benefit most from a program will be the most likely to take it up.

**Figure 2.** With Treatment Effect Heterogeneity, Power Falls Less Steeply with Take-Up the More Positively Correlated Take-up is with Individual Treatment Effects



Source: Authors’ analysis based on the study implementation data provided by BBVA Bancomer.  
Note: Illustration from simulated treatment effect where control mean is 0.69, control standard deviation is 0.46, sample size is 5,000 units divided equally between treatment and control, and individual treatment effects are drawn from a random normal distribution with mean 0.06 and standard deviation 0.03. Corr denotes the correlation between the order in which units take up treatment and their treatment effect. (a) Power vs. Take-up Rate, (b) LATE vs. Take-up Rate. See appendix S1 for more details.

19 Appendix S1 also highlights that when treatment heterogeneity is very large and take-up is strongly correlated with treatment, it can be possible for power to actually increase at first when take-up falls from 100 percent—because those with large negative effects no longer take-up treatment and thus do not drag down the average.



### Standard Approaches to Overcoming Low Power

One obvious solution to this concern is to try and increase take-up rates.<sup>20</sup> There are several approaches possible for doing this. A first approach is to first screen individuals on their interest in attending the course, and then only randomize amongst those who are most interested. There are two downsides to such an approach from the point of view of a financial institution. First, this screening can be costly (e.g., it might require calling all the clients, or mailing them all fliers to send back in), and may still not generate large enough a sample to have sufficient statistical power. This approach was tried with credit card clients in Mexico by Bruhn, Lara Ibarra, and McKenzie (2014), but led to very small samples when restricting to those who replied to a mail promotion or Facebook advertisement. A second disadvantage from the financial institution's viewpoint, is that it then requires denying (or delaying) the provision of training to clients who have explicitly indicated strong interest.

A second approach is then to try to boost take-up among those offered the program. This can be done by offering additional incentives to take up the program, and by costly efforts to call participants multiple times and encourage and remind them to participate. Such an approach was also tried by Bruhn, Lara Ibarra, and McKenzie (2014), offering a variety of payments, testimonials, and transportation assistance. While this can increase statistical power, it also comes with two downsides. The first is cost: financial institutions may be willing to offer financial literacy services, but not to incur a lot of additional costs in getting people to take these services. Second, if scale-up of such a program will not involve offering these same incentives, then the sample for whom the treatment effect is estimated will not be representative of the treatment effect for individuals who will take up the program when it is offered at scale with no take-up incentives.

As a consequence of these issues, the interest here is in identifying methods that can be used to measure impact in a large-scale randomized experiment that aims to identify a treatment effect for the types of people who will take up the program when it is offered at scale and low cost. This then means that take-up will be low, and so the question is how to learn about the program in such a context.

## 4. Experimental Treatment Impacts

The offer of a financial education workshop or of coaching was randomly assigned, and so comparing post-treatment outcomes for the treatment group to the control group gives an unbiased estimate of the intention-to-treat (ITT) effect, which is the effect of being offered the program. Consider outcome  $Y_{i,t}$  measured for client  $i$  in period  $t$ . McKenzie (2012) shows that with multiple rounds of follow-up data, maximum power comes from estimating an average effect  $\gamma$  over the entire nine-month post-intervention period  $t = 1, 2, \dots, 9$  via the following Ancova specification:

$$Y_{i,t} = \gamma \text{TreatmentOffered}_i + \sum_{s=1}^9 \delta_s 1(s = t) + \theta \bar{Y}_{i,PRE} + \sum_a \lambda_a 1(i \in a) + \varepsilon_{i,t} \quad (10)$$

Where  $\bar{Y}_{i,PRE}$  is the mean of the outcome over the pretreatment periods,  $\lambda_a$  are strata fixed effects,  $\delta_s$  are time period fixed effects, and the standard errors  $\varepsilon_{i,t}$  are clustered at the client level.

Under the assumption that the invitation to financial education or coaching has no impact on outcomes for those who do not take up the treatment the local-average treatment effect (LATE) can also be estimated by replacing *TreatmentOffered* with *TreatmentReceived* in equation (10), and then instrumenting the receipt of treatment with its randomly assigned offer. This identifies the local-average treatment effect

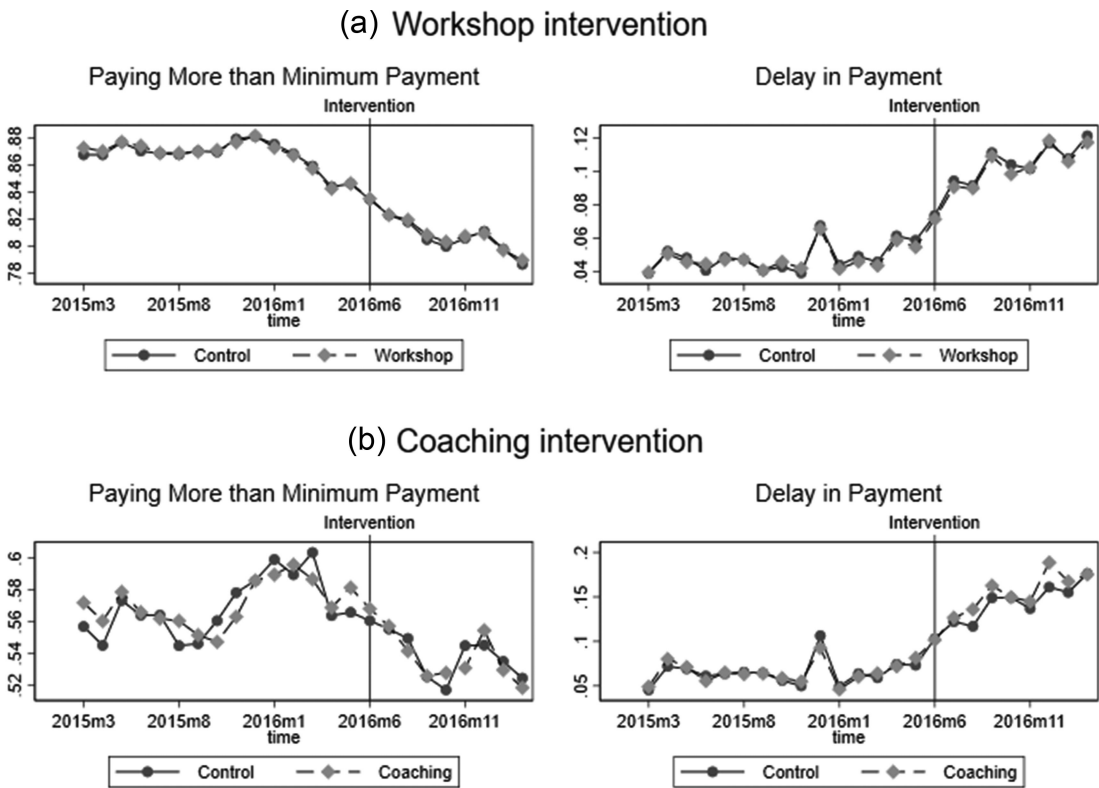
20 Another approach that governments can take is to make financial education mandatory in high schools, thereby ensuring high take-up. Cole, Paulson, and Shastry (2016) and Bruhn et al. (2016) provide evaluations of such programs in the United States and Brazil respectively. Conceivably a financial institution could require financial education before granting a credit card, but this would likely reduce demand for the cards, and is not observed in practice.

