

Learning the Impact of Financial Education When Take-Up Is Low

Gabriel Lara Ibarra

David McKenzie

Claudia Ruiz Ortega



WORLD BANK GROUP

Development Research Group

Finance and Private Sector Development Team

&

Poverty and Equity Global Practice Group

November 2017

Abstract

Financial education programs are increasingly offered by governments, nonprofits, and financial institutions. However, voluntary participation rates in such programs are often very low, posing 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 nonexperimental 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.

This paper is a product of the Finance and Private Sector Development Team, Development Research Group and the Poverty and Equity Global Practice Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at <http://econ.worldbank.org>. The authors may be contacted at glaraibarra@worldbank.org, dmckenzie@worldbank.org and cruizortega@worldbank.org.

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the International Bank for Reconstruction and Development/World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

Learning the Impact of Financial Education When Take-Up Is Low[#]

Gabriel Lara Ibarra
World Bank

David McKenzie
World Bank

Claudia Ruiz Ortega
World Bank

JEL Classification Codes: D14, G21, G28, O12.

Keywords: financial literacy; credit-card behavior; low take-up

[#] We thank Adolfo Albo, Juan Luis Ordaz, David Cervantes and the BBVA Bancomer Financial Education Team for their collaboration on the experiment and for providing us with the anonymized customer data used in this paper. We would also like to thank the participants at the EduFin Summit 2017 in Mexico City for comments and suggestions. The authors of the paper received no funding for this work from BBVA Bancomer, and maintained independence in their 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.

1. Introduction

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, non-profits, and financial institutions have launched efforts to provide financial education in many countries around the world (Fernandes et al, 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 et al. (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, to receive only 119 responses. In Peru, Chong et al. (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 not unique to financial education, but are 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).¹

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. This can be one reason why many experimental evaluations of financial literacy training struggle to find a significant effect (Fernandes et al, 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 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.

¹ Take-up rates can be substantially higher in underdeveloped economies in which consumers have few alternatives for the product or service. For example, Bursztyn et al. (2017) report 21 percent take-up for a platinum credit card offer in Indonesia. Nevertheless, Karlan et al. (2010) 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. Crépon et al. (2015) use a pilot phase to then over-sample households with a higher ex-ante propensity to borrow, but find that take-up is not that much higher in this group than in the average population, highlighting the difficulty of predicting in advance who will take-up borrowing.

The key question this paper attempts to address is whether we can still obtain reliable measures of the impact of financial education when take-up rates are too low to enable estimation by experimental methods alone. Our 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, but with very wide confidence intervals for the estimated treatment effect on the treated (the impact of actually receiving training or coaching).

Nevertheless, we argue 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. We show 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, we use the rich time-series data we have 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 pre-treatment. 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 we can show the two groups we select 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.

We find that both attending the financial education workshops, and receiving personalized coaching, do lead to changes in financial outcomes for participants. Attending the workshop or receiving coaching results in a 6 to 11 percentage point increase in the likelihood of paying more than the minimum payment, a 3 percentage point lower likelihood of not paying by the payment due date, and a 3 percentage point increase in the likelihood of also having a deposit account at the bank. 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.

The remainder of the paper is structured as follows: Section 2 discusses the context of the experiment, details of the two interventions, the experimental assignment, and the low take-up; Section 3 provides the treatment impacts using pure experimental methods, showing these to be

uninformative; Section 4 then discusses our methodology for combining experimental with non-experimental methods and the resulting estimates of treatment impact; and Section 5 concludes.

2. The Intervention, Samples, and Take-up

2.1 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 27.5 million in December of 2016 – a number 12.4 percent higher than that registered at the end of 2015.² In a country with 76 million adults, this implies that there is one credit card for every three adults.³ As of June 2016, the balance in credit card debts represented 39.4 percent of the total credit to personal consumption issued in the country.⁴

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 et al. 2015). Equally worrying, Ponce et al. (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 our study was rolled out in several cities across Mexico by BBVA Bancomer, our partner bank. BBVA Bancomer has a dominant position in the credit card market in Mexico. In 2016, BBVA Bancomer and Citigroup Banamex held more than half the credit card market in Mexico and around the same share of the total balance in such credits.⁵

Clients of our 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 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 2014.⁶

2.2 The Interventions

In this paper, we tackle the question of how effective is financial literacy in improving 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

² <https://www.publimetro.com.mx/mx/economia/2017/05/03/numero-tarjetas-credito-se-dispara-12-4-condusef.html>

³ Based on INEGI's population Count 2015 and including those 20 years and older.

⁴ <http://www.banxico.org.mx/sistema-financiero/publicaciones/reporte-de-tasas-de-interes-efectivas-de-tarjetas-/%7B3A787547-BAAA-A0F2-4D1A-151E96D32321%7D.pdf>

⁵ Based on comparable market of credit cards (Banxico, 2016).

⁶ Informe Anual del Observatorio de Salarios 2016 “Los salarios y la desigualdad en México” IBERO/EQUIDE

structured “in a one-on-one setting, with content personalized for each consumer”. Due to the more personal approach, individuals may internalize better 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. We now describe each of the treatment arms in detail.

Financial Literacy course

The first treatment arm is a financial literacy training course that is made freely available by our partner bank at a large-scale throughout Mexico and online. BBVA Bancomer launched the financial literacy courses in 2008, winning an award for innovativeness in fostering financial education in 2011. The courses offered include savings, retirement savings, credit card use, mortgages, life insurance, as well as a series of workshops for small and medium enterprises. As of November 2016, over 4.8 million workshop sessions have provided training to about 1.2 million participants.

The courses are offered in person and online. All courses follow generally the same structure: two-hour-long 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 that 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.

Under this treatment arm, we focus on the provision of the course covering modules on *Credit Card Use and Financial Health*. The first module of this two-hour course delves into the use of credit cards, associated fees, and how to decipher a credit card statement. Hands-on exercises make participants go through the explanation of what the different parts 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. The second module focuses on good credit card debt management practices. In this module, individuals learn about the credit score and credit history and their determinants, including failure to pay old debts or keeping high balances frequently. Participants are asked to go through a self-evaluation of their financial health, and based on this assessment a discussion is held on the steps individuals can take to preserve and improve their credit management.⁷ Throughout the modules, 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 are presented in Appendix 3b, and emphasize the importance of paying on time, and paying more than the minimum payment.

Invitations and attendance to the workshops were conducted from July through December 2016. Invitations were done through email and BBVA Bancomer call center. Given the scope of the study, participants were contacted and invited to participate in the face-to-face training in their city

⁷ The detailed topics covered by the training are included in Appendix 3a.

of residency. Clients from Mexico City, Puebla, Guadalajara, Morelia, Cuernavaca, Mérida and Tijuana were included in the workshop 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.⁸

Personalized coaching

A second treatment arm was based on 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 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 individuals who they would call and invite to engage in these coaching sessions. If an individual agreed to participate, the asesores would ask them about the doubts and questions they may have about their 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 followed closely the contents and advice that the workshops had, such as: payment more than the minimum to lessen the total interest payment overall, remembering the payment date, and budgeting as to avoid unnecessary use of credit: “the 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 fitted 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. Overall, participants were expected to be engaged in this treatment in a span of two calendar months. The list of topics and guide for the discussion are presented in appendix 3c.

