Do Savings Nudges Cause Borrowing? Evidence from a Mega Study

November 10, 2021

Abstract

We study whether savings nudges have the unintended consequence of additional borrowing in high-interest credit. We use data from an experiment that encouraged 3 million bank customers to save. We train a machine-learning algorithm to predict individual-level treatment effects and focus on individuals who are predicted to save the most in response to the nudge and who had a credit card. For them, we find a 6.1% increase in savings (208 USD PPP per month). Individuals increase their savings by spending less instead of borrowing more: for every dollar saved we can rule out increases of more than one cent in interest expenses. However, individuals who were carrying credit card debt at baseline also respond to the treatment with an increase in savings of similar magnitude and do not use the new savings to pay off existing debt. Our results suggest that individuals process saving and borrowing decisions in different mental accounts, and have implications for policy design.

Keywords: savings nudges, credit card borrowing, heterogeneous treatment effects, causal forest, co-holding puzzle JEL codes: G5, D14

1 Introduction

A vast number of policies aimed at increasing savings are currently in place, with a growing number of them involving nudges (Benartzi et al., 2017). However, when policymakers or researchers evaluate these interventions, they often focus on the immediate savings outcome and do not consider other margins of adjustment (Beshears and Kosowsky, 2020). In this paper, we are interested on the effect of saving nudges on high-interest unsecured debt. Knowing whether or not increases in savings are financed with debt is of central importance for policymakers and researchers alike. If people were to borrow more in response to the savings nudge, they would be worse off. And we know that a non-trivial fraction of households hold liquid savings while simultaneously carrying credit card debt (Telyukova, 2013).

To analyze whether or not savings nudges increase borrowing, we use a large-scale field experiment paired with comprehensive and accurate panel data of individual bank accounts and credit cards. We obtain the data from a bank in Mexico, Banorte, which ran a randomized experiment with 3,054,438 customers. 2,679,545 customers were selected as the treatment group and received (bi-)weekly ATM and SMS messages encouraging them to save for 7 weeks during the Fall of 2019, while the remaining 374,893 customers in the control group received no messages.

To provide a meaningful analysis of the impact of savings on credit card debt, we focus on individuals whose observable characteristics predict a large treatment effect from the savings nudge. Specifically, we train a causal forest (Wager and Athey, 2018; Athey et al., 2019) to predict treatment effects at the individual level. We then focus on individuals in the top quartile of the distribution of predicted treatment effects who have a credit card. These individuals increase their savings substantially. For them, we ask whether the increased savings were accompanied by an increase in borrowing. We then pay special attention to the response of those individuals who were carrying credit card debt at baseline.

If we would manually search for sub-populations with large treatment effects instead of using the causal forest, we would run into over-fitting problems. In a manual search, we would spuriously attribute large treatment effects to sub-populations in which some observations exhibit unusually large savings due to idiosyncratic shocks. Any in-sample search for large treatment effect would inevitably pick up noise, and specifically noise that leads to spuriously large effects and could affect borrowing outcomes as well. In contrast, the causal forest is based on a split-sample procedure repeated 2,000 times, in which one sample is used to partition the covariate space and another sample is used to estimate the corresponding treatment effects (Athey and Imbens, 2016). Furthermore, individual level predictions are leave-one-out predictions, and the ranking of observations according to their predicted treatment effect is based on a cross-fitted procedure over two folds (Chernozhukov et al., 2018). This eliminates the possibility that pre-treatment covariates predict a large treatment effect as a result of idiosyncratic shocks that could also affect other outcomes, including borrowing decisions.

For individuals in the top quartile of the distribution of predicted treatment effects who have a credit card, the treatment lead to a 6.1% increase in savings, from a base of 31,702 MXN (3,392 USD) in the control group. This represents an increase of 1,948 MXN (208 USD). On average, this group decreased their interest payments by 1.45% from a basis of 222 MXN with a standard error of 3.53%. We can thus rule out an increase in borrowing costs of more than 12 MXN with 95% statistical confidence. We compare this to the increase in savings and conclude that, for every 1 MXN in savings, we can rule out increase in borrowing costs of more than 1 cent (12/1,948). We then focus on the subset of individuals who roll over credit card debt as measured by their interest payments in the 6 months previous to the intervention. For them, we also see increases in savings of similar magnitude and, for every 1 MXN in savings, we can rule out increases in borrowing costs or more than 2 cents. It turns out that the magnitude of the treatment effect on savings is uncorrelated with the probability of rolling-over credit card debt in the first place.

To further illustrate the pitfalls of overfitting and how the causal forest overcomes it, we compare our results with the estimates for saving and borrowing for sub-populations with the largest ex-post observed treatment effects. To do so, we identify individuals who belong to the experimental blocks with the largest observed treatment effects. Individuals in this group could have increased their savings due to idiosyncratic shocks that also affect their borrowing. Indeed, for them, we find treatment effects on borrowing that are large and negative. This suggests that individuals who responded a lot to the treatment were actually cleaning up their finances and thus also reduced their borrowing.

We then document two more findings. First, spending decreases, as measured by ATM withdrawals and expenses with debit and credit cards. Second, we document that individuals who carried credit card debt at baseline do not use their new savings to pay off existing debt. The latter finding suggest that the savings nudge exacerbated the simultaneous holding of low-interest savings and high-interest debt.

Co-holding of liquid savings and credit card debt is common in Mexico and in the US. In our sample, the average credit card interest rate is 35.2% and checking accounts pay 0% interest. Despite the large price differences, we find that 26% of individuals who pay credit card interest keep average daily balances higher than 50% of their income in their checking accounts (over the 6 months previous to the intervention). Similarly, in the 2001 US Survey of Consumer Finances (SCF), 27% of households reported revolving an average of 5,766 USD in credit card debt paying 14% interest and simultaneously holding an average of 7,338 USD in liquid assets with a return of around 1%. A household in the SCF puzzle group loses, on average, 734 USD per year from the costs of revolving debt, which amounts to 1.5% of its total annual after-tax income (Telyukova, 2013).

The literature proposes different explanations for why households that accumulate credit card debt may not use their liquid savings to pay it off. Telyukova (2013) argues that this occurs because households anticipate needing that money in situations where credit cards cannot be used, such as making mortgage or rent payments.Haliassos and Reiter (2005) argue that individuals individuals choose to hold credit card debt and savings simultaneously in two separate mental accounts, to cope with limited self-control. Individuals who accumulate credit card debt do not pay it off with their savings because they want to constrain their impatience or the spendthriftness of their other self or spouse. If the debt were to be repaid, the impatient party would simply accumulate credit card debt again, effectively spending the savings.

Models based on mental accounting predict a null effect on debt from an increase in savings. By keeping savings in a separate mental account, individuals effectively remove a certain amount of money (labeled as savings) from their consumption-borrowing problem. Therefore individuals reduce consumption rather than increase borrowing, when they allocate whatever resources they have left. In contrast, in models without mental accounting, the future availability of money set aside for savings is taken into account in the consumption-borrowing problem. For example, in a transaction-convenience model, if individuals are nudged to set more money aside for transaction purposes, they know that money will still be available in the future. In this case, the total amount of resources available for consumption did not change. Thus, individuals borrow against future resources leaving their consumption levels essentially unchanged.

To provide further evidence for a preference-based explanation behind the co-holding puzzle, we show, as mentioned, that savings nudges reduce consumption. Additionally, we show that those individuals who co-hold, overlap most strongly with the highest quartile of the predicted treatment effects: that is, the co-holding individuals are also most responsive to the savings nudge without increasing their credit card borrowing. Finally, we note that one of the messages made reference to Banorte being "the safest money box", and this message carries a large treatment effect.