(LATE), which is the effect of receiving training or coaching when offered it, and not otherwise. If no one in the control group takes up the treatment, then this also gives the treatment-effect-on-the-treated (TOT). This is the case for the coaching intervention, but it is possible that a few individuals in the control group for the training intervention may have attended a workshop without being invited.

Figure 3 plots the trajectory of two key outcomes—paying more than the minimum payment, and having a delay in payment—over time by treatment status. The top two figures show this for the sample assigned to workshops, and the bottom two figures for the sample assigned to coaching. In both cases it is seen that the treatment and control groups track each other very closely over time before the intervention (as would be expected by randomization with a large sample), and continue to track each other closely after the intervention. With such low take-up, the average for the treatment group as a whole is dominated by the behavior of those who do not receive the treatment.

The ITT estimates in equation (10) of being offered the financial education workshop or the coaching are all small in magnitude, and very close to zero (table 3). That is, the offer of treatment has a very small, and insignificant impact on financial behavior. Underneath each ITT, the study then reports the LATE/TOT and a 95 percent confidence interval around it. It is seen that the confidence intervals are very wide for the impact of actually taking up either treatment; as a result, the experiment is not very informative about the impact of these interventions. For example, the 95 percent confidence interval for the impact of coaching on whether or not the client pays more than the minimum payment ranges from −25 percentage points

Figure 3. Evolution over Time of Fraction of Clients Paying above the Required Minimum and Fraction of Clients with Delay in their Payment by Experimental Treatment Status



Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.  
Note: Figures show outcome trajectories by assignment to treatment status, comparing those randomly assigned to the workshop intervention to their assigned control group (top panels) and those randomly assigned to the coaching intervention to their assigned control group (bottom panel).

Table 3. Experimental Estimates of Treatment Effects

	Share of debt paid by due date	Client classified as not in good standing	Delay in payment	Pays more than minimum	Has basic deposit account with bank	Profitable client (bai)	Profitable client (nibt)	Log of monthly balance	Log of monthly spending
Panel A: Impact of workshops									
ITT	-0.002 (0.004)	-0.001 (0.003)	0.000 (0.001)	0.001 (0.002)	-0.001 (0.001)	0.003** (0.001)	0.004*** (0.002)	0.020** (0.010)	0.012 (0.015)
LATE	-0.254	-0.069	0.053	0.149	-0.148	0.424**	0.529***	2.494**	1.440
95% confidence interval	[-1.2, 0.7]	[-0.7, 0.6]	[-0.2, 0.3]	[-0.2, 0.5]	[-0.5, 0.2]	[0.1, 0.8]	[0.1, 0.9]	[0.0, 5.0]	[-2.1, 5.0]
Sample size	799,816	248,411	865,572	798,314	858,891	660,084	660,084	842,944	865,572
Mean	0.526	0.110	0.054	0.806	0.698	0.812	0.786	9.297	5.382
Panel B: Impact of coaching									
ITT	-0.009 (0.012)	0.010 (0.008)	0.010 (0.006)	-0.002 (0.008)	-0.002 (0.006)	-0.002 (0.006)	-0.000 (0.006)	-0.014 (0.045)	0.029 (0.066)
LATE	-0.122	0.146	0.140	-0.033	-0.025	-0.027	-0.000	-0.206	0.411
95% confidence interval	[-0.4, 0.2]	[-0.1, 0.4]	[-0.0, 0.3]	[-0.2, 0.2]	[-0.2, 0.1]	[-0.2, 0.2]	[-0.2, 0.2]	[-1.5, 1.1]	[-1.4, 2.2]
Sample Size	43,100	30,777	47,632	43,017	48,058	36,736	36,736	46,940	47,632
Mean	0.271	0.146	0.101	0.537	0.851	0.821	0.805	9.474	3.657

Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.

Note: Robust standard errors in parentheses, clustered at the client level. \*, \*\*, \*\*\* denote significance at the 10, 5, and 1 percent levels respectively. Estimation is by Ancova, and includes mean of outcome over baseline periods, time period fixed effects, and strata fixed effects.

to +18 percentage points. The control mean is 54 percent, so this is equivalent to almost halving the percent paying on time, or increasing it by one-third.<sup>21</sup>

This is where standard analysis using experimental methods would stop. The study would conclude that there is no significant impact of either intervention on the main outcomes, but that there is insufficient power to rule out a wide range of positive and negative impacts. The study therefore turns to combining non-experimental methods with the experiment to obtain more informative results.

## 5. Combining Experimental and Non-Experimental Methods to Measure Impact for Those Who Actually Take Up Treatment

### Empirical Approach

This study's solution is to combine elements from the idea of principal stratification analysis ([Frangakis and Rubin 2002](#)), with propensity score matching and difference-in-differences. The idea behind principal stratification is to first divide individuals into strata based on their compliance status with the assigned treatment. In this context, the fact that none of the control group are able to receive coaching (or receive training in the specified period), means it is possible to classify individuals as either "Compliers," who will take up treatment when offered it, and not take it up otherwise; and "Noncompliers," who will not take up treatment when offered. The object of interest is then the complier average causal effect (CACE), which is the effect of the coaching or training for those who will take it up when offered.