As with the workshops, coaching sessions took 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 our treatment arms are easily scalable and of relatively low cost from the point of view of a financial institution. Each participant in the workshop cost 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 should not be necessarily seen as a high toll to pay. While credit card holders are already profitable to the bank, the expected effects of financial education through better (and

⁸ They were given up to 4,000 points, which are valued at approximately 300 pesos (US\$16).

increased) use of their credit cards may further strengthen the business case to continue BBVA Bancomer financial literacy program.

2.3 Outcomes and Measurement

We obtained administrative data from BBVA Bancomer on 136,104 credit card clients from December 2014 to February 2017. The data set contains a rich set of 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 our partner bank and their background characteristics such as age and gender.

We center our analysis on a set of outcomes that we believe 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 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.

We first test whether participants of the interventions are more likely to follow these rules by analyzing key outcomes directly related to them. These outcomes are: i) the likelihood of paying more than the minimum payment; ii) the likelihood of paying past the payment due date; and iii) total credit card purchases. As the interventions highlight the importance of saving, we also analyze whether participants are more likely to own a basic deposit account after the interventions. Finally, we investigate how profitable these financial education interventions are for our partner bank. We do this by analyzing a variable that measures the profitability of each client for the bank at every month.^{9,10}

2.4 Samples and Random Assignment

Given that we expect financial education interventions to have a greater impact among participants with riskier credit card management practices, we stratified the sample 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 our 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, we produce a complementary risk measure that classifies clients

⁹ Our 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.

¹⁰ 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. 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.

according to how often their payment exceeds the minimum required by the bank. We define clients as “with frequent low payments” if more than one-third of the time they pay the required minimum or less.¹¹

While our partner bank had no capacity constraints to deliver the financial education workshops, only 300 coaching interventions could be given.¹² Therefore, we 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.¹³ 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, we 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, our final sample consists of 114,226 clients.

Table 1 presents the summary statistics of our sample, divided by the group of clients assigned to the workshops (Panel A) and the set of clients assigned to coaching (Panel B). We divide each panel in 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 our 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 our 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 we observe 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 our partner bank. In terms of profitability, each month the bank obtains approximately 1,000 Mexican pesos for each client in the sample.

¹¹ On average, 27 percent of individuals in the sample paid the minimum required payment or less in the pre-intervention period.

¹² 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.

¹³ 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.

Currently, BBVA Bancomer offers eleven different credit cards in Mexico (Appendix 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% of the balance) before incurring late fees. Even though 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 ranging from 68.2% to 91.6%. 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.

2.5 Take-up

As in many other settings analyzing financial behavior, the implementation of the study faced a major challenge regarding take-up rates. 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 3,636 were assigned to the coaching treatment arm. 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). 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% 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%). From these, only about a third picked up the phone and less than a sixth agreed to participate in the coaching sessions (14% 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%. While these take-up rates seem dire, as discussed in the introduction, they are unfortunately not unusual in the 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%.

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.

We find evidence of selective participation in our study. 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 (figure 1 panel A) or the coaching (Figure 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 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.

We now turn to the description of the methodological approaches used to estimate the effects of the treatment arms on financial behavior. We first describe the pure experimental approach that is applied in most RCTs. Next, to overcome the low take-up rate problem, we apply a combination of non-experimental methods to get at a cleaner estimate of the impact of the workshops and coaching sessions. We describe these approaches in detail below.

3. Pure Experimental Results

3.1 The Challenge of Low Take-up for Statistical Power

Consider a simple comparison of treatment and control means in a randomized experiment which allocates a proportion P of subjects to the treatment, and $1-P$ to the control. This intent-to-treat effect can then be estimated in a regression of the form:

$$Y_i = a + bT_i + \varepsilon_i \tag{1}$$

Where T_i is a dummy variable denoting assignment to treatment, and the error ε is i.i.d. with variance σ^2 . Let c be the take-up rate in the treatment group, s the take-up rate in the control group ($s=0$ if no one in the control group gets the treatment), and E the impact of treatment for those who actually receive the treatment (the treatment effect on the treated). The sample size N needed to detect effect size E at significance level α and power β is then (e.g. Duflo et al, 2008):

$$N = \left[\frac{(t_{1-\beta} + t_{\alpha/2})}{E(c-s)} \right]^2 \frac{\sigma^2}{P(1-P)} \tag{2}$$

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).

We see that the sample size required is proportional to the inverse of the difference in take-up rates squared $1/(c-s)^2$. The consequence is that low take-up rates dramatically increase the sample size 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.

3.2 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 γ over the entire nine-month post-intervention period $t=1,2,\dots,9$ via the following Ancova specification:

$$Y_{i,t} = \gamma OfferedTreatment_i + \sum_{s=1}^9 \delta_s 1(s = t) + \theta \bar{Y}_{i,PRE} + \sum_a \lambda_a 1(i \in a) + \varepsilon_{i,t} \quad (3)$$

Where $\bar{Y}_{i,PRE}$ is the mean of the outcome over the pre-treatment periods, λ_a are strata fixed effects, δ_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, we can also estimate the local-average treatment effect (LATE) by replacing *OfferedTreatment* with *ReceivedTreatment* in equation (3), and then instrumenting the receipt of treatment with its randomly assigned offer. This identifies the local-average treatment effect (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 2 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 we see 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.

Table 3 then reports the results of estimating the ITT effect in equation (3) of being offered the financial education workshop (panel A), and of being offered the coaching (panel B). These estimates are all small in magnitude, and very close to zero. That is, the offer of treatment has a very small, and insignificant impact on financial behavior. Underneath each ITT, we then report the LATE/TOT and a 95 percent confidence interval around it. We see 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 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.

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

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

4.1 Empirical Approach

The basic challenge for identifying the impact of training and coaching for those who actually took part in these interventions is that take-up is not random. Section 2 showed that those who attended the workshops or received coaching differ in current and past financial behaviors from those who did not. As a result, simply comparing those who took-up the interventions to the full set of individuals in the control group would yield biased results.

Our solution is to use the richness of the financial data available on credit card clients to combine experimental and non-experimental methods. We use propensity score matching to match individuals in the treatment group who took up the treatment 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, we estimate the following equation for the matched sample:

$$Y_{i,t} = \gamma ReceivedTreatment_{i,t} + \beta EverTreated_i + \sum_{s=-18}^9 \delta_s 1(s = t) + \sum_a \lambda_a 1(i \in a) + \varepsilon_{i,t} \quad (4)$$

Where $ReceivedTreatment_{i,t}$ takes on value one for individuals in the treatment group in post-intervention periods, and zero otherwise; $EverTreated_i$ is an indicator of whether individual i is in the treatment group and ever took-up the treatment; 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, similar to the problem of an Ashenfelter dip in labor economics experiments. 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. Our rich data and experiment help in addressing both concerns. We have up to 18 months of pre-intervention financial data for these clients, 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), we have a plausible reason why some individuals do not take up treatment – they were not invited to under the random invitations.