2 Literature Review

Our paper offers three separate contributions to the literature: (i) a large-scale randomized controlled trial with a rich set of outcome variables demonstrating that nudges to increase savings do not increase borrowing (ii) a careful application and discussion of state-of-the-art machine learning techniques (iii) new evidence on the co-holding puzzle.

Our paper is thus related to a large literature on the savings nudges, via automatic enrollment into 401(k) savings plans (Choi et al., 2004), or via SMS or Fintech apps (Karlan et al., 2016; Gargano and Rossi, 2020; Akbaş et al., 2016; Rodríguez and Saavedra, 2015). This literature documents positive treatment effects on savings of varying magnitude. However, the additional savings effect may be offset by future dissaving (Choukhmane, 2019) or other positions on household balance sheets. To the best of our knowledge, the only research papers looking at other positions on household balance sheets are Beshears et al. (2019) and Chetty et al. (2014).

In both of these studies, credit card borrowing is measured via biannual snapshots of balances from a credit bureau. However, snapshots of credit card balances do not reveal how much high-interest unsecured debt is actually rolled over. After all, credit card balances also reflect how much individuals spend in a given month and not only how much debt they carry. If, e.g., an individual gets unemployed, and decreases spending but also stops repaying their credit card, then the credit card balance would be negatively correlated with his or her rolled over debt. In our study, we directly observe how much credit card debt is actually rolled over and carries interest. Additionally, we can look at the treatment effect on spending, ATM withdrawals and repayments of credit card debt. Finally, we focus on a different type of nudge. While Chetty et al. (2014) and Beshears et al. (2019) study the consequences of automatic enrollment, we study the consequences of informational nudges, which are lighter and less active interventions but still highly effective for certain sub-populations. Due to their wide popularity, studying unintended effects of information nudges designed to promote savings can be important for public and private stakeholders.

Because we show that savings nudges exacerbate the co-holding puzzle, we also contribute to a growing literature looking at unintended effects of nudges in different domains, ranging from financial accounts (Beshears et al., 2015; Goldin et al., 2017; Medina, 2020) to health outcomes (Wisdom et al., 2010) and energy conservation (Costa and Kahn, 2013; Allcott and Kessler, 2019).

Our paper relates to the literature on credit card borrowing. Laibson et al. (2003) argue that credit card debt constitutes "a puzzle" for standard life-cycle models in which fully rational agents would rather forgo the benefits of consumption smoothing than borrow at prevailing credit card interest rates. Laibson et al. (2012) document that many households hold credit card debt and retirement assets. The authors explain this coexistence with time-inconsistent decision making by households, which makes them patient in the long run but impatient in the short run. Thus, households want to lock away their wealth in retirement assets to keep from consuming it. Kaplan and Violante (2014) explain the same phenomenon in a fully rational model in which households save at a higher return in their illiquid assets and then borrow in response to income fluctuations. Retirement assets in these models are illiquid because they involve a significant penalty for early withdrawal. But because that is not the case for liquid savings accounts, these models cannot explain why households co-hold savings and credit card debt.

It is puzzling that households co-hold credit card debt and perfectly liquid assets. Gross and Souleles (2002) first documented the phenomenon and noted that the transaction demand for liquidity may contribute to it. Maki (2002) studied whether households may run up credit card debt strategically in preparation for a bankruptcy filing, to be discharged during the filing while keeping assets in liquid form in order to convert them to exemptible assets. However, Telyukova (2013) indicates that most puzzle households are unlikely to file for bankruptcy. Beyond the models in Telyukova (2013) and Haliassos and Reiter (2005), proposed explanations for the co-holding puzzle include financial literacy (Gathergood and Weber, 2014), mental accounting (Gathergood and Olafsson, 2020), or the variability of credit limits (Fulford, 2015) .

A number of authors from different fields such as marketing or consumer psychology have argued in favor of spending or self-control considerations in borrowing behavior. Hoch and Loewenstein (1991) argue that self-control problems occur when the benefits of consumption are experienced earlier and are dissociated from the costs. The findings of Shefrin and Thaler (1988), Prelec and Simester (2001), and Wertenbroch (2001) suggest that liquidity enhances both the probability of making a purchase and the amount one is willing to pay for a given item being purchased over and above any effects due to the relaxation of liquidity constraints. Soman and Cheema (2002) present experimental and survey evidence that consumers interpret available credit lines as indications of future earning potential when deciding consumption expenditures.

3 Background on the Mexican Credit Card Market

The Mexican credit card market has expanded rapidly. As of June 2017, there were 17.9 million general-purpose credit card accounts in good standing holding a positive balance in a population of 124 million, whereas only 13 million cards were in circulation in 2009. In spite of this trend, credit card penetration in Mexico has remained low relative to other countries. In 2014, only 18% of adults had credit card accounts, while the equivalent figures in Brazil, Argentina, and the US were 32%, 27%, and 60%, respectively. Furthermore, the number of credit cards per individual cardholder remains relatively low compared to the US. According to a nationally representative survey, the average credit cardholder has 1.27 cards. Among individuals reporting to have at least one credit card, 79% have only one credit card, 15% have 2, and the rest have more than 2 cards.[1] Interest rates are high compared with those in the US. By the end of 2017, the average credit card interest rate in Mexico had a spread of 26.4% above the federal short-term interest rate, which was 7.17%. Banorte's average credit card interest is 35.2%.

There are 16 banks participating in the credit card market, offering 140 products. The five largest banks hold 85% of the market, the two largest products hold more than 25% of the market, and the six largest products cover just above 50%. Credit cards represent 22% of the consumer credit portfolio measured by balance, inclusive of mortgage debt at the end of 2015.[2]

[1] INEGI, Encuesta Nacional de Inclusion Financiera, 2018.
[2] Banco de Mexico, multiple reports, including: https://www.banxico.org.mx/publicaciones-y-prensa/

Several recent papers have looked at the credit card market in Mexico. (Castellanos et al., 2018; Ponce et al., 2017; Medina and Negrin, 2021). As a result, we are confident that the characteristics of the market are general enough to allow making portable inferences about consumer behavior.

4 Experimental Design and Data Description

4.1 Experiment

We analyze the results of a large-scale experiment to promote savings that was run by the Mexican bank Banorte. The experimental pool consisted of 3,054,438 customers taken as a random sample from the universe of Banorte customers satisfying the following characteristics: 1) Individuals had a payroll account with Banorte.[3] 2) Individuals kept an average daily balance of at least 50 MXN over the 2 months previous to the intervention. 3) Individuals had a valid cell phone number to receive SMS.

Out of this experimental pool, 374,893 customers were randomly selected to be in a control group. Clients in the control group received no messages. Clients in the treatment group were randomly assigned to receive 1 of 7 messages that proved to be effective in previous experiments run by the bank. Half of the treated customers were cross-randomized to receive the messages on a weekly basis, while the other half were assigned a bi-weekly frequency (i.e. one message every other week).[4] The intervention lasted 7 weeks, from September 13 to October 27, 2019.

With this large-scale intervention Banorte expanded significantly the pool of customers targeted by savings nudges. Previous experiments implemented by Banorte using the same treatment messages focused on a pool of customers selected to have higher balances or a richer portfolio of Banorte products (and thus a deeper relation with the bank). Because the experimental pool was so large and was selected with minimal constraints, we can study heterogeneous treatment

rib-tarjetas-de-credito/rib-tarjetas-credito--tasas-i.html and https://www.banxico.org.mx/publicaciones-y-prensa/reportes-sobre-las-condiciones-de-competencia-en-l/%7B9A9ADEB4-7D4E-8307-B645-DB78A8A91ADE%7D.pdf

[3]Payroll accounts are deposit accounts in which individuals receive their paychecks. These accounts are very common in the Mexican market. In contrast to regular deposit accounts, these accounts are offered to employees of companies who have an arrangement with the bank to disburse salary payments. Employees in turns are waved minimum balances and offered access to credit products with special terms. There are no restrictions as to what can be done with a payroll account: those who hold a payroll account also have access to all other products offered by the bank through standard application procedures.