Since the study observes take-up status in the treatment group, the identity of the compliers in the treatment group is known, and the proportion of compliers and of noncompliers in the control group.<sup>22</sup> However, it is not known which individuals in the control group would comply had they been offered treatment. The LATE experimental estimation detailed above makes no assumptions about which control individuals will be compliers, but only uses their proportion—an approach [Page et al. \(2015\)](#) refer to as the moment-based approach to estimating the CACE. In contrast, if researchers are willing to make further assumptions about which individuals in the control group will be compliers (which [Page et al. \(2015\)](#) refer to as the model-based approach), then power can increase, but at the cost of additional assumptions, and bias if these assumptions are violated. For example, if take-up could be perfectly predicted based on observable characteristics, then one could model take-up among the control group, and simply compare these individuals to the compliers in the treatment group.

However, in practice it can be difficult to precisely identify which control individuals will be compliers purely based on modeling selection on observables. This study's solution is to use the richness of the financial data available on credit card clients to combine experimental methods with propensity score matching and difference-in-differences. This only makes it necessary to identify a group of individuals who would have similar trends to the compliers, rather than having to identify exactly which individuals in the control group would comply. Propensity score matching is used to match individuals in the treatment group who took up the treatment (compliers) to similar individuals in the control group, and then difference-in-differences on this matched sample to estimate the impact of attending the workshop or receiving coaching. That is, the study estimates the following equation for the matched sample of compliers

- 21 This study also investigates if the ITT effects of workshops and coaching are stronger for the sample of clients with higher predicted take-up rates and find no statistically significant differences for these clients (see supplementary online appendix S4).
- 22 This illustrates where applying this study's approach to cases where the treatment-control gap in take-up is low, but where some of the control group also take the treatment would be more difficult, since then two other groups need to be identified: the always-takers (who take the treatment regardless of group), and the defiers (who do the opposite of their treatment assignment).

in the treatment group and matched controls:

$$Y_{i,t} = \gamma \text{TreatmentReceived}_{i,t} + \beta \text{Complier}_i + \sum_{s=-18}^9 \delta_s 1(s = t) + \sum_a \lambda_a 1(i \in a) + \varepsilon_{i,t} \quad (11)$$

Where *TreatmentReceived*<sub>*i,t*</sub> takes on value 1 for the post-intervention periods in which the compliers in the treatment group have received their treatment, and 0 otherwise; *Complier*<sub>*i*</sub> is an indicator of whether individual *i* is a complier in the treatment group (as opposed to a matched control); and the time fixed effects are now included for up to 18 months pre-treatment, as well as 9 months post-treatment. The standard errors are again clustered at the client level.

There are several concerns that typically apply when applying propensity score matching. The first is a concern of omitted variables: individuals who look similar in terms of baseline observable variables might differ in terms of unobserved characteristics that also matter for client outcomes. A particular concern here is that of dynamic selection. For example, people might be more willing to engage in financial education if they suddenly find themselves struggling with their credit card, whereas those who have been experiencing problems for a while may be less likely to participate. Matching on current behavior only would not be able to distinguish between these two types. Secondly, a critique underlying all matching studies is to explain why, if these two groups are so similar, only one group ended up taking the intervention. The rich data and experiment help in addressing both concerns. The study has up to 18 months of pre-intervention financial data for these clients and so can match not only on current financial behavior, but on the monthly trajectory of this behavior over many months. This helps alleviate concerns about dynamic selection. Moreover, by only matching to individuals in the control group (and not those in the treatment group who did not take up treatment), the study has a plausible reason why some individuals do not take up treatment—they were not invited to under the random invitations.

Difference-in-differences further makes it possible to difference out any time-invariant unobservable differences between the two groups. Thus, if those who participate in training or coaching always tend to be better re-payers than those who do not, this can be differenced out. The underlying assumption for difference-in-difference analysis is that of a common trend, so that the two groups would follow the same time paths as each other in the absence of an intervention. This assumption is more credible if the individuals are more similar to begin with (which is where matching helps), and if it is seen that the two groups have the same dynamics prior to the intervention. The monthly administrative data make it possible to not only test whether the two groups follow similar linear trends prior to the intervention, but also to test whether they follow the same nonlinear trend. To test this, the study estimates over the pre-intervention period:

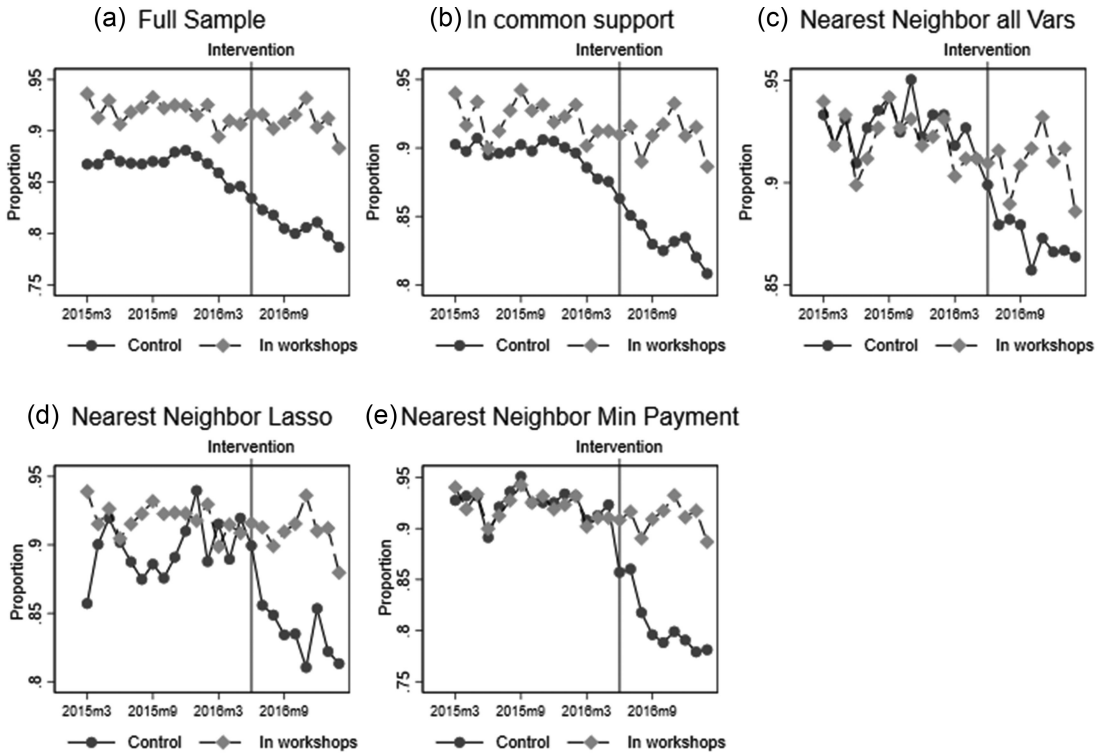
$$Y_{i,t} = \sum_{s=-18}^{-1} \beta_s 1(s = t) \text{Complier}_i + \sum_{s=-18}^{-1} \delta_s 1(s = t) + \sum_a \lambda_a 1(i \in a) + \varepsilon_{i,t} \quad (12)$$

And test that all the  $\beta$ 's are jointly zero.