Difference-in-differences further enables us 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, we can difference this 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 we see the two groups have the same dynamics prior to the intervention. Researchers relying on survey data typically do not have multiple rounds of pre-intervention data with which to test this assumption. In contrast, the monthly administrative data enables us 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 non-linear trend. To test this, we estimate over the pre-intervention period:

$$Y_{i,t} = \sum_{s=-18}^{-1} \beta_s 1(s = t) EverTreated_i + \sum_{s=-18}^{-1} \delta_s 1(s = t) + \sum_a \lambda_a 1(i \in a) + \varepsilon_{i,t} \quad (5)$$

And test that all the β 's are jointly zero.

As we move away from the pure experiment, there is no one universally agreed control group. We examine several different plausible ways of choosing this control group. We then view 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.

We begin 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, we estimate the propensity score as a function of gender, and pre-treatment monthly levels of all five outcomes. This involves matching on 73 variables in total, and eliminates 38% of the control group and 22% of the treatment group for coaching. We then go 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, may not match especially well on any particular single outcome. We therefore also consider 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 our

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 3 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. We see 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 we condition 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 2 clients are selected by all three methods, so we are forming multiple plausible counterfactuals.

Appendix Table 1 shows how these three nearest neighbor matches achieve samples from the control group which are much more comparable on baseline observables to those who took up treatment than is the case for the full control sample. The first column shows baseline means for those who took-up treatment. We 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 locations, with differences less than 0.2 standard deviations typically considered to indicate balance. We see that normalized differences exceed this level for pre-intervention averages in our 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 our three nearest neighbor methods. As a result, we cannot reject equality of means of our financial outcomes averaged over all pre-intervention periods when using nearest neighbor.

4.2 Impacts of Workshops

Table 4 presents the difference-in-difference estimates of the impact of workshops on our five outcomes of interest (Panels A through E). Each column presents the results of one of the five approaches we use to form a control group. For each outcome and approach, we test whether the treated and control groups followed common linear and non-linear trends in the pre-intervention period. The p-values of the tests are included in the table. In Figure 3, we show 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 our preferred specification, the nearest neighbor approach that matches on the monthly pre-intervention data for a specific outcome of interest. As Figure 4

shows, this approach allows us 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 we cannot reject that the outcomes of clients who took the workshop and the control group formed by our preferred approach followed common linear and non-linear trends before the interventions. The economic impact of the estimates is also robust across the different matching approaches. The results of our 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 our partner bank.

4.3 Impacts of Coaching

Table 5 provides the difference-in-difference results for our five different approaches to forming a control group, along with tests of whether the two groups follow a common linear trend, and a common non-linear trend, before the intervention. Figure 5 shows the trajectory of mean outcomes for the group that received coaching, compared to our different matched control groups.

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 we cannot reject that the matched control groups display parallel linear or non-linear trends pre-intervention. Using our preferred specification in the last column (which matches Figure 5), we find 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.

4.4 Discussion

This combination of nearest neighbor matching, difference-in-differences, and the random assignment enables us 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, we detect treatment effects of the interventions

that were not able to be detected using experimental methods alone. It is worth re-iterating how little power experimental methods will have to detect treatment effects of the size we find, given the low take-up levels. For example, our power to detect the 5.9 percentage point improvement in the likelihood of paying more than the minimum that we see for the coaching treatment, given that take-up is 6.8 percent, is only 17.9 percent.¹⁴ Note that our 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. We have 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, etc., 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, we see 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.

5. Conclusions

The reliable estimation of treatment effects in impact evaluations relies heavily on the implementation efforts of teams to get individuals assigned to a treatment to be effectively treated (and to the extent possible ensure 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.

In this study, we take advantage of the richness of administrative data from a financial institution to implement a novel 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, our approach improves the empirical evaluation of the treatment in several ways. The approach improves upon the simple application of experimental methods. The experimental

¹⁴ This uses the autocorrelation of approximately 0.4 in our 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)`.

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, our 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 U.S. where response rates were estimated to reach 0.2% in 2006,¹⁵ and the banking/finance industry in the U.K., 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% in 2016.¹⁶

¹⁵ Federal Deposit Insurance Corporation “Credit Card Activities Manual” https://www.fdic.gov/regulations/examinations/credit_card/ch5.html [accessed October 15h 2017].

¹⁶ Sign-up.to 2017 report <https://www.signupto.com/email-marketing-benchmarks/email-benchmark-2017/> [accessed October 15th 2017].

References

- Banco de México (2016) “Indicadores Básicos de Tarjetas de Crédito. Datos a junio de 2016” <http://www.banxico.org.mx/sistema-financiero/publicaciones/reporte-de-tasas-de-interes-efectivas-de-tarjetas-/0%7B3A787547-BAAA-A0F2-4D1A-151E96D32321%7D.pdf> [accessed October 20th 2017].
- Brown, Amy and Kimberly Gartner (2007) “Early Intervention and Credit Cardholders”, Center for Financial Services Innovation, Chicago, IL. http://cfsinnovation.com/system/files/imported/managed_documents/earlyintervention.pdf [accessed 15 March, 2013].
- Bruhn, Miriam, Gabriel Lara Ibarra and David McKenzie (2014) “The Minimal Impact of a Large-Scale Financial Education Program in Mexico City.” *Journal of Development Economics*, 108: 184-89, 2014.
- Bursztyn, Leonardo Bruno Ferman, Stefano Fiorin, Martin Kanz and Gautam Rao (2017) “Status Goods: Experimental Evidence from Platinum Credit Cards”. Mimeo. World Bank.
- Chong, Alberto, Dean Karlan, and Martin Valdivia (2010) “Using radio and video as a means for financial education in Peru.” *Innovations for Poverty Action*. Accessed: 4 September 2017 <http://www.povertyactionlab.org/evaluation/using-radio-and-video-means-financial-education-peru>.
- Crépon, Bruno, Florencia Devoto, Esther Duflo and William Pariente (2015) “Estimating the Impact of Microcredit on those who take it up: Evidence from a randomized experiment in Morocco”, *American Economic Journal: Applied Economics* 7(1): 123-50.
- Duflo, Esther, Rachel Glennerster and Michael Kremer (2008) “Using randomization in development economics: A toolkit” In: Paul Schultz, T., John, Strauss (Eds.), *Handbook of Development Economics, Vol. 4*. North Holland, Amsterdam, NH, pp. 3895–3962.
- Fernandes, Daniel, John G. Lynch, Jr., and Richard G. Netemeyer (2014) “Financial Literacy, Financial Education, and Downstream Financial Behaviors”, *Management Science*, 60(8): 1861-1883
- Grodzicki, Daniel (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
- Karlan, Dean, Jonathan Morduch and Sendhil 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, Leora, Annamaria Lusardi and Peter 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

Lusardi, Annamaria and Olivia Mitchell (2014) “The economic importance of financial literacy: theory and evidence”, *Journal of Economic Literature* 52(1): 5-44.

Lusardi, Annamaria and Peter Tufano (2015) “Debt literacy, financial experiences, and overindebtedness”, *Journal of Pension Economics and Finance* 14(4): 332-68.

McKenzie, David (2012) “Beyond baseline and follow-up: The case for more T in experiments”, *Journal of Development Economics* 99: 210-221.

Miller, Margaret, Julia Reichelstein, Christian Salas, and Bilal Zia (2015) “Can You Help Someone Become Financially Capable? A Meta-Analysis of the Literature”, *World Bank Research Observer*, 30(2):

Mottola, Gary (2013) “In Our Best Interest: Women, Financial Literacy, and Credit Card Behavior.” *Numeracy* 6 (2), Article 4.