[4]Users in the treatment group were further cross-randomized across two additional dimensions. First, half of them would stop receiving the messages for 2 weeks after 2 months of receiving them, and then the messages would resume. Second, half of the consumers in the treatment group would receive the same message throughout the duration of the intervention, and the other half would receive alternating messages every 4 weeks. Due to logistical considerations, these last two treatment variations were not implemented and as a result, these treatments were pooled and all individuals in the treatment group received the same message for the duration of the treatment, without interruptions.

effects overcoming the implicit selection issues of experimenting only with individuals for which the treatment is expected to work Athey et al. (2021).

The treatment messages were as follows:

> **Message 1:** "Congratulations. Your average balance over the last 12 months has been great! Continue to increase your balance and strengthen your savings."
>
> **Message 2:** "Increase the balance in your Banorte Account and get ready today for year-end expenses!"
>
> **Message 3:** "Join customers your age who already save 10% or more of their income. Commit and increase the balance in your Banorte Account by $XXX this month."[5]
>
> **Message 4:** "In Banorte, you have the safest money box! Increase your account balance by $XXX this payday and reach your goals."
>
> **Message 5:** "Increase your balance this month by $XXX and reach your dreams. Commit to it. You can do it by saving only 10% of your income."
>
> **Message 6:** "The holidays are coming. Commit to saving $XXX in your Banorte Account and avoid money shortfalls at year-end!"
>
> **Message 7:** "Be prepared for an emergency! Commit to leaving 10% more in your account. Don't withdraw all your money on payday."

We categorize the messages as follows: First, we have messages alluding to short-term savings goals (Messages 2, 6, and 7). Second, we have messages about savings more generally (Messages 1, 3, and 5). Third, we have one message that alludes to self-control problems and locking away the money (Message 4).

For each customer in the experimental pool, we observe all information routinely collected by the bank, including all transactions and balances in all checking accounts and credit cards, information from the credit bureau, and other demographic characteristics. Our measure of income corresponds to the income estimated by the bank based on all regular deposits received by customers in their checking account. Given that our entire sample has a payroll account with Banorte, the estimates of income are particularly precise: Banorte is the bank hired by employers to disburse paychecks and has thus first hand information on salaries. In terms of credit bureau data, Banorte performs bi-monthly credit checks for all customers for who they have a valid credit check authorization. This includes all customers who have at least one credit product with Banorte, e.g., a Banorte credit card.[6] All continuous variables are winsorized at the 1 and 99 percentiles.

[5] XXX was a personalized amount representing 10% of the balance in the last 3 months.

[6] Periodic credit checks are an input to market personalized credit offers which is a core function of the consumer

4.2 Descriptive Statistics

4.2.1 Descriptive statistics for treatment and control groups

Table 1 shows descriptive statistics for treatment and control groups with and without credit cards. We can see that the average age is 45 years, the average monthly after-tax income is approximately 13,500 MXN (1,441 USD), and the clients have banked with the bank for 7 years on average.[7] Additionally, their average checking account balance is approximately 19,384 MXN and about 30% of credit card holders pay credit card interest.

Beyond showing these descriptive statistics for all individuals, we also show them separately for the ones who have a credit card with Banorte. These individuals have about 30% more income and 60% higher checking account balances than the average client. Their average credit card balance is 21,914 MXN (2,339 USD) with a median of 6,056 MXN (646 USD). The average individual with a credit card pays 169 MXN (18 USD) in interest costs per month (this average includes individuals who do not pay any interest and the median interest payment is 0). Individuals also have substantial borrowing capacity on their cards, 102,278 MXN on average and 40,000 MXN as the median.

4.2.2 Randomization checks

The experiment was stratified along a number of dimensions: income quartiles, age quartiles, median of tenure with the bank, quartiles of baseline savings, dummy for clients for whom Banorte is the main bank, dummy for clients considered predominantly digital (30% or less of debit card charges made through cash withdrawals), median of ATM transactions, terciles of debit card transactions, and a dummy variable indicating if an individual had a credit card. The baseline refers to the 6 months previous to the intervention. Table 2 shows that there is covariate balance across a number of variables of interest. More specifically, Table 2 shows the same descriptive statistics separately for the treatment and control groups and also shows the results of the randomization check. The randomization appears successful, as none of the differences between the two groups are statistically significant except for age: the treatment group is 1 month younger than the control group. We argue that this difference is due to chance and not economically meaningful.

---

relationship management team. These credit checks do not affect individuals' credit scores (they are analogous to soft inquiries in the US).

[7] As of the end of 2019, 1 MXN corresponded to 0.107 USD PPP, based on OECD coversion rates available at https://www.oecd-ilibrary.org/economics/data/aggregate-national-accounts/ppps-and-exchange-rates_data-00004-en.

4.2.3 Co-Holding of Credit Card Debt and Liquid Savings

In terms of the co-holding of savings and credit card debt, Table 3 shows the fraction of individuals who pay credit card interest and their balances on checking accounts, credit cards, and interest payments by deciles of account balances over income. Here, we restrict the sample to only individuals who have a credit card. We can see that, even among those individuals in the higher deciles of checking account balances, 20% to 30% pay credit card interest. The 30% of individuals with the highest checking account balances could repay their entire credit card debt and save around 1,300 MXN per month (139 USD). After all, as mentioned, Banorte's average credit card interest is 35.2% and the return on checking accounts is 0%.

Note that, we also observe savings account balances but they are rarely used and most individuals do not have one.[8]

We now look at all individuals rolling over credit card debt and define the co-holding puzzle group as individuals holding more than 50% of their income in their checking accounts and paying credit card interest. About 26% of individuals who pay credit card interest are in the puzzle group. This corresponds to about 8% of all individuals who have a credit card. Table 4 compares individuals in the puzzle group to the rest of those who pay credit card interest. The puzzle group is slightly older but has similar monthly income and tenure with the bank. They mostly differ in their checking account and credit card balances and seem to roll over more debt.

Additionally, note that, individuals appear to hold debt persistently: there is a correlation of 80% between rolling over debt in any given month and doing so in the previous month.

5 Methodology

For every customer, we observe the daily balances in their checking accounts at the end of each day. We calculate the average of daily checking account balances over the 7-week treatment period as our main dependent variable. The outcome variable is the natural log of checking account balances plus one. We then analyze the effects of the experiment using two approaches. First, we evaluate the effects of the savings nudges on checking account balances for the entire population. For this, we use standard ordinary least squares (OLS) specifications comparing treatment to control outcomes.

Second, we use machine learning techniques to predict individual treatment effects. Specifically, we estimate a causal forest, as discussed in Athey et al. (2019). In turn, we look at the sub-population with the largest predicted treatment effects. For them, we will study the borrowing consequences of saving by looking at treatment effects on savings and credit card outcomes.

[8]Less than 1% of users in our sample have a saving account, and the average balance on them is 57 MXN.

The typical way to estimate heterogeneous treatment effects in low-dimension settings is by interacting a variable that captures a heterogeneity of interest (for example, a dummy variable for observations above or below the median age) with the treatment indicator. The interaction coefficient then identifies the incremental effect of the treatment on individuals above or below the median age. If there are several potential explanatory variables, the dimensionality of the problem grows significantly, since one would need to interact all variables of interest with each other and with the treatment variable. Researchers then run the risk of overfitting or capturing spurious heterogeneous treatment effects, that is, an interaction shows up as significant by pure chance.