In moving away from the pure experiment, there is no one universally agreed control group. Several different plausible ways of choosing this control group are examined. The study then views the resulting estimates as more credible if these different methods give similar results, even though they end up choosing different individuals from the control group to match to those who actually take up treatment in the treatment group.

The study begins by estimating the difference-in-differences estimator using the full control group. If there is self-selection into treatment, those receiving training or coaching will differ in levels, and potentially trends, from this full control group. A first step towards refining the control group to more comparable individuals is to restrict the analysis to individuals in the common support of the propensity score. For this approach, the propensity score is estimated as a function of gender, and pretreatment monthly

**Figure 4.** Illustration of the Five Different Approaches to Forming a Counterfactual, for the Workshop Treatment and Outcome of Paying More than the Minimum Payment



Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.

Note: (a) Full sample compares means for all individuals receiving coaching to the full control group. (b) In common support shows means for the sample within the common support of a propensity score estimated using the full history of all five outcomes plus a control for gender. (c) Nearest neighbor all vars then shows means after single nearest-neighbor matching without replacement within this common support using the propensity score estimated with all variables. (d) Nearest-neighbor lasso using lasso regression to pick the variables used to form the propensity score, then matches to the nearest neighbor with this propensity score. (e) Nearest neighbor min payment forms a propensity score only on the history of paying more than the minimum payment, and forms the nearest neighbor from this score.

levels of all five outcomes. This involves matching on 73 variables in total, and eliminates 38 percent of the control group and 22 percent of the treatment group for coaching. The study then goes further by choosing the nearest neighbor within this common support for each client who received treatment. Using all the outcomes simultaneously to form these matches has the advantage of making clients similar on average in terms of existing financial behavior, but, because it is attempting to match on so many variables, this approach may not match especially well on any particular single outcome. The analysis therefore also considers two alternatives to forming the propensity score and then matching on the nearest neighbor. The first is to use Lasso to choose a parsimonious set of variables to match on. This chooses 8 of the 73 variables to use in forming the match. The second, and this study's most preferred approach, is to match just on the month-by-month pre-intervention data for an outcome at a time. This last method ensures that the control individuals look as similar as possible on levels and dynamics as those receiving treatment, but does mean, in contrast to the other approaches, that different controls are used for each outcome.

Figure 4 illustrates how the five different approaches define counterfactuals for the coaching treatment and outcome of paying more than the minimum payment owed. The top left panel compares the full control group to those receiving treatment. It is seen that the group which received coaching starts from a much higher mean level than the control group, reflecting positive selection into training in terms of



pre-existing credit behavior. The trends seem broadly similar pre-intervention, suggesting difference-in-differences may be able to control for this selection. This difference in baseline means becomes smaller, but is still there, when the study conditions on being in the common support. In contrast, all three nearest-neighbor approaches look much more similar on baseline levels, and appear to match reasonably well on baseline trends. These different nearest-neighbor approaches do select different individuals from the control group: only two clients are selected by all three methods, so multiple plausible counterfactuals are being formed.

Tables S2.5 and S2.6 in the supplementary online appendix show how these three nearest-neighbor matches achieve samples from the control group that are much more comparable on baseline observables to those who took up treatment than is the case for the full control sample. In both tables, the first column shows baseline means for those who *took-up* treatment. The study then follow Imbens and Rubin (2015) in considering the normalized difference  $(\bar{X}_T - \bar{X}_C)/\sqrt{(\bar{s}_T^2 + \bar{s}_C^2)/2}$  as a measure of balance, where  $\bar{X}_j$  and  $\bar{s}_j^2$  are the sample mean and variance of the variable for those receiving treatment ( $j = T$ ) and the comparison subsample from the control group ( $j = C$ ) respectively. These normalized differences provide a scale-invariant measure of the difference in means, with differences less than 0.2 standard deviations typically considered to indicate balance. It is seen that normalized differences exceed this level for pre-intervention averages in the key outcomes when using the full control sample, or the sample within the common support, but are all less than this when using any of the three nearest-neighbor methods. As a result, it is not possible to reject equality of means of the financial outcomes averaged over all pre-intervention periods when using nearest neighbor.

### Impacts of Workshops

Table 4 presents the difference-in-difference estimates of the impact of workshops on the five outcomes of interest (panels A through E). Each column presents the results of one of the five approaches that are used to form a control group. For each outcome and approach, the study tests whether the treated and control groups followed common linear and nonlinear trends in the pre-intervention period. The  $p$ -values of the tests are included in the table. Figure 4 shows a graphic representation of the different approaches. The figure shows the trajectory of the outcome of paying more than the minimum payment for the clients that participated in the workshops and the clients assigned to the different control groups.

The last column of the table presents the preferred specification, the nearest-neighbor approach that matches on the monthly pre-intervention data for a specific outcome of interest. As fig. 5 shows, this approach makes it possible to generate a control group that tracks very closely the evolution of each outcome for the treated group in the pre-intervention period. After the intervention, the mean outcomes of the clients that participated in the workshops begin to separate from the mean outcomes of the control group. While the outcomes of clients in the control group deteriorate over time (i.e., lower fraction of clients paying more than the minimum required and increased likelihood of delayed payment), the outcomes of clients who took the workshops remained stable.

The results from table 4 show that these differences are statistically significant for all outcomes, except for bank profitability. The  $p$ -values of the common trends tests suggest that the study cannot reject that the outcomes of clients who took the workshop and the control group formed by the preferred approach followed common linear and nonlinear trends before the interventions. The economic impact of the estimates is also robust across the different matching approaches. The results of the preferred specification suggest that participating in the workshop results in an 11 percentage point increase in the likelihood of paying more than the minimum payment, a 3.4 percentage point reduction in the likelihood of delaying payment, 63.7 percent higher monthly spending on the credit card, and a 2.7 percentage point increase in the likelihood of owning a deposit account with the partner bank.