Ponce, Alejandro, Enrique Seira and Guillermo Zamarripa (2017) “Borrowing on the Wrong Credit Card? Evidence from Mexico” *American Economic Review*, 107:

Willis, Lauren (2011) “The financial education fallacy” *American Economic Review Papers & Proceedings* 101 (3), 429–434.

Figure 1. Take up rates

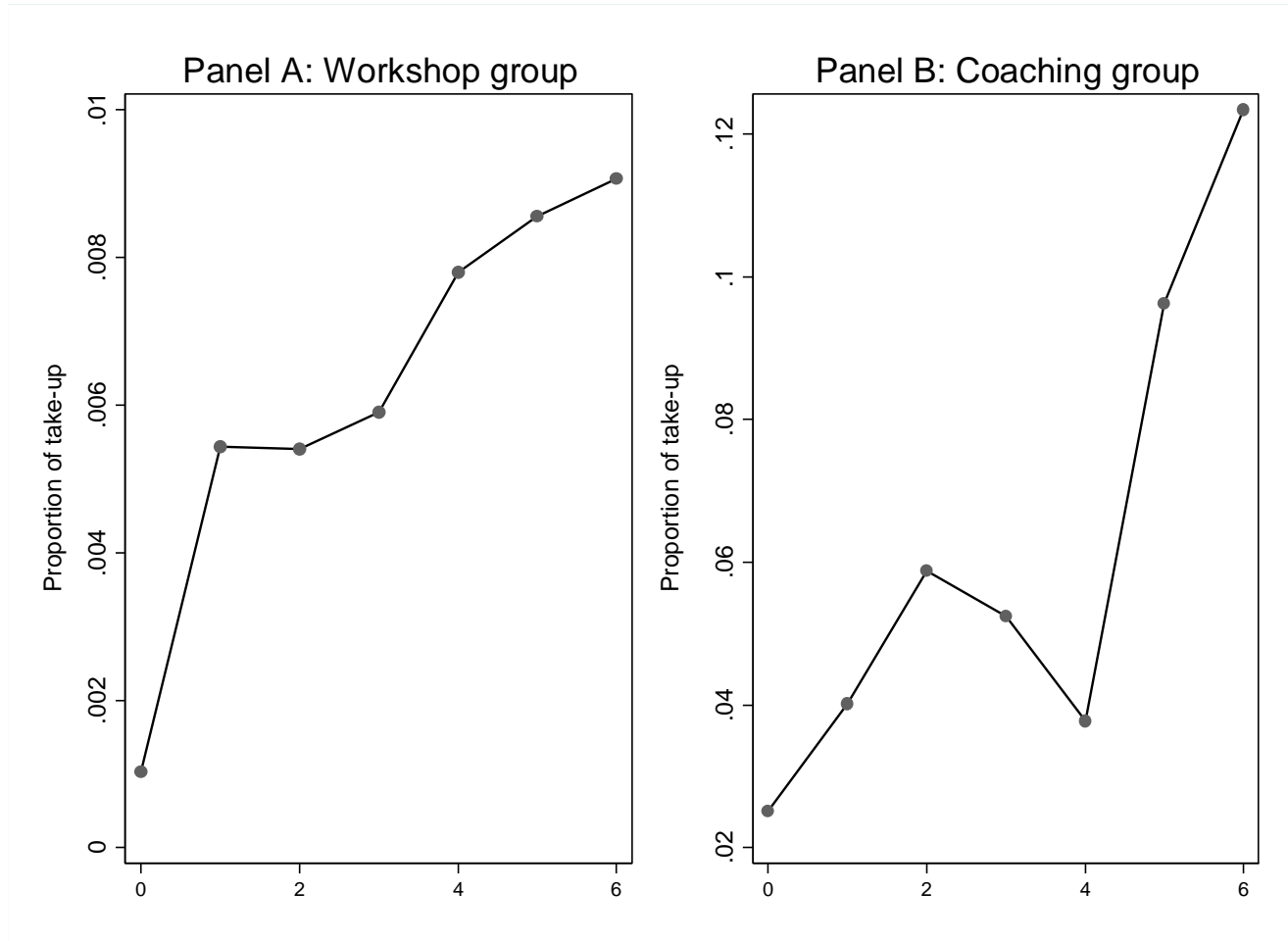


Figure 2: Evolution over time of fraction of clients paying above the required minimum and fraction of clients with delay in their payment by treatment status