The causal forest algorithm allows us to identify heterogeneity in treatment effects without concern about invalidating inference due to overfitting. This method is tailored to efficiently predict the causal effects of a treatment for a rich set of different sub-populations through three distinctive features: sample splitting, orthogonalization, and optimization on an objective function designed to capture treatment effect heterogeneity.

Causal forests are different from off-the-shelf machine learning methods in three ways: 1) they estimate treatment effects with a repeated split sample method by which one sample is used to identify splitting rules and a different sample is used to estimate treatment effects (Athey and Imbens, 2016), 2) the splitting rule for the trees is defined to find sub-populations with different treatment effects instead of predicting levels of the outcome of interest in the treatment and control groups separately, and 3) orthogonalization methods are used to ensure covariate balance across multiple sub-populations. Appendix A.1 provides additional details on the intuition behind causal forests as well as the specific implementation of causal forests in conjunction with the generalized random forest algorithm developed by Athey et al. (2019). We will further discuss the rationale and findings of applying a causal forest in Subsection 6.2.

6 Results

6.1 Aggregate Effects of the Intervention

We study the treatment effect of the intervention on saving for the entire experimental pool as well as the treatment effect on saving and borrowing for individuals who have a credit card. To do so, we estimate Equation (1).

$$Y_i = \alpha_s + \beta * treatment_i + \epsilon_i \tag{1}$$

where α_s represents fixed effects for randomization blocks and β identifies the treatment effect of the intervention as the difference in outcomes between the treatment and control groups.

Table 5 shows the treatment effects across all treatments by treatment message and treatment frequency. Column 1 shows that, on average, there is a significant 0.6% increase in savings from a basis of 21,867 MXN. Column 2 displays the effects by treatment message, showing that only Message 2, individually, has a positive treatment effect. Column 3 shows that only the treatment with weekly messages has a positive treatment effect on its own. However, even though not all treatments lead to significant effects on their own, all treatment messages and frequencies have similar coefficients that are not statistically significantly different from each other.

Columns 4 and 5 show the treatment effect for individuals who have a credit card. In this column, we again pool all treatments into one single dummy variable that takes the value of 1 if a given individual was assigned to any of the treatments. Here, we find a significant 1.4% increase in savings from a basis of 24,331 MXN, which represents an increase in savings of 340 MXN. We then explore if there is any increase in credit card interest payments but do not find a significant effect. This null effect is tightly estimated: We can rule out an increase in credit card interest of more than 0.3% with 95% statistical confidence on a basis of 213.84 MXN, that is, we can rule out an increase of more than 0.64 MXN in borrowing costs. Thus, in this aggregate specification, for every 1% increase in savings, individuals with credit cards incurred less than a 0.64/340 or 0.19% increase in borrowing costs.

Note that these treatment effects (TE) are intention-to-treat estimates because individuals may or may not have seen the messages; if they do see the messages, they then choose how much to respond. The fact that we find a positive and significant effect in a randomized setting implies that at least some individuals saw the messages and that their behavior was affected by them.

Consistent with the previous literature on savings nudges via SMS, the documented impact is relatively small (Karlan et al., 2016). The fact that there was a stronger effect on savings among credit cardholders suggests that there may be some sub-populations with a stronger response than others. We thus study heterogeneity of treatment effects in the following subsection.

6.2 Heterogeneous Effects and Sub-Population Analysis

We pay special attention to heterogeneous treatment effects for two reasons. First, previous work has found moderate effects of nudging interventions via SMS on savings. We argue that this occurs because the average effect masks heterogeneities, with some individuals responding strongly while others remain unaffected. Our setting allows us to characterize sub-populations who respond to savings nudges and provide insights on how to perform targeted interventions. Second, any meaningful test of the effect of savings nudges on borrowing requires to first have a strong effect on savings. To identify individuals with the highest response to the treatment, we use a causal forest.

6.2.1 Characterizing Individuals with the Largest Predicted Responses to Savings Nudges

As described, the causal forest provides a predicted treatment effect for each individual in the sample (both treatment and control groups). We first train a pilot causal forest with 2,000 trees using all 161 pre-treatment variables available for the analysis. These variables include past financial behavior (for example, the previous 6 months of checking account balances, credit card balances, and credit card interest payments), demographic variables, and a number of geographic dummies. Then, following Athey and Wager (2019), we train a second forest only on the 52 variables with a variable importance larger than 1%. For this second causal forest estimation, Figure 1 shows the variable importance of the variables used in the analysis. This will be the basis for our subsequent analysis.

Figure 2 shows the distribution of predicted treatment effects. We can see that there seems to be dispersion in the distribution of predicted treatment effects, suggesting that there indeed be heterogeneity in treatment effects. Appendix A.1.1 provides a formal test for the validity of individual treatment effects to study actual treatment effects (Chernozhukov et al., 2018), both in the aggregate and as a sorting score. In addition to the formal test, we perform a cross-fitted ranking of predicted treatment effects and calculate the actual treatment effect for each group (Chernozhukov et al., 2018; Abadie et al., 2018). Specifically, we split the sample into two folds, for each fold we train a causal forest and use the resulting model to predict treatment effects on the other fold. We then assign observations within each fold to quartiles of the corresponding distribution of predicted treatment effect. The resulting ranking is a mapping from pre-treatment covariates to quartiles of predicted treatment effects, in which the outcome variable of observations in each fold is not used in the prediction and ranking process of observations in the same fold.

Figure 3 how how the treatment effects on savings are larger for individuals with larger predicted treatment effects, suggesting that the predicted treatment effects are a valid sorting score for the actual treatment effects. We note that while there are several observations with a negative *predicted* treatment effect, none of the quartile splits shows a negative treatment effect. In essence, the forest is identifying two groups of individuals: a large first group with a zero treatment effect (quartiles 1 to 3 of predicted treatment effect), and a second smaller group with a positive and significant treatment effect (the top quartile of predicted treatment effects).The first group is predicted with a large amount of noise, since the predictions span a large range of negative and positive numbers that all have an actual treatment effect of zero, and are indistinguishable from each other. In contrast the top quartile of predicted treatment effects have a positive and significant treatment effect. Following an analogous cross-fitting procedure, we further split observations into quintiles of predicted treatment effects (See Figure 4). We can see that these groups are once again properly sorted, based on their actual treatment effects. Figures A6 and A7 show that the treatment effect of the intervention on observations in the top segments of the distribution of predicted treatment effects is economically and statistically significantly different than the treatment effect of the observation in the bottom segment of this distribution.

To characterize individuals with the largest predicted response to savings nudges, Table A2 compares the baseline characteristics of individuals in the top and bottom quartiles of the distribution of the predicted treatment effects. Compared to individuals in the bottom quartile of the distribution of the predicted treatment effects, individuals with the highest predicted response are about one year older and have a higher income, longer tenure with the bank, larger checking account balances, larger credit card balances, and larger credit card limits.

6.2.2 Addressing Overfitting Concerns

We now focus on individuals in the top quartile of the predicted treatment effect distribution who have a credit card. For them, we calculate the treatment effect on savings and then borrowing. We note that the individual predictions produced by the causal forest are based on pre-treatment covariates and result from a procedure based on sample splitting and orthogonalization. In addition, the ranking into quartiles is defined with a cross-fitted procedure over two folds in wich the dependent variable in each fold is never used in the prediction and ranking of observations in the same fold. We do not search manually for large treatment effects over multiple partitions of the entire dataset, since, in that case, our analysis would suffer from a type of "reverse endogeneity" overfitting (that is, we would pick a group of individuals who displayed large savings in response to the treatment when in reality the large savings arised only by noise). Instead, our predictions are based on 2,000 causal trees, each trained with a different sample, which is further split into a splitting sample and an estimation sample. Individuals in the top quartile of the predicted treatment effects are those whose observable characteristics consistently predict high treatment effects across the multiple training samples and folds.