**Table 4.** Estimated Treatment Effects for Those Who Did Receive Workshops

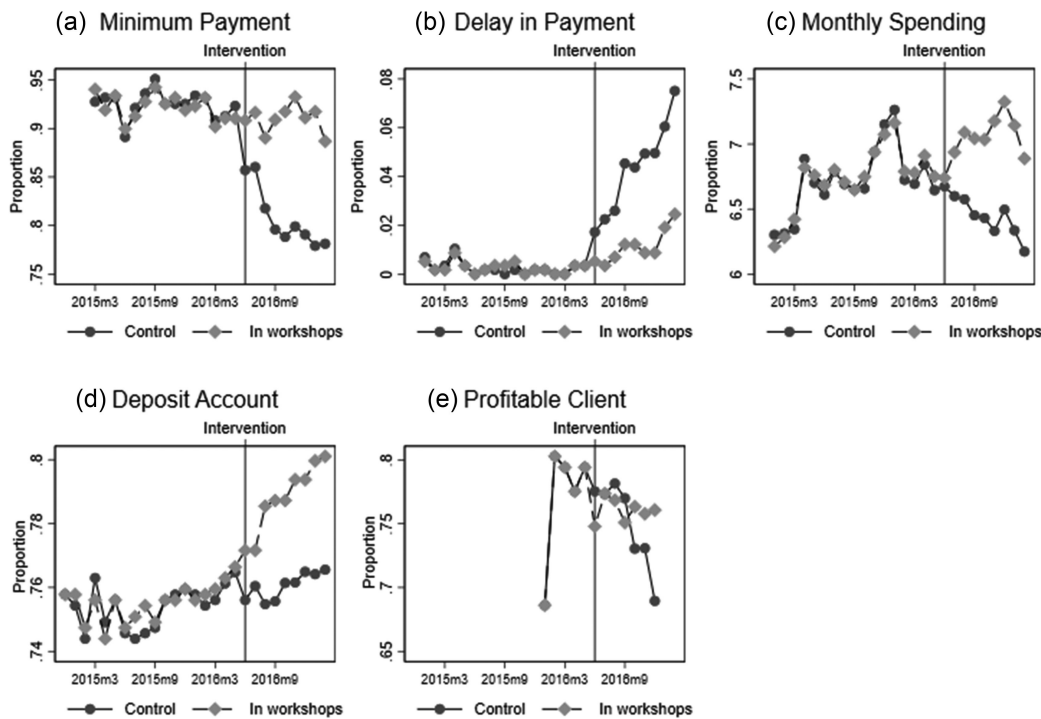
	Full control sample	In common support	Nearest-neighbor matching		
			On all variables	Using lasso	On outcome
Panel A: Pay more than minimum payment					
Receive workshop*post-intervention	0.050*** (0.007)	0.051*** (0.008)	0.043*** (0.012)	0.053*** (0.012)	0.107*** (0.015)
Sample size	826,664	647,267	22,161	25,773	22,225
Mean	0.806	0.831	0.871	0.834	0.802
p-values for test common linear pre-trend	0.864	0.592	0.567	0.0241	0.599
p-values for test common nonlinear pre-trend	0.380	0.475	0.981	7.69e-05	0.989
Panel B: Delay in payment					
Receive workshop*post-intervention	-0.037*** (0.003)	-0.036*** (0.003)	-0.020*** (0.007)	-0.038*** (0.008)	-0.034*** (0.008)
Sample Size	967,442	707,101	24,121	28,389	29,998
Mean	0.0539	0.0489	0.0332	0.0546	0.0464
p-values for test common linear pre-trend	0.225	0.126	0.262	0.829	0.711
p-values for test common nonlinear pre-trend	0	0.450	0.811	0.0164	0.991
Panel C: Log monthly spending on card					
Receive workshop*post-intervention	0.455*** (0.089)	0.408*** (0.096)	0.417*** (0.139)	0.454*** (0.125)	0.637*** (0.126)
Sample Size	967,442	707,101	24,121	28,389	29,997
Mean	5.382	5.963	6.845	6.478	6.425
p-values for test common linear pre-trend	0.224	0.636	0.973	0.621	0.680
p-values for test common nonlinear pre-trend	0.253	0.762	0.970	0.00904	0.998
Panel D: Has a deposit account					
Receive workshop*post-intervention	0.028*** (0.009)	0.029*** (0.010)	0.044*** (0.016)	0.028** (0.014)	0.027** (0.013)
Sample size	1,003,455	732,321	25,061	29,418	31,079
Mean	0.698	0.687	0.747	0.673	0.761
p-values for test common linear pre-trend	0.370	0.615	0.766	0.393	0.922
p-values for test common nonlinear pre-trend	0.226	0.812	0.776	0.0830	0.995
Panel E: Profitable client for the bank					
Receive workshop*post-intervention	0.024** (0.011)	0.023** (0.011)	-0.005 (0.016)	0.018 (0.015)	0.021 (0.016)
Sample size	449,122	326,999	11,153	13,122	13,948
Mean	0.786	0.811	0.778	0.819	0.746
p-values for test common linear pre-trend	0.083	0.368	0.871	0.840	1.000
p-values for test common nonlinear pre-trend	0.173	0.0687	0.978	0.327	1.000

Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.

Note: Robust standard errors in parentheses, clustered at the client level. \*, \*\*, \*\*\* denote significance at the 10, 5, and 1 percent levels only. The five columns show estimated treatment impacts of taking part in the workshops treatment, using different control groups. Column (1) uses all clients randomly assigned to the control; column (2) uses those within the common support when matching on all pre-intervention variables; column (3) uses single nearest-neighbor matching within this common support; column (4) uses single nearest-neighbor matching with the common support when using lasso to select variables for propensity score, and then nearest-neighbor matching within the common support.

Finally, the study investigates if the impact of workshops differs for riskier clients. As table S2.7 in the supplementary online appendix shows, in the months following the interventions, the subset of high-risk clients in the control group shows worse performance across all outcomes. Thus, to test if workshops were particularly helpful for riskier clients, the study includes additional variables in equation (11) that interact the risk category of a client with the *TreatmentReceived* variable. No statistically significant evidence of heterogeneous effects across clients of different risk categories are found (table S2.8).

**Figure 5.** Trajectories of Financial Outcomes of those Receiving Workshops Compared to Nearest-Neighbor Matched Control Group



Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.  
Note: Propensity score matching used to construct a nearest-neighbor matched control sample using outcome-specific pre-intervention variables. Fewer months pre- and post-intervention are available for the outcome of being a profitable client for the bank. (a) Minimum payment, (b) Delay in payment, (c) Monthly spending, (d) Deposit account, (e) Profitable client.