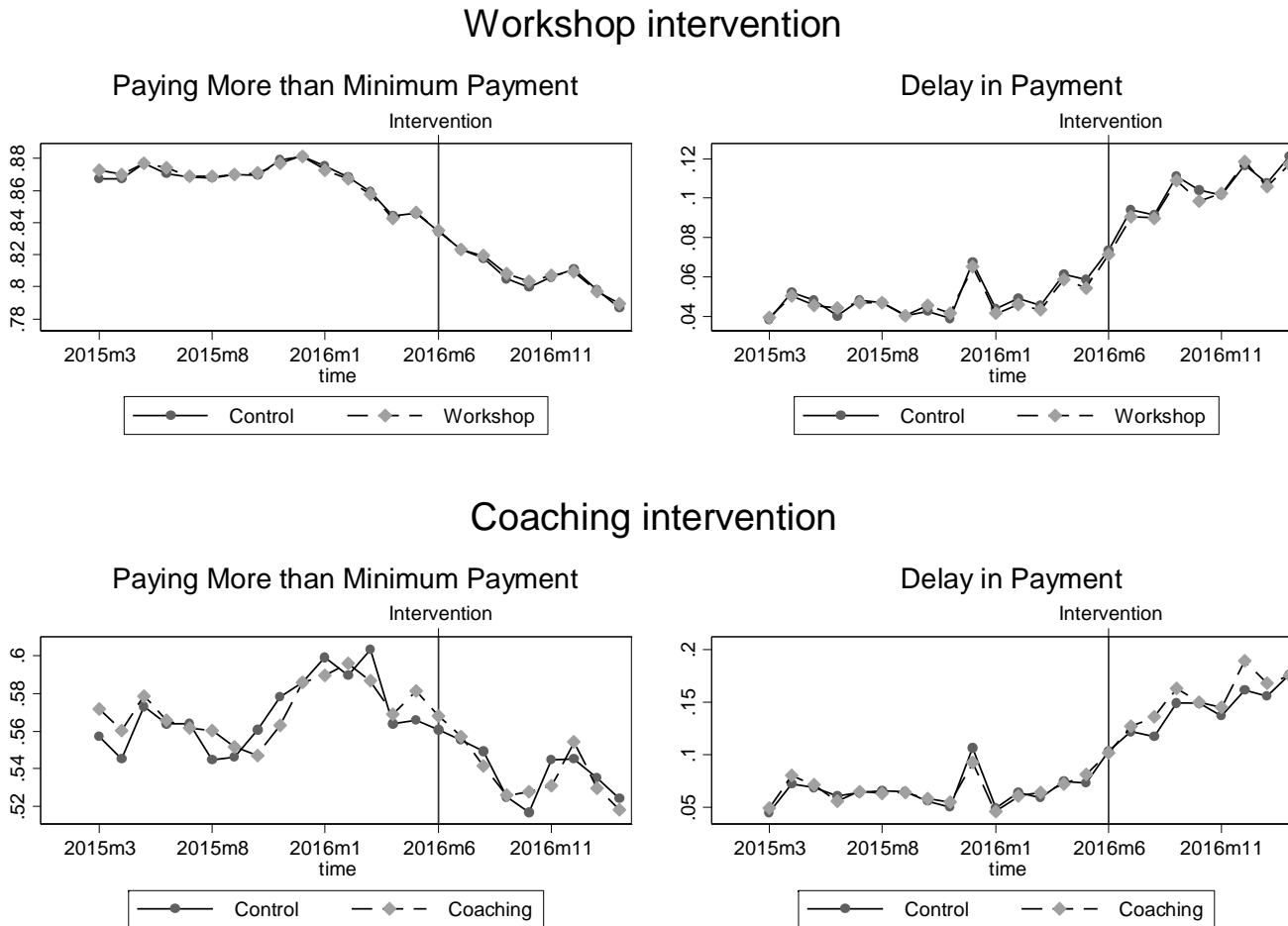
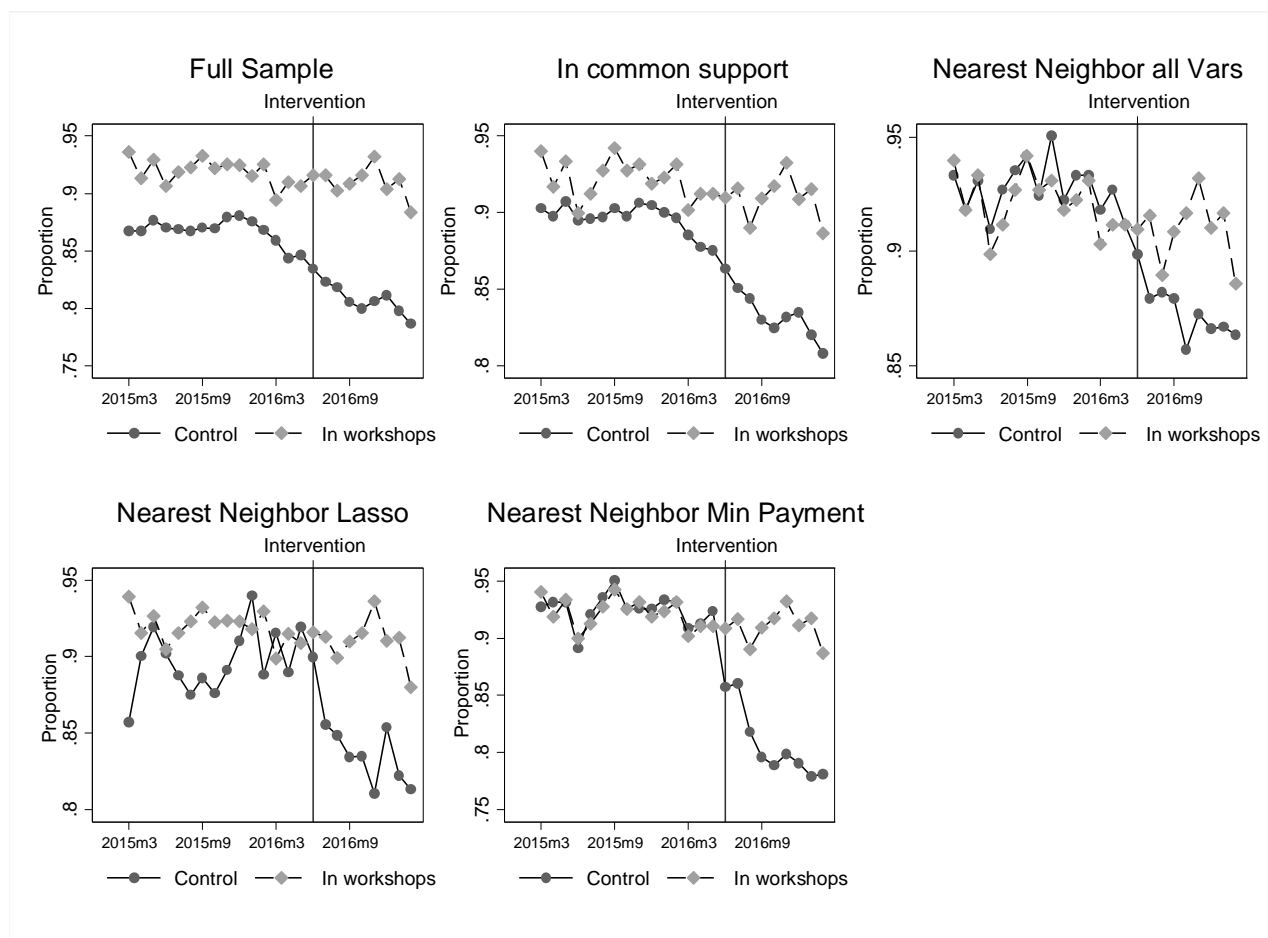
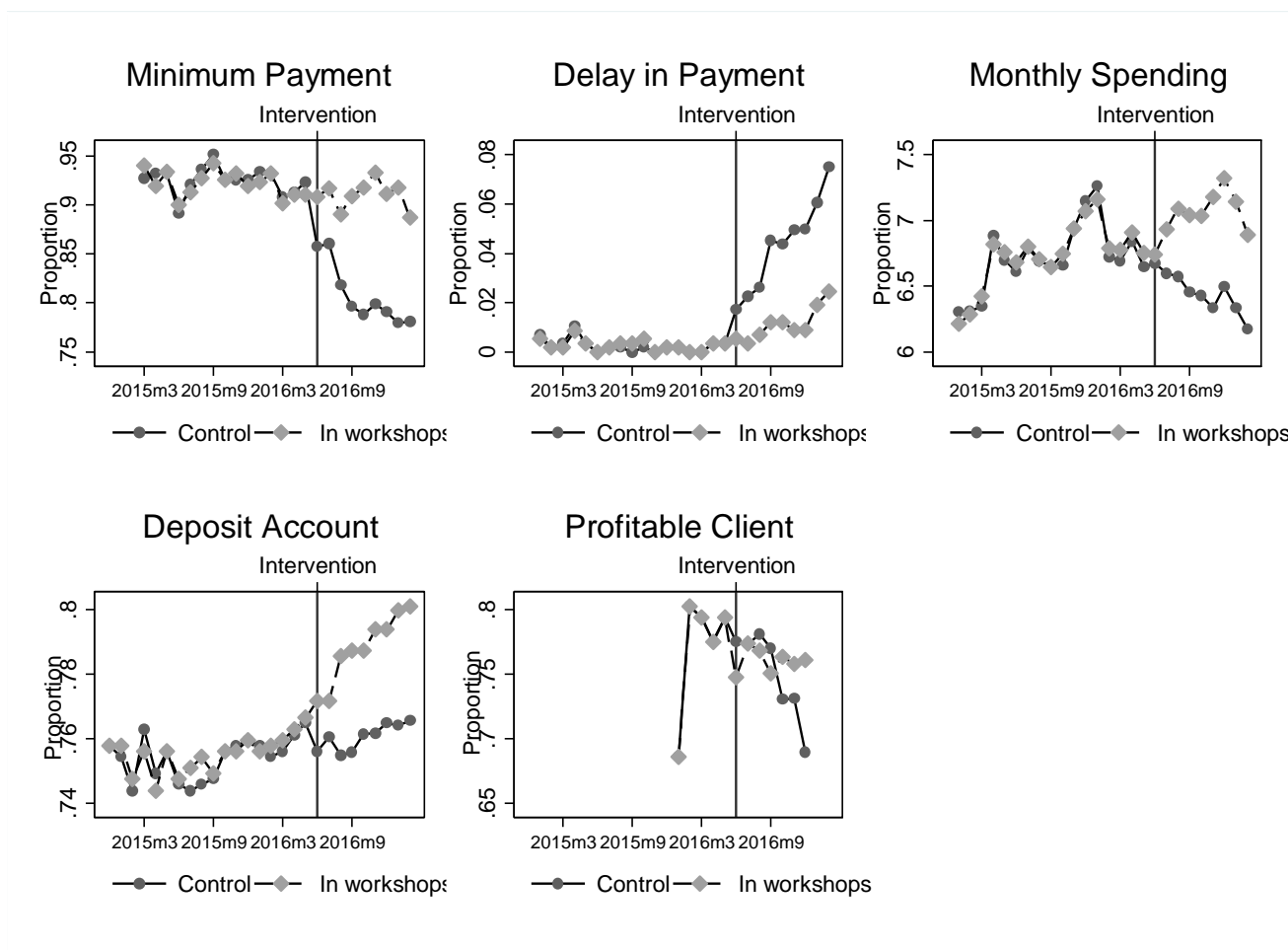