6.2.3 Ensuring Covariate Balance

Furthermore, since the top quartile of the predicted treatment effects is an arbitrary sample cut from the perspective of the experimental design, covariate balance between the treatment and control groups is not ensured. Therefore, instead of calculating treatment effects with a simple regression of treatment status on the outcome, we adjust our treatment effect estimates by treatment propensity or covariate imbalance using a variation of the Adjusted Inverse Probability Weighted (AIPW) estimator of Robins et al. (1994), as implemented by Athey et al. (2019) in the grf package of R. AIPW estimators are based on calculating the propensity to be in the treatment group given observable characteristics (Glynn and Quinn, 2010). Under perfect covariate balance, treatment propensity is constant across all observable characteristics. But while successful randomization guarantees that this is true on average, perfect covariate balance is not necessarily present across all partitions of the sample. AIPW effectively controls for these imbalances, thus improving the precision of our estimates.

6.3 Results for the Top Quartile of Predicted Treatment Effect Individuals

6.3.1 Effects on Saving and Borrowing

Table 6 shows the treatment effects on saving and borrowing for individuals in the top quartile of the predicted treatment effect distribution. For all continuous variables we take the natural log of one plus the variable, since they are non-negative.[9] Panel A considers all individuals who have a credit card, while Panel B focuses on the subset of individuals paying credit card interest.

We first discuss the results in Panel A. In Column (1), we can see the savings results for the top quartile of predicted treatment effect individuals who have a credit card. Here, the estimated increase in savings is 6.14% on a baseline savings of 31,702 MXN, that is, 1,948 MXN. On average, this group of individuals decreased their credit card balances by 1.41% from a basis of 17,120 MXN and a standard error of 1.07%, as can be seen in Column (2). However, as discussed, the mere credit card balance is not very informative about the actual credit card debt rolled over. Therefore, in Column (4), we can look at interest payments and see a decrease of 1.45% from a basis of 222 MXN. Column (4) shows a standard error of 3.53%. We can thus rule out an increase in borrowing costs of more than 12 MXN with 95% statistical confidence.

We can compare this to the increase in savings and conclude that, for every 1 MXN in savings, we can rule out a 12/1,947 increase in borrowing costs. In other words, we can rule out a larger than 1 cent increase in borrowing costs in response to a 1 dollar increase in savings.

In Column (3), we can see the effect of credit card balances from the credit card bureau, which also includes non-Banorte credit cards. The coefficient estimate and standard errors paint a similar picture. For each 1 dollar in savings, we can rule out a very small increase in borrowing with statistical confidence. Note that the credit bureau reports the credit card balances at the end of the month, whereas we use the average daily balances for Banorte credit cards. Nevertheless, the fact that we tightly estimate small effects reassures us that individuals do not borrow using other cards instead of their Banorte credit cards.

In Column (5), we can see the estimated effect for the likelihood of paying interest in a given month. Here, we can rule out an increase of 0.87% on a baseline probability of 46%. Thus, for every 1 MXN in savings, the increase in the likelihood to borrow is only 0.0087/1,904, or 0.0004%.

[9] In principle, credit card balances can be negative when borrowers pay more than the outstanding balance at the end of each month. However, in our data less than 1% of observations have a negative balances. These negative values are replaced by zero due to winsorization at the 1 and 99 percentiles, as described in Subsection 4.1.

Finally, in Column (6), we report results for credit card payments, that is, individuals repaying their outstanding credit card balances or rolled-over credit card debt. Here, we also document a very small and tightly estimated treatment effect. Individuals save more but do not repay more of their outstanding credit card balances or debt.

We now turn to the results in Panel B of Table 6, which corresponds to individuals in the top quartile of the predicted treatment effect distribution that also pay credit card interest at baseline, i.e., we observe positive average interest payments in the 6 months previous to the intervention. For this group, we have an increase in savings of 5.57% (1,295 MXN) on a baseline of 23,244 MXN. In turn, we can rule out an increase of 26.08 MXN in borrowing costs. To conclude, for every 1 MXN in savings, we can thus rule out increases larger than 2 cents (27/1,315) increase in credit card borrowing costs. Table A3 repeats the analysis expressing the dependent variables in pesos, instead of natural log. The results are consistent, although the coefficients are estimated with less precision, since the dependent variables are highly skewed. We thus use the natural log specification as our preferred one, thorough the analysis.

6.3.2 Effects on Spending and Income

We want to know whether individuals increased their savings without increasing their borrowing by decreasing their spending or increasing their income. Table 7 shows the treatment effects on deposits, ATM withdrawals, and spending for individuals in the top quartile of the predicted treatment effect distribution. We can see that the treatment effect appears to work through a 5.1% decrease in monthly ATM withdrawals and a slightly smaller but still significant 4.7% decrease in debit card spending. This is true for all individuals with a credit card and also for the subset of those paying credit card interest. We thus conclude that a decrease in spending, and in particular discretionary spending that may be financed by cash, was responsible for the increase in savings.

6.3.3 Customers with Banorte as their Main Bank

Finally, we also replicate the analysis for individuals for whom Banorte is likely to be their main bank. After all, it could be that individuals who have other bank accounts offset their additional savings using those other accounts. We say that Banorte is likely to be the main bank of a given individual when the following three conditions are satisfied: he or she receives her payroll on a Banorte payroll account (identified as such by the Mexican transaction system), he or she has a credit card with Banorte, and he or she has no credit (of any type) outside of Banorte, according to the credit bureau records. Table 8 shows the saving and borrowing results for this group of individuals. Panel A shows the results for all clients in the top quartile of the predicted treatment effect distribution and for whom Banorte is likely to be their main bank (who therefore have a credit card). We can rule out increases of more than 1 cent in borrowing cost for every additional MXN saved as a result of the savings nudge. Panel B shows the results for the subset of individuals who also incurred credit card interest at baseline. For them, we can rule out increases of more than 2 cents in borrowing cost for every additional 1 MXN saved.

6.3.4 Effects by Treatment Message

We now want to understand whether the effects on saving and borrowing differ across treatment messages. To explore the relationships between saving and borrowing across each of the seven messages included in the experiment, we focus on the 126,458 individuals in the top quartile of the predicted treatment effect distribution who have a credit card. For them, we calculate the treatment effect of receiving each specific message on their saving and borrowing.

Table A4 shows that the borrowing effect is small and tightly estimated for all individual messages. At the individual message level, the savings effect is significant for Messages 2, 3, 4, and 5 (the messages are displayed in Subsection 4.1). This finding may help with the interpretation of our results. In particular, the effect does not seem to be constrained to messages alluding to short-term savings motives. Message 2, "[...] get ready for year-end expenses," is the only message with an individually significant coefficient that alludes to saving for the short term, whereas the other messages with individually significant coefficients do not. Additionally, Message 6 "[...] avoid money shortfalls at year-end." and Message 8 "[...] emergency [...]" did not have significant effects, even though they refer to specific short-term savings goals.

Given that the short-term versus long-term savings messages are not statistically significantly different in pair-wise comparisons, we may be in a position to extrapolate our specific savings nudge experiment to other savings nudges or more forceful measures aimed to increase savings for the longer term.

Finally, Message 4 "[...] you have the safest money box [...] reach your goals" carried a large treatment effect and alluded to the safeness of the savings and reaching general goals. This message and its effect are in line with the behavioral hypothesis of mental accounting and constraining oneself to save more.