Impacts of Coaching

Table 5 provides the difference-in-difference results for the five different approaches to forming a control group, along with tests of whether the two groups follow a common linear trend, and a common nonlinear trend, before the intervention. Figure 6 shows that after the time of the intervention, the control group is becoming progressively more likely to not pay more than the minimum payment, to delay in their credit card payments, to no longer have a deposit account with the bank, to cut back on spending, and are becoming less profitable clients for the bank. The coaching treatment is halting these trends from occurring, so that those who receive coaching appear more similar to their pre-intervention levels.

Table 5 shows that these impacts are statistically significant after matching for all but having a deposit account, are reasonably robust in magnitude to different plausible ways of defining this matched control group, and that the study cannot reject the possibility that the matched control groups display parallel linear or nonlinear trends pre-intervention. Using the preferred specification in the last column (which matches fig. 6), it is found that receiving coaching results in a 5.9 percentage point increase in the likelihood of paying more than the minimum payment, a 2.6 percentage point reduction in the likelihood of delaying payment, 51.9 percent higher monthly spending on the credit card, and a 7.8 percentage point increase in the likelihood that the client is profitable for the bank.

Discussion

This combination of nearest-neighbor matching, difference-in-differences, and the random assignment makes it possible to find a subset of clients within the full experimental control group who look similar on baseline observables to those who take up the interventions, and who also follow similar pre-intervention

trends. Using this strategy, the study detects treatment effects of the interventions that could not be detected using experimental methods alone. It is worth reiterating how little power experimental methods will have to detect treatment effects of the size that are found, given the low take-up levels. For example, the study's power to detect the 5.9 percentage point improvement in the likelihood of paying more than the minimum that is seen for the coaching treatment, given that take-up is 6.8 percent, is

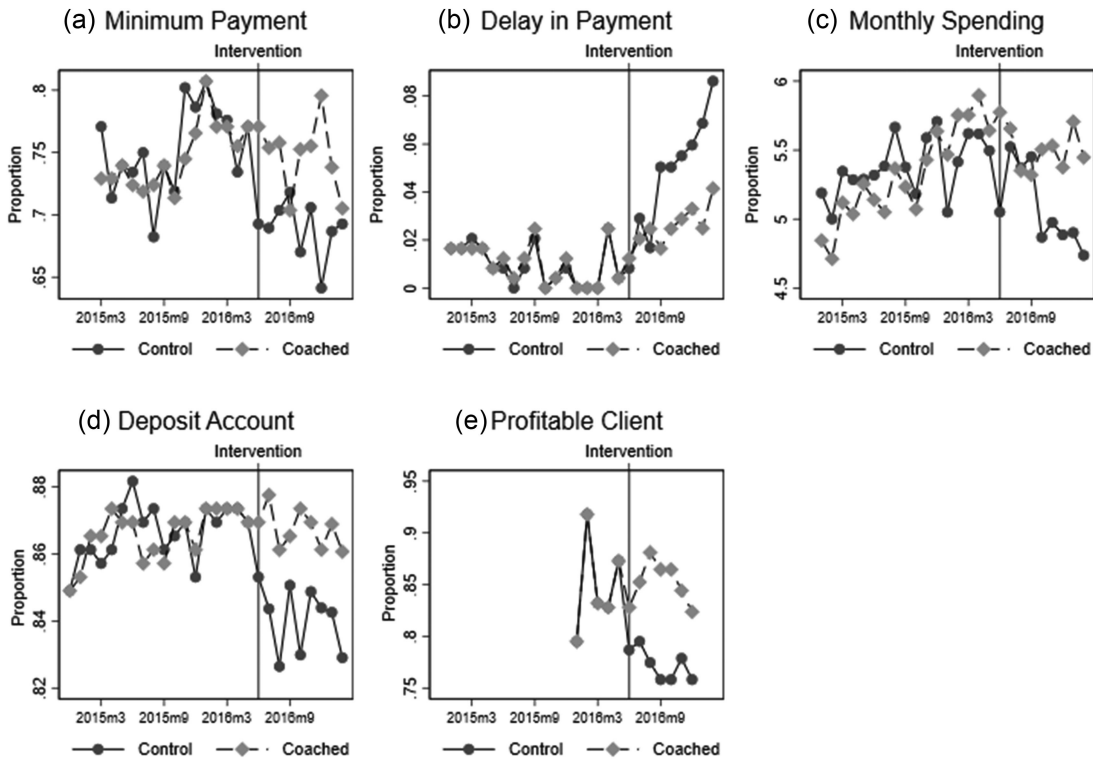
**Table 5.** Estimated Treatment Effects for Those Who Did Receive Coaching

	Full control sample (1)	In common support (2)	Nearest-neighbor matching		
			On all variables (3)	Using lasso (4)	On outcome (5)
Panel A: Pay more than minimum payment					
Receive coaching*post-intervention	0.036** (0.017)	0.055*** (0.019)	0.068** (0.027)	0.040 (0.026)	0.059** (0.028)
Sample size	59,043	41,743	9,104	10,380	9,151
Mean	0.537	0.582	0.687	0.661	0.689
p-values for test common linear pre-trend	0.074	0.417	0.107	0.474	0.800
p-values for test common nonlinear pre-trend	0.283	0.985	0.641	0.358	0.811
Panel B: Delay in payment					
Receive coaching*post-intervention	−0.058*** (0.009)	−0.061*** (0.010)	−0.064*** (0.019)	−0.042*** (0.016)	−0.026* (0.013)
Sample size	70,498	45,485	9,894	11,390	12,560
Mean	0.537	0.582	0.687	0.661	0.689
p-values for test common linear pre-trend	0.074	0.417	0.107	0.474	0.800
p-values for test common nonlinear pre-trend	0.283	0.985	0.641	0.358	0.811
Panel C: Log monthly spending on card					
Receive coaching*post-intervention	0.396** (0.178)	0.270 (0.195)	0.585** (0.267)	0.448* (0.250)	0.418* (0.247)
Sample size	70,498	45,485	9,894	11,390	12,510
Mean	3.657	4.559	5.207	5.562	5.092
p-values for test common linear pre-trend	0.006	0.306	0.895	0.701	0.0700
p-values for test common nonlinear pre-trend	0.031	0.853	0.961	0.798	0.907
Panel D: Has a deposit account					
Receive coaching*post-intervention	0.032*** (0.012)	0.030** (0.014)	0.033 (0.021)	0.019 (0.019)	0.028 (0.019)
Sample size	73,805	47,261	10,270	11,835	13,179
Mean	0.851	0.838	0.815	0.862	0.839
p-values for test common linear pre-trend	0.476	0.771	0.502	0.584	0.938
p-values for test common nonlinear pre-trend	0.035	0.886	0.890	0.0631	0.860
Panel E: Profitable client for the bank					
Receive coaching*post-intervention	0.065*** (0.015)	0.059*** (0.015)	0.061** (0.024)	0.033* (0.020)	0.078*** (0.022)
Sample size	32,987	21,064	4,572	5,268	5,856
Mean	0.805	0.834	0.825	0.854	0.770
p-values for test common linear pre-trend	0.786	0.621	0.753	0.557	1.000
p-values for test common nonlinear pre-trend	0.079	0.432	0.962	0.0749	1.000

Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.