Figure 3: Illustration of the Five Different Approaches to Forming a Counterfactual, for the Workshop Treatment and Outcome of Paying More than the Minimum Payment



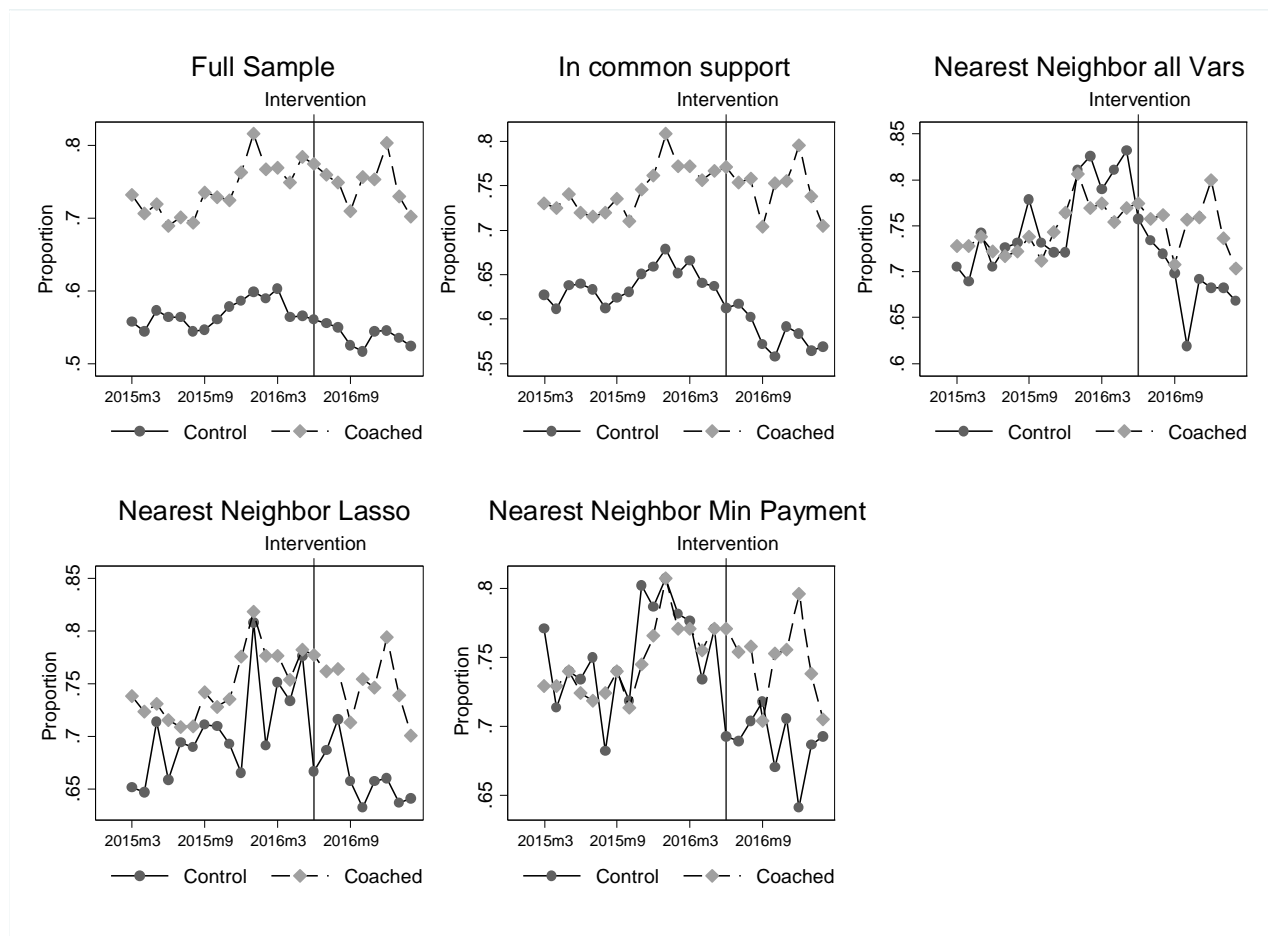
Notes: Full Sample compares means for all individuals receiving coaching to the full control group; in common support shows means for the sample within the common support of a propensity score estimated using the full history of all five outcomes plus a control for gender; Nearest neighbor all vars then shows means after single nearest neighbor matching without replacement within this common support using the propensity score estimated with all variables; 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; 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.

Figure 4: Trajectories of financial outcomes of those receiving workshops compared to nearest neighbor matched control group