6.4 Analysis of Methods to Identify Sub-Populations with Large Treatment Effects that are subject to Overfitting Problems

We study heterogeneous treatment effects using a causal forest (Athey et al., 2019). This method allows us to derive valid inferences for the treatment effects of the intervention across different sub-populations and to identify the sub-population with the largest treatment effect without concerns of overfitting. We now contrast this method with methods of identifying heterogeneous treatment effects based on randomization strata or ex-post observed treatment effects.

6.4.1 Potential for Prediction Error and Persistence of Credit Card Debt

Predicted treatment effects estimate actual treatment effects with error, and it is possible that some individuals with a large predicted treatment effect on savings may not respond to the nudge. To rule out the possibility that the treatment effects on borrowing outcomes are driven by individuals for whom the predictions of the causal forest are not accurate, we investigate the relation between the “prediction errors" and the treatment effect of the intervention on borrowing outcomes. In contrast to standard prediction exercises, we do not observe individual level prediction errors since actual treatment effects are never observed at the individual level. We define “prediction errors" as the difference between the simple average of individual-level predicted treatment effects of observations in a given group and the (average) treatment effect of observations in the same group, calculated with the AIPW method.

We implement the same two-fold cross-fitted procedure described above to assign observations in the top quartile of predicted treatment effects into ten decile groups.[10] For each group, we focus on individuals who had a credit card and paid interest at baseline and calculate the corresponding prediction error and the average treatment effect of the intervention on credit card interest. Figure A8 shows a scatter plot of these two variables. We can see that, as expected, prediction errors are uncorrelated with treatment effects on borrowing outcomes. The prediction errors is thus the result of noise, which is uncorrelated with the treatment.

We also note that while individuals who paid interest on their credit cards during the baseline period have a 73% probability of incurring interest during the treatment period, it is possible that the treatment effect on savings documented in Panel B of Table 6 is driven by those for whom interest payments were not persistent. To investigate this possibility, we examine the correlation between the fraction of individuals actually paying credit card interest during the treatment period and the magnitude of the treatment effect on savings, across different groups of observations. As before, we split individuals in the top quartile of predicted treatment effects into decile groups. For each group, we focus on individuals who had a credit card and paid interest at baseline and the treatment effect on checking account balances. Figure A9 shows a scatter plot of these two variables. We can see that there is no clear relationship between them, suggesting that indeed, some individuals increased their savings as a result of the nudges despite carrying credit card interest.

[10] Specifically, we rank observations into 40 groups, and focus the analysis on the top 10 groups, which we interpret as deciles within the top quartile of the distribution

6.4.2 Heterogeneity by Experimental Strata

A standard way to study heterogeneous treatment effects is to split the sample based on strata from the experimental design. Table 9 shows the treatment effects on savings across the experimental strata. We find limited heterogeneity across the sub-populations that were pre-selected for heterogeneity analysis before the experiment was run. Individuals with pre-treatment checking account balances in the top quartile are the ones with the largest treatment effects. For them, we find a 1.8% increase in savings (-0.006 + 0.024).

Looking at experimental strata is a useful approach to estimate how a treatment affects a sub-population of interest that is specified before the experiment takes place. However, this method is inappropriate when trying to identify the sub-population with large treatment effects. To show this, we replicate our base saving and borrowing analysis focusing on individuals in the top quartile of pre-treatment checking account balances who have a credit card. After all, pre-treatment checking account balances had high variable importance, as seen in Figure 1. For them, Table A5 shows that there is no treatment effect on savings or borrowing. Pre-treatment checking account balances are a coarse predictor of treatment effects, and they could be bundling together individuals with large and small responses to the treatment. We thus conclude that, on average, individuals in the top quartile of pre-treatment checking account balances have a large and significant response to the savings nudge, but individuals with a credit card who had pre-treatment checking account balances in the top quartile did not show a statistically significant increase in savings.

Comparing treatment effects across experimental strata thus appears inefficient when searching for the group with the largest effects because it is based on very coarse partitions of the covariate space. Instead, one could split the sample based on the ex-post observed treatment effects. For example, one could further split the sample of individuals in the top quartile of pre-treatment checking account balances by overlaying strata dimensions and ultimately calculating the treatment effects for each strata block.[11] We now show that such attempts to perform more granular partitions without adjusting for overfitting (as the causal forest does) leads to substantial bias.

6.4.3 The Pitfalls of Overfitting: Heterogeneity by Observed Treatment Effects at the Strata-Block Level

To illustrate the bias that may arise from overfitting, we split the sample into 6,104 non-empty mutually exclusive groups defined by the interaction of all experimental strata. For each group, we calculate treatment effects, and we assign to each observation in the group the treatment effect of its group. We then split the sample into quartiles based on the treatment effect assigned to each ob-

[11] We note that this is not the standard way in which people calculate heterogeneous treatment effects (and we are not aware of any study that has done so), but we use this as a limiting case of what would happen when trying to find heterogeneous treatment effects with a rich set of explanatory variables without adjusting for the risk of overfitting.

servation. The top quartile corresponds to the 25% of observations which belong to strata blocks with the highest observed treatment effect. For them, we calculate treatment effects on checking account balances, credit card interest, and credit card balances regressing the corresponding outcome variable on a treatment indicator.

The results are presented in Table 10. Column 1 shows the number of observations included in this section of the analysis. Columns 2 to 4 show the treatment effects for individuals in strata blocks with the largest observed treatment effects. We see that the increases in savings are very large. When considering all individuals, we find a 24% increase in savings. When considering only individuals with a credit card, we find a 44% increase in savings. When considering only individuals who have a credit card and who paid interest at baseline, we find a 52% increase. Additionally, these individuals show large decreases in borrowing, measured both in terms of interest (Column 3) and balances (Column 4).

In contrast, Columns 5 to 8 show the results obtained from the causal forest. Column 5 shows the number of observations included in this part of the analysis. Column 6 shows that, as described before, the increases in savings are in the order of 2 to 6%. Columns 7 and 8 show the corresponding treatment effects on borrowing and borrowing cost. These estimates, which are free of overfitting bias, are significantly smaller than the ones in Columns 2 to 4. The large overestimation we find is consistent with the discussion of Abadie et al. (2018), who also found that sample splitting reduces bias in the context of endogenous stratification.

In Table A6, we compare the overlap between the observations assigned to quartiles of predicted individual treatment effects calculated with the causal forest and the observations assigned to quartiles of the observed treatment effects, calculated for each strata block. We conclude that there is little overlap.

7 Explanations for the Credit Card Debt Puzzle

Our empirical findings imply that some individuals who are paying credit card interest respond to savings nudges with substantial increases in savings. These savings are not used to pay off credit card debt over the two billing cycles subsequent to receiving the nudge (as documented in Table 6), thus exacerbating the co-holding of low-interest savings and high-interest debt. As we discussed, the literature proposes different explanations for this behavior.

In A.4, we outline a toy model to demonstrate that a null effect on credit card borrowing after an increase in savings is inconsistent with the predictions of rational models explaining the credit card debt puzzle, e.g., Telyukova (2013). In this model, the agent has an optimal plan for consumption and savings; if she needs to save more because of transaction purposes in the future, her optimal present consumption is not affected, and she will simply borrow more to maintain the same level of optimal present consumption.

Instead, in a second toy model, we propose mental accounting and rules of thumb as a potential explanation, following Haliassos and Reiter (2005). The theoretical idea is the following: an individual has a spending account (that is, her credit card) as well as an account for savings. On her credit card, she will spend up to some personal limit. Once she gets close to that personal limit, she feels constrained and can restrict her overspending more successfully. If this individual would take her savings and repay her credit card debt, she would feel unconstrained and rack up more credit card debt. Individuals thus prefer to hold liquid savings while simultaneously holding consumer debt instead of paying off their credit card debt. They separate these two accounts mentally to cope with their overspending and self-control problems. In the model, when the agent is nudged to save more, this money is placed in a different mental account. She then effectively allocates fewer resources between the present and future and adjusts her present consumption downwards instead of borrowing more.