Note: Robust standard errors in parentheses, clustered at the client level. \*, \*\*, \*\*\* denote significance at the 10, 5, and 1 percent levels only. The five columns show estimated treatment impacts of taking part in the coaching treatment, using different control groups. Column (1) uses all clients randomly assigned to the control; column (2) uses those within the common support when matching on all pre-intervention variables; column (3) uses single nearest-neighbor matching within this common support; column (4) uses single nearest-neighbor matching with the common support when using lasso to select variables for propensity score, and then nearest-neighbor matching within the common support.

**Figure 6.** Trajectories of Financial Outcomes of Those Receiving Coaching Compared to Nearest-Neighbor Matched Control Group



Source: Authors' analysis based on the study implementation data provided by BBVA Bancomer.

Note: Propensity score matching used to construct a nearest-neighbor matched control sample using outcome-specific pre-intervention variables. Fewer months pre- and post-intervention are available for the outcome of being a profitable client for the bank. (a) Minimum payment, (b) Delay in payment, (c) Monthly spending, (d) Deposit account, (e) Profitable client.

only 17.9 percent.<sup>23</sup> Note that the estimated treatment effects are all within the (very wide) confidence intervals seen for the LATE in [table 3](#).

It is important to note that these treatment effects are for the set of clients who will take up the interventions when invited. It has been seen there is positive selection into participation, so that individuals who have the worst initial financial behavior in terms of late payments, not paying more than the minimum required, and so forth are less likely to participate. The treatment effect may be larger for these individuals if they could be induced to participate, since they have more room for improvement, or potentially smaller if such individuals are less likely to implement the changes suggested in the workshops.

Finally, the cost per client of providing these programs was 131 MXP (7 USD) per person coached, and 86 MXP (5 USD) per person participating in the workshop. Using the impact on paying on time, this equates to a cost of  $7 \text{ MXP} / 0.026 = 269 \text{ MXP}$  per additional client induced by coaching to pay on time. If this were the only impact, this would appear an expensive way for banks to get clients to pay in a timelier manner. However, it is seen that the training and coaching get clients to pay their accounts on time and pay more of their bills, but do not get them to cut back on spending. In fact, perhaps because they are not experiencing as many payment problems, they spend more on their cards. The result is that this training does increase the likelihood these clients remain profitable for the bank.

23 This uses the autocorrelation of approximately 0.4 in the data, and the following command in Stata: `sampsi 0.69 0.694012, n1(2504) n2(3626) sd1(0.46) pre(14) post(9) r01(0.4) r1(0.4) r0(0.4)`.

## 6. Conclusions

Reliable estimation of treatment effects in impact evaluations relies heavily on the implementation efforts of teams to get individuals assigned to a treatment to take up treatment (and on ensuring those assigned to the control group do not receive treatment). In settings where individuals have little incentive to participate and those that do tend to be self-selected, the identification of the program effects through experimental methods is challenging. Unfortunately, and despite their recent popularity, financial education programs constitute a perfect example of these issues.

This study takes advantage of the richness of administrative data from a financial institution to implement an estimation approach that overcomes the low-take-up problem in RCTs. The availability of monthly administrative data over a two-year period for a large pool of individuals assigned to the control group allows for a clearer estimation of the impacts of the workshops and coaching sessions. By selecting a group of individuals within the control group that is statistically comparable in their financial behavior (previous to the treatment) to that of the effectively treated group, this approach improves the empirical evaluation of the treatment in several ways. The approach improves upon the simple application of experimental methods. The experimental method estimate gets “diluted” as it compares a large pool of individuals assigned to the treatment group where only a handful were effectively treated, to a large pool of individuals assigned to the control group. By using several rounds of administrative data, this study’s approach also allows the verification of the parallel trends assumption required for the application of non-experimental approaches.

The combination of experimental and non-experimental methods along with rich administrative data present a new avenue for empirical applications of impact evaluations when take-up is low. Examples of widespread reach-out efforts with small response rates include credit card offers in the United States where response rates were estimated to reach 0.2 percent in 2006,<sup>24</sup> and the banking/finance industry in the UK, where the overall click rate (i.e., the percent of clicks that are made among all the emails sent) of email marketing campaigns was estimated at 0.48 percent in 2016.<sup>25</sup>

## References

- Banco de México. 2016. Indicadores Básicos de Tarjetas de Crédito. Datos a junio de 2016.
- Banco de México. 2017. Indicadores Básicos de Tarjetas de Crédito. Datos a junio de 2017.
- Banerjee, A., E. Duflo, R. Glennerster, and C. Kinnan. 2015. “The Miracle of Microfinance? Evidence from a Randomized Evaluation.” *American Economic Journal: Applied Economics* 7 (1): 22–53.
- Brown, A., and K. Gartner. 2007. “Early Intervention and Credit Cardholders.” *Center for Financial Services Innovation*, Chicago, IL. [http://cfsinnovation.com/system/files/imported/managed\\_documents/earlyintervention.pdf](http://cfsinnovation.com/system/files/imported/managed_documents/earlyintervention.pdf).
- Bruhn, M., and D. McKenzie. 2014. “Entry Regulation and the Formalization of Microenterprises in Developing Countries.” *World Bank Research Observer* 29 (2): 186–201.
- Bruhn, M., L. de Souza Leão, A. Legovini, R. Marchetti, and B. Zia. 2016. “The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil.” *American Economic Journal: Applied Economics* 8 (4): 256–95.
- Bruhn, M., G. Lara Ibarra, and D. McKenzie. 2014. “The Minimal Impact of a Large-Scale Financial Education Program in Mexico City.” *Journal of Development Economics* 108: 184–9.
- Chemin, M. 2018. “Informal Groups and Health Insurance Take-up Evidence from a Field Experiment.” *World Development* 101 (C): 54–72.
- Chetty, R., J.-N. Friedman, N. Hilger, E. Saez, D.-W. Schanzenbach, and D. Yagan. 2011. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project STAR.” *Quarterly Journal of Economics* 126 (4): 1593–660.