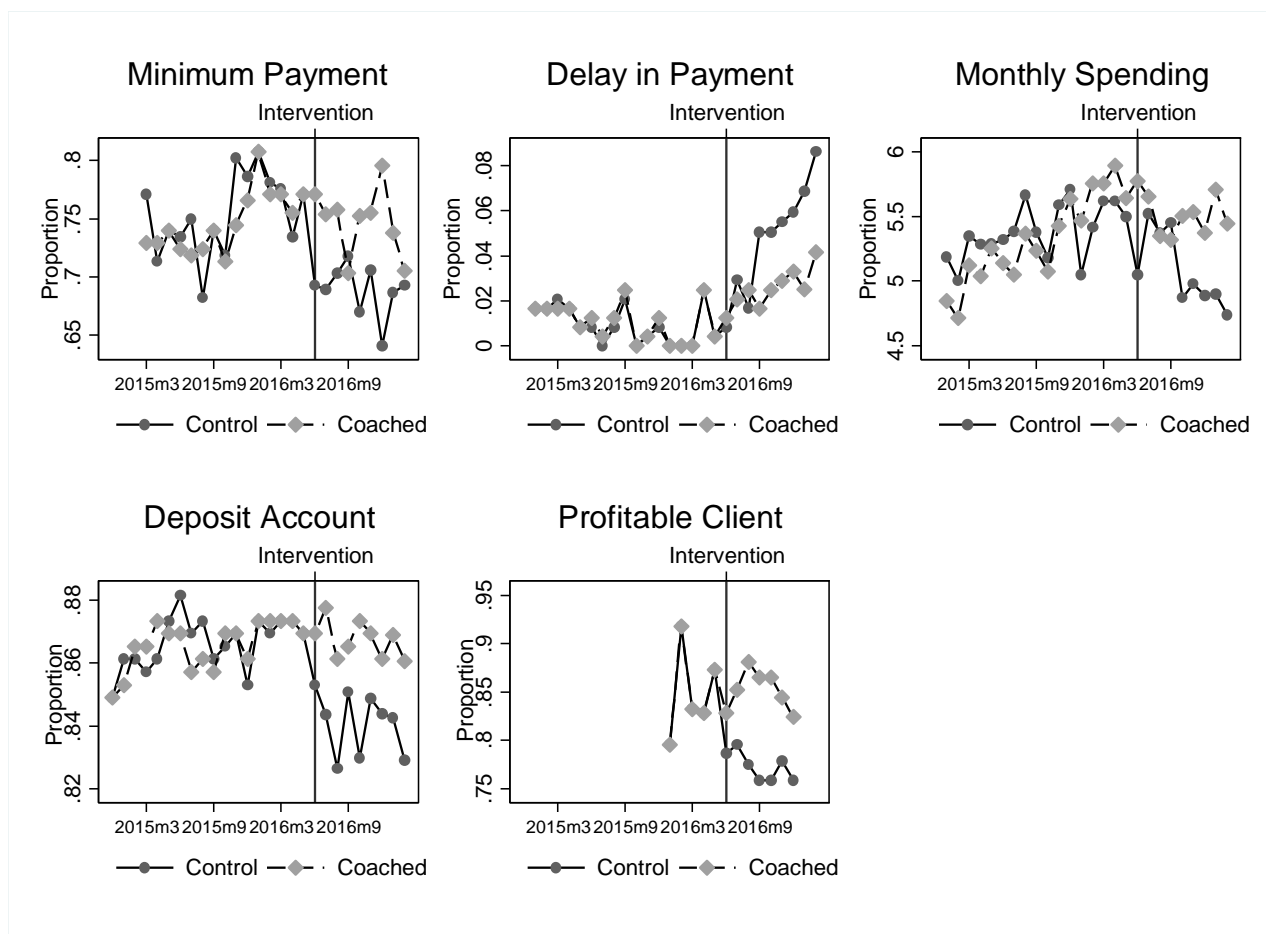
Notes: Propensity score matching used to construct a nearest neighbor matched control sample using outcome-specific pre-intervention variables. Fewer months pre- and post-intervention are available for the outcome of being a profitable client for the bank.

Figure 5: Illustration of the Five Different Approaches to Forming a Counterfactual, for the Coaching Treatment and Outcome of Paying More than the Minimum Payment



Notes: Full Sample compares means for all individuals receiving coaching to the full control group; 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; 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; 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; 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.

Figure 6: Trajectories of financial outcomes of those receiving coaching compared to nearest neighbor matched control group



Notes: 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.

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
<i>Time unvarying characteristics of clients</i>								
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 our partner bank	12	12	13	-1*	12	11	12	0
<i>Variables used for stratification</i>								
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
<i>Time varying characteristics of clients (Average over pre-intervention period)</i>								
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***
Owns deposit account	0.69	0.69	0.75	-0.06***	0.88	0.88	0.86	0.02
Profitability to the bank	1,056	1,073	1,142	-86	1,881	1,974	2,399	-518**

Notes: The table presents the summary statistics of our sample of 114,226 clients for the period of December 2014 to May 2015. 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%, 5% and 1% levels, respectively.

Table 2: The Take-up Challenge

	Workshops		Coaching	
	Number of clients	Percent	Number of clients	Percent
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%	509	14.0%
Actually received treatment	583	0.8%	246	6.8%

Source: Own calculations from the study implementation data.

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

Notes: 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.

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 the 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 non-linear 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 non-linear 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 non-linear 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 non-linear 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 non-linear pre-trend	0.173	0.0687	0.978	0.327	1.000

Notes: 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.

Table 5: Estimated Treatment Effects for those who did receive coaching

	Full control sample	In common support	Nearest Neighbor Matching		
			on all variables	using lasso	on outcome
Panel A: Pay more than the 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 non-linear 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 non-linear 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 non-linear 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 non-linear 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 non-linear pre-trend	0.079	0.432	0.962	0.0749	1.000

Notes: 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.

Appendix 1: Comparison of Pre-Intervention Means for Workshop Intervention

	Treatment	Full Control Sample		In Common Support		NN All Vars		NN Lasso		NN Outcome	
	Received Workshop	Normalized Difference	p-value	Normalized Difference	p-value	Normalized Difference	p-value	Normalized Difference	p-value	Normalized Difference	p-value
Female	0.511	-0.016	0.705	-0.015	0.753	-0.021	0.743	-0.150	0.013		
Age	45.995	-0.040	0.345	-0.065	0.166	-0.174	0.008	-0.074	0.218		
Years as Client	15.189	-0.096	0.072	-0.104	0.074	-0.260	0.000	-0.173	0.004		
Mean Min Pay	0.916	-0.286	0.000	-0.147	0.004	0.038	0.559	-0.121	0.045	0.020	0.755
Mean Delay in Paying	0.003	0.153	0.002	0.003	0.950	0.020	0.763	0.175	0.004	-0.010	0.863
Mean Log Spending	6.732	-0.441	0.000	-0.346	0.000	-0.011	0.873	-0.108	0.075	-0.007	0.910
Mean Deposit Account	0.751	-0.145	0.001	-0.147	0.002	0.035	0.592	-0.202	0.001	-0.003	0.961
Mean Profitable Client	0.770	0.181	0.000	0.200	0.000	-0.032	0.626	0.252	0.000	0.000	1.000
Sample Size	583	36946		26811		465		547		469	

Notes: The control group varies with outcome for the last approach (NN outcome), and so normalized differences and p-values are shown using the outcome specific control group.

Appendix 2: Comparison of Pre-Intervention Means for Coaching Intervention

	Treatment	Full Control Sample		In Common Support		NN All Vars		NN Lasso		NN Outcome	
	Received Coaching	Normalized Difference	p-value	Normalized Difference	p-value	Normalized Difference	p-value	Normalized Difference	p-value	Normalized Difference	p-value
Female	0.354	0.251	0.000	0.182	0.019	-0.076	0.461	0.320	0.001		
Age	45.618	0.030	0.661	0.039	0.618	0.155	0.131	0.066	0.492		
Years as Client	14.557	-0.074	0.248	-0.028	0.709	0.035	0.733	0.039	0.682		
Mean Min Pay	0.737	-0.561	0.000	-0.346	0.000	0.032	0.753	-0.150	0.117	0.024	0.813
Mean Delay in Paying	0.011	0.280	0.001	0.018	0.811	0.045	0.661	0.094	0.325	-0.029	0.747
Mean Log Spending	5.283	-0.423	0.000	-0.316	0.000	-0.027	0.790	0.087	0.361	0.020	0.830
Mean Deposit Account	0.862	0.068	0.292	0.030	0.689	-0.047	0.645	0.041	0.665	0.002	0.981
Mean Profitable Client	0.846	0.086	0.190	0.093	0.203	0.047	0.648	0.087	0.360	0.000	1.000
Sample Size	246	2504		1563		190		219		192	

Notes: The control group varies with outcome for the last approach (NN outcome), and so normalized differences and p-values are shown using the outcome specific control group.