The first toy model thus predicts that savings nudges should increase borrowing whereas that is not true in the second toy model. In our setting, we find that individuals do not respond with borrowing when nudged to save. We thus interpret this as evidence in favor of preference-based explanations for the co-holding puzzle.

Additionally, we have three more pieces of evidence in favor of a preference-based explanation of the credit card debt puzzle.

First, in Figure 5, we plot the fraction of the co-holding puzzle population, defined as the fraction of individuals paying credit card debt interest and holding more than 50% of their income in their checking accounts, for each quartile of the savings score distribution. We can see that most co-holding individuals are in the highest quartile of the savings score distribution (approximately 40%). By focusing the analysis on the top quartile of the predicted treatment effects, we are capturing a relevant fraction of the puzzle population. This also speaks to the idea that co-holding is a psychological mechanism to exercise self-control, as it also makes individuals more susceptible to savings nudges.

Second, as we show, individuals increase their savings because they cut their spending, especially ATM withdrawals, which is likely to be used for discretionary consumption.

Third, as mentioned, Message 4 "[...] you have the safest money box [...] reach your goals" carries a large treatment effect and alludes to the safeness of the savings and reaching general goals. This message and its effect are in line with the behavioral hypothesis of mental accounting and constraining oneself to save more.

8 Conclusion

We estimate whether or not nudging individuals to save more has the unintended consequence of additional borrowing in high-interest unsecured consumer credit. We analyze the effects of a large-scale experiment in which 3.1 million bank customers were nudged to save more via bi-weekly SMS and ATM messages over 7 weeks. We uncover strong heterogeneities in the magnitude of the treatment effects. Compared to the control group, the subset of customers in the top quartile of the predicted treatment effect distribution increased their savings considerably. However, this increase in savings was not accompanied by an increase in rolled over high-interest unsecured consumer debt. We thus conclude that savings nudges do not cause borrowing.

However, some individuals increased their savings in response to the nudge even when they held credit card debt, thus exacerbating the co-holding of high interest debt and low interest savings, referred to as the co-holding or credit card debt puzzle.

Our results can help us to understand the mechanism behind the co-holding puzzle. We argue that a null effect in credit card borrowing is more consistent with the predictions of behavioral rather than rational explanations of the credit card debt puzzle. In particular, we argue that our results are consistent with the idea that individuals hold savings and credit card debt simultaneously to deal with self-control problems via mental accounting. That is, they can maintain a rule of not touching their savings (that are parked in a separate mental account) but are simultaneously indebted due to overspending. This explanation is also consistent with the observed decrease in spending and the fact that the co-holding puzzle population strongly overlaps with the population that respond a lot to the savings nudge, which exacerbates their co-holding problems.

References

Abadie, A., M. M. Chingos, and M. R. West (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics 100*(4), 567–580.

Akbaş, M., D. Ariely, D. A. Robalino, and M. Weber (2016). How to help the poor to save a bit: Evidence from a field experiment in kenya.

Allcott, H. and J. B. Kessler (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics 11*(1), 236–76.

Athey, S., V. Hadad, N. Keleher, O. Medina, R. Nissan, R. Rosemberg, M. Schaelling, J. Spiess, and J. Wright (2021). Computational applications to behavioral science. Technical report.

Athey, S. and G. W. Imbens (2016). Recursive partitioning for heterogeneous causal effects. *Proceedings of the National Academy of Sciences 113*, 7353–7360.

Athey, S., J. Tibshirani, S. Wager, et al. (2019). Generalized random forests. *The Annals of Statistics 47*(2), 1148–1178.

Athey, S. and S. Wager (2019). Estimating treatment effects with causal forests: An application. *arXiv preprint arXiv:1902.07409*.

Benartzi, S., J. Beshears, K. L. Milkman, C. R. Sunstein, R. H. Thaler, M. Shankar, W. Tucker-Ray, W. J. Congdon, and S. Galing (2017). Should governments invest more in nudging? *Psychological science 28*(8), 1041–1055.

Bertaut, C., M. Haliassos, and M. Reiter (2009). Credit Card Debt Puzzles and Debt Revolvers for Self Control. *Review of Finance 13(4)*, 657–692.

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, and K. L. Milkman (2015). The effect of providing peer information on retirement savings decisions. *The Journal of finance 70*(3), 1161–1201.

Beshears, J., J. J. Choi, D. Laibson, B. C. Madrian, and W. L. Skimmyhorn (2019). Borrowing to save? the impact of automatic enrollment on debt. Technical report, National Bureau of Economic Research.

Beshears, J. and H. Kosowsky (2020). Nudging: Progress to date and future directions. *Organizational Behavior and Human Decision Processes 161*, 3–19.

Castellanos, S. G., D. J. Hernández, A. Mahajan, and E. Seira (2018). Expanding financial access via credit cards: Evidence from mexico. Technical report, National Bureau of Economic Research.

Chernozhukov, V., M. Demirer, E. Duflo, and I. Fernandez-Val (2018). Generic machine learning inference on heterogenous treatment effects in randomized experiments. Technical report, National Bureau of Economic Research.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen, and T. Olsen (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from denmark. *The Quarterly Journal of Economics 129*(3), 1141–1219.

Choi, J. J., D. Laibson, B. C. Madrian, and A. Metrick (2004). For better or for worse: Default effects and 401 (k) savings behavior. In *Perspectives on the Economics of Aging*, pp. 81–126. University of Chicago Press.

Choukhmane, T. (2019). Default options and retirement saving dynamics. In *112th annual conference on taxation. NTA*.

Costa, D. L. and M. E. Kahn (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association 11*(3), 680–702.

Fulford, S. L. (2015). How important is variability in consumer credit limits? *Journal of Monetary Economics 72*, 42–63.

Gargano, A. and A. G. Rossi (2020). There's an app for that: Goal-setting and saving in the fintech era. *Available at SSRN*.

Gathergood, J. and A. Olafsson (2020). The co-holding puzzle: New evidence from transaction-level data.

Gathergood, J. and J. Weber (2014). Self-control, financial literacy & the co-holding puzzle. *Journal of Economic Behavior & Organization 107*, 455–469.

Glynn, A. N. and K. M. Quinn (2010). An introduction to the augmented inverse propensity weighted estimator. *Political analysis 18*(1), 36–56.

Goldin, J., T. Homonoff, and W. Tucker-Ray (2017). Retirement contribution rate nudges and plan participation: Evidence from a field experiment. *American Economic Review 107*(5), 456–61.

Gross, D. B. and N. S. Souleles (2002). Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data. *The Quarterly Journal of Economics 117*(1), 149–185.

Haliassos, M. and M. Reiter (2005). Credit card debt puzzles. Technical report, CFS Working Paper.

Hoch, S. J. and G. F. Loewenstein (1991). Time-inconsistent preferences and consumer self-control. *Journal of consumer research 17*(4), 492–507.

Kaplan, G. and G. Violante (2014). A Model of the Consumption Response to Fiscal Stimulus Payments. *Econometrica 82*, 1199–1239.

Karlan, D., M. McConnell, S. Mullainathan, and J. Zinman (2016). Getting to the top of mind: How reminders increase saving. *Management Science 62*(12), 3393–3411.

Laibson, D., A. Repetto, and J. Tobacman (2003). A Debt Puzzle. *in Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, eds., Knowledge, Information, and Expectations in Modern Economics: In Honor of Edmund S. Phelps, Princeton: Princeton University Press*.