24 FDIC (2007).

25 Sign-up.to (2017).



- Chetty, R., J.-N. Friedman, and J.-E. Rockoff. 2014. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.
- Chong, A., D. Karlan, and M. Valdivia. 2010. "Using Radio and Video as a Means for Financial Education in Peru." *Innovations for Poverty Action*. <http://www.povertyactionlab.org/evaluation/using-radio-and-video-means-financial-education-peru>.
- Cole, S., A. Paulson, and G.-K. Shastri. 2016. "High School and Financial Outcomes: The Impact of Mandated Personal Finance and Mathematics Courses." *Journal of Human Resources* 51 (3): 656–98.
- Duflo, E., R. Glennerster, and M. Kremer. 2008. "Using Randomization in Development Economics: A Toolkit." In *Handbook of Development Economics*, edited by P. Schultz and T.-J. Strauss, Vol. 4, 3895–962. North Holland, Amsterdam, NH.
- Einav, L., and J. Levin. 2014. "Economics in the Age of Big Data." *Science* 346 (6210): 1243089.
- Federal Deposit Insurance Corporation. 2007. "Credit Card Activities Manual." Accessed October 15, 2017. [https://www.fdic.gov/regulations/examinations/credit\\_card/ch5.html](https://www.fdic.gov/regulations/examinations/credit_card/ch5.html).
- Fernandes, D., J.-G. Lynch, Jr., and R.-G. Netemeyer. 2014. "Financial Literacy, Financial Education, and Downstream Financial Behaviors." *Management Science* 60 (8): 1861–83.
- Frangakis, C., and D. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics* 58 (1): 21–29.
- Giné, X., R. Townsend, and J. Vickrey. 2008. "Patterns of Rainfall Insurance Participation in Rural India." *World Bank Economic Review* 22 (3): 539–66.
- Goldstein, M. 2011. "A Disappointment with Encouragement." *Development Impact (blog)*. April 5. <http://blogs.worldbank.org/impactevaluations/a-disappointment-with-encouragement>.
- Grodzicki, D. 2015. "Competition and Customer Acquisition in the U.S. Credit Card Market." [https://editorialexpress.com/cgi-bin/conference/download.cgi?db\\_name=IIOC2015&paper\\_id=308](https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIOC2015&paper_id=308).
- Heckman, J., S. Urzua, and E. Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88 (3): 389–432.
- Imbens, G.W., and D.B. Rubin. 2015. *Causal Inference in Statistics, Social, and Biomedical Sciences*. New York: Cambridge University Press.
- Informe Anual del Observatorio de Salarios. 2016. "Los salarios y la desigualdad en México." Universidad Iberoamericana Puebla. Accessed October 21, 2016. <http://redsalarios.org/app/uploads/594c4919032ff.pdf>.
- Karlan, D., J. Morduch, and S. Mullainathan. 2010. "Take-up: Why Microfinance Take-Up Rates are Low & Why it Matters." Financial Access Initiative Framing Note. <http://www.arabic.microfinancegateway.org/sites/default/files/mfg-en-paper-take-up-why-microfinance-take-up-rates-are-low-why-it-matters-jun-2010.pdf>.
- Klapper, L., A. Lusardi, and P. van Oudheusden. 2015. "Financial Literacy Around the World: Insights from the Standard & Poor's Rating Services Global Financial Literacy Survey." [http://gflec.org/wp-content/uploads/2015/11/3313-Finlit\\_Report\\_FINAL-5.11.16.pdf?x28148](http://gflec.org/wp-content/uploads/2015/11/3313-Finlit_Report_FINAL-5.11.16.pdf?x28148).
- Levinsohn, J., N. Rankin, G. Roberts, and V. Schöer. 2013. "Wage Subsidies to Address Youth Unemployment in South Africa: Evidence from a Randomised Control Trial." Working paper, Stellenbosch University.
- Lusardi, A., and O. Mitchell. 2014. "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature* 52 (1): 5–44.
- Lusardi, A., and P. Tufano. 2015. "Debt Literacy, Financial Experiences, and Overindebtedness." *Journal of Pension Economics and Finance* 14 (4): 332–68.
- McKenzie, D. 2018. "Can Business Owners Form Accurate Counterfactuals? Eliciting Treatment and Control Beliefs about Their Outcomes in the Alternative Treatment Status." *Journal of Business & Economic Statistics* 36 (4): 714–22.
- McKenzie, D. 2012. "Beyond Baseline and Follow-Up: The Case for More T in Experiments." *Journal of Development Economics* 99 (2): 210–21.
- Miller, M., J. Reichelstein, C. Salas, and B. Zia. 2015. "Can You Help Someone Become Financially Capable? A Meta-Analysis of the Literature." *World Bank Research Observer* 30 (2): 220–46.
- Mottola, G. 2013. "In Our Best Interest: Women, Financial Literacy, and Credit Card Behavior." *Numeracy* 6 (2): 4.
- Page, L., A. Feller, T. Grindal, L. Miratrix, and M.-A. Somers. 2015. "Principal Stratification: A Tool for Understanding Variation in Program Effects Across Endogenous Subgroups." *American Journal of Evaluation* 36 (4): 514–31.
- Ponce, A., E. Seira, and G. Zamarripa. 2017. "Borrowing on the Wrong Credit Card? Evidence from Mexico" *American Economic Review* 107 (4): 1335–61.

- Seira, E., A. Elizondo, and E. Laguna-Müggenburg. 2017. "Are Information Disclosures Effective? Evidence from the Credit Card Market." *American Economic Journal: Economic Policy* 9 (1): 277–307.
- Sign-up.to. 2017. "Email Marketing Benchmark Report." Accessed October 15, 2017. <https://www.signupto.com/email-marketing-benchmarks/email-benchmark-2017/>.
- Taubman, S.-L., H.-L. Allen, B.-J. Wright, K. Baicker, and A.-N. Finkelstein. 2014. "Medicaid Increases Emergency-Department Use: Evidence from Oregon's Health Insurance Experiment." *Science* 343 (6168): 263–8.
- Willis, L. 2011. "The Financial Education Fallacy." *American Economic Review Papers & Proceedings* 101 (3): 429–34.