Appendix 3a. Contents of Credit Card Financial Literacy course
















Topic	Description
Credit cards	<p>This session explains that debt can be useful if you know how to use it correctly. The session also covers how to apply for a loan and the different types of loans there are.</p> <p>Objectives:</p> <ol style="list-style-type: none"> 1. Participants know what debt is 2. Participants know what a credit card is 3. Participants know good habits with their credit cards <p>Content:</p> <ul style="list-style-type: none"> • Types of loans available • Advantages and disadvantages of each type of loan • What a credit card is • Credit cards' elements • The Personal Identification Number (NIP) • The bank statement and how to read it • Credit cards APR (Annual Percentage Rate) • Healthy use of credit cards <p>Duration: 1 hour</p> <p>Exercise:</p> <ul style="list-style-type: none"> • Case study: identifying what kind of credit is the best for each situation • Interactive exercise: identifying the parts of a bank statement
Healthy Credit rating	<p>This session focuses on understanding what a credit rating is and how to keep a good credit score.</p> <p>Objectives:</p> <ol style="list-style-type: none"> 1. Participants learn what a credit rating is 2. Participants learn how to obtain their own credit rating 3. Participants learn what they can do if they have credit problems <p>Content:</p> <ul style="list-style-type: none"> • Credit ratings and the importance of having good credit rating • The credit bureau • How and where to get a credit report • What my credit report means • Self-diagnose credit health • Advise to maintain or improve your credit rating <p>Duration: 1 hour</p> <p>Exercise:</p> <ul style="list-style-type: none"> • Case study: helping someone to solve their financial problems

Source: BBVA Bancomer Financial Education department

Appendix 3b. Golden Rules for having good credit health from BBVA Bancomer's Financial Literacy Course

15 GOLDEN RULES

Know them to keep or improve your financial health.
Consider them when using your credit card. Having them in mind will help take maximum advantage of your card:

-  **1** Write down everything that you **earn** and **spend**. This will help you plan all your purchases and save.
-  **2** When recording **expenses**, divide them into fixed (such as electricity) and variable (such as the coffee that you regularly buy)
-  **3** Purchases with your credit card are a loan that you must pay back. Consider them as part of your **expenses** and **avoid delays**.
-  **4** Know and keep in mind the **fees** and **interest rates of your credit cards**. This will help you manage it better and know its total cost.
-  **5** Always consider your **closing and due dates**, to know when to make purchases and avoid delays.
-  **6** Remember that when you buy with your credit card the **right after the closing date, you have up to 50 days to pay with no interests**.
-  **7** Failing to cover at least the **minimum payment** will put you at risk of not being able to use your credit card and **paying additional fees**.
-  **8** Failing to cover at least the **minimum payment** by the **due date** will result in **late fees and interests**.
-  **9** Remember that credit cards purchases at '**months without interest**' **are not considered in your minimum payment**.
-  **10** When you pay **more than the minimum**, your debt will decline and you will pay sooner. **The more you pay, the more your debt drops**.
-  **11** Know and use available means of payment (such as automatic payments) to **never delay in the payment of your credit card**.
-  **12** Only use credit cards that your **income and expenditures** level can allow you to pay and manage.
-  **13** Avoid loan payments **higher than 30% your monthly income**; this will help take care of your credit health.
-  **14** If you have an **extra income** throughout the month, try to use part of it to **pay or reduce** the debt in your credit card.
-  **15** You are not alone. If you have problems with your credits, you can approach your **account executive or seek for help in your bank**.

Source: BBVA Bancomer *Adelante con tu futuro* ©

Appendix 3c. Description of Coaching Sessions

Sessions	Objectives	Topics to be covered	Tools
Session 1: Diagnostic	Identify financial issues and concerns based on questions and analysis of credit situation	<ol style="list-style-type: none"> 1. Introduce the coaching sessions 2. Invite the client to participate in the sessions 3. Mention the 5-session program 4. Detect the client's issues/concerns 5. Bring solutions to the issues/concerns identified 	<ul style="list-style-type: none"> - Client's information - Diagnostic questionnaire from the Credit Health workshop
Session 2: Budget	Prepare savings plan considering variables that prevent the client to keep up with her payments	<ol style="list-style-type: none"> 1. Analyze client's expenditures 2. Classify expenditures into fixed and variable expenditures 3. Self-evaluation 4. Request to prepare the coaching worksheet for the next session 	<ul style="list-style-type: none"> - Coaching format for creating a budget
Session 3: Credit	Analyze key aspects of a credit card and credit card statement to come up with an action plan that helps improve her credit card payment behavior	<ol style="list-style-type: none"> 1. Review topics from previous session 2. Explain main parts of the bank statement 3. Highlight key elements to help improve expenditure and payment behavior, and overall use of the credit card 4. Request that coaching worksheet is ready for the next session 	<ul style="list-style-type: none"> - Client's bank statements - Credit card
Session 4: Healthy credit	Suggest several alternatives that the client can implement to pay down her loans	<ol style="list-style-type: none"> 1. Verify client's behavior on credit card payments, starting from analysis of last session's bank statement 2. Evaluate client's credit health after incorporating recommendations into purchasing and payment behavior 	<ul style="list-style-type: none"> - Coaching worksheet - Client information
Session 5: Were goals met?	Measure effectiveness of BBVA Bancomer coaching sessions among clients who stayed in the program	<ol style="list-style-type: none"> 1. Evaluate debt registry after learning how to properly use the credit card and the benefits of staying with a healthy credit history 	<ul style="list-style-type: none"> - Bank statements - Client information

Notes: Sessions 1-4 were scheduled approximately two weeks apart. The final session was scheduled one month after the fourth session. Source: Adapted from BBVA Bancomer guidance tables.

Appendix 4. Current credit card offerings by BBVA Bancomer Mexico

<i>Name of card</i>	<i>Minimum payment</i>	<i>Minimum income to be demonstrated</i>	<i>Late fee</i>	<i>Annual fee</i>	<i>Annual Percentage Rate (APR)*</i>
Platinum	20 %	50,000	377	2177	34.9%
Visa infinite	20 %	150,000	-	5275	18.6%
Oro	20 %	20,000	377	972	75.3%
Afinidad UNAM	20 %	12,000	377	972	88.0%
e	25 %	5,000	348	580	ND
Azul	20 %	6,000	377	631	90.2%
IPN	20 %	6,000	377	631	91.6%
Congelada	20 %	4,000	235	290	115.6%
Educacion	20 %	6,000	377	631	68.2%
Mi Primera Tarjeta	20 %	6,000	-	-	79.9%
Rayados	20 %	6,000	377	631	83.0%

Source: Comisión Nacional para la Protección y Defensa de los usuarios de Servicios Financieros (accessed October 13th 2017)

* APR does not include taxes