Laibson, D., A. Repetto, and J. Tobacman (2007). Estimating Discount Functions with Consumption Choices over the Lifecycle.

Laibson, D., A. Repetto, and J. Tobacman (2012). Estimating Discount Functions with Consumption Choices over the Lifecycle. *Working Paper*.

Maki, D. M. (2002). The growth of consumer credit and the household debt service burden. In *The impact of public policy on consumer credit*, pp. 43–68. Springer.

Medina, P. (2020). Side-Effects of Nudges: Evidence from a Randomized Intervention in the Credit Card Market. *Review of Financial Studies*.

Medina, P. C. and J. L. Negrin (2021). The hidden role of contract terms: Evidence from credit card minimum payments in mexico. *Management Science, forthcoming*.

Ponce, A., E. Seira, and G. Zamarripa (2017). Borrowing on the wrong credit card? evidence from mexico. *American Economic Review 107*(4), 1335–61.

Prelec, D. and D. Simester (2001). Always leave home without it: A further investigation of the credit-card effect on willingness to pay. *Marketing letters 12*(1), 5–12.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association 89*(427), 846–866.

Rodríguez, C. and J. Saavedra (2015). Nudging youth to develop savings habits: Experimental evidence using sms messages. *CESR-Schaeffer Working Paper* (2015-018).

Romano, J. P. and M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

Shefrin, H. M. and R. H. Thaler (1988). The Behavioral Life-Cycle Hypothesis. *Economic Inquiry 26*(4), 609–643.

Soman, D. and A. Cheema (2002). The effect of credit on spending decisions: The role of the credit limit and credibility. *Marketing Science 21*(1), 32–53.

Telyukova, I. A. (2013). Household need for liquidity and the credit card debt puzzle. *Review of Economic Studies 80*(3), 1148–1177.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

Wertenbroch, K. (2001). Self-rationing: Self-control in consumer choice.

Wisdom, J., J. S. Downs, and G. Loewenstein (2010). Promoting healthy choices: Information versus convenience. *American Economic Journal: Applied Economics 2*(2), 164–78.

Figures and Tables

Table 1: Descriptive Statistics

All Individuals (N= 3,054,503)	Mean	Std dev	P25	P50	P75
Age (years)	44.72	16.35	31.00	43.00	56.00
Monthly Income ($)	13,499.86	13,711.68	6,116.67	9,866.88	15,005.78
Tenure (months)	81.67	73.16	22.00	59.33	125.37
Checking Account Balance ($)	19,384.03	52,565.83	729.00	2,295.69	10,402.39
Fraction with Credit Card	0.12	0.32	0.00	0.00	0.00
Credit Card Interest ($)	20.04	120.24	0.00	0.00	0.00
Credit Card Balance ($)	3,879.84	16,602.93	0.00	0.00	0.00
Credit Card Limit ($)	17,168.81	67,247.74	0.00	0.00	0.00
Individuals with Credit Cards (N=362,223)					
	Mean	Std dev	P25	P50	P75
Age (years)	43.15	13.04	33.00	42.00	53.00
Monthly Income	19,744.77	18,653.78	9,071.32	13,912.75	22,718.28
Tenure (months)	103.65	73.12	43.27	86.43	148.53
Balance Checking Account	32,191.10	70,646.63	1,581.29	5,157.02	23,069.07
Credit Card Interest	168.91	311.01	0.00	0.00	170.01
Credit Card Balance	21,914.28	34,666.06	85.17	6,055.66	25,297.75
Credit Card Limit	102,277.57	137,313.20	14,000.00	40,000.00	123,999.00

This table presents summary statistics for all individuals in the experiment, and for the subset of individuals who have a credit card. For each individual, we consider information from the 6 months previous to the intervention. Monthly income, balances, and interest payments are in Mexican Pesos (MXN). 1 MXN = 0.107 USD PPP.

Table 2: Covariate Balance

	Control	Treatment	P-value of Difference
Age (Years)	44.73	44.71	0.1604
Monthly Income	13,506.49	13,497.15	0.7030
Tenure (Months)	87.75	80.94	0.3950
Checking Account Balance	19,322.25	19,394.21	0.3629
Ln (Checking Account Balance)	8.02	8.02	0.3180
Credit Card Interest	20.31	20.23	0.2849
Ln(Credit Card Interest)	0.26	0.25	0.3760
Credit Card Balance	3,858.71	3,884.17	0.3526
Ln(Credit Card Balance)	1.32	1.33	0.6653
Credit Card Limit	17,203.11	17,199.28	0.7031
N	357,567	2,696,936	

This table presents a covariate balance test in which we estimate Equation 1 with different dependent variables (as specified in Column 1). Columns 1 and 2 present the average value of each dependent variable for Treatment and Control groups. Column 3 shows the p-value of regressing the corresponding outcome on the treatment indicator with strata fixed effects and robust standard errors. The p-value of an F-test from regressing the treatment indicator on all of the covariates with strata fixed effects is 0.3615. Monthly income and balances are in Mexican Pesos (MXN). 1 MXN = 0.047 USD.

Table 3: Checking and Credit Card Account Balances for Individuals Who Have a Credit Card, By Deciles of Average Daily Balance on Checking Accounts Over Income

	All Clients with Credit Card				*Clients Paying Credit Card Interest*			
Decile	N	Checking Account Balance over Income (Average)	Fraction Of Clients with non-zero Credit Card Balance	Fraction Of Clients Paying Credit Card Interest	N	Checking Account Balances (Average)	Credit Card Balances (Average)	Credit Card Interest (Average)
All	362223	1.81	0.61	0.31	111999	27,818.18	32,929.68	1,120.90
1	36223	0.01	0.62	0.42	15141	340.20	29,917.08	1,018.99
2	36222	0.05	0.56	0.37	13445	1,086.67	24,165.70	854.02
3	36222	0.08	0.59	0.37	13351	2,054.23	26,525.30	956.52
4	36223	0.13	0.61	0.36	13115	3,204.63	27,805.94	1,001.48
5	36222	0.20	0.64	0.35	12546	5,293.93	31,556.76	1,107.03
6	36222	0.33	0.64	0.32	11475	8,467.78	35,507.68	1,215.31
7	36223	0.58	0.63	0.28	10054	15,266.06	38,101.32	1,280.91
8	36222	1.16	0.62	0.24	8757	29,971.89	42,637.44	1,366.57
9	36222	2.81	0.59	0.21	7529	66,548.62	43,713.88	1,381.63
10	36222	12.73	0.58	0.18	6586	295,446.99	45,925.31	1,463.94

This table presents statistics about credit card borrowing and checking account balances for individuals who have a credit card and pay interest, holding different levels of checking account balances over income. Individuals are split into deciles of checking account balances over income. For observations in each decile group, we first present the average of checking account balances over income as well as the fraction of individuals with a non-zero credit card balance and the fraction of individuals paying credit card interest. We then focus on individuals who are paying credit card interest. For them, we present average checking account balances, average credit card balances, and average monthly interest charges. Monthly balances and interest charges are in Mexican Pesos (MXN). 1 MXN = 0.107 USD PPP.

Table 4: Individuals Paying Credit Card Interest With Checking Account Balances Over or Below 50% of Their Income

	No-Puzzle (Less than 50%)	Puzzle (50% or more)	P-value of Difference
Age (Years)	42.72	48.03	0.0000
Monthly Income	19,602.03	21,339.81	0.0000
Tenure (Months)	100.89	134.53	0.0000
Checking Account Balance	29,243.58	65,127.67	0.0000
Credit Card Interest	137.87	515.77	0.0000
Credit Card Balance	19,855.37	44,921.26	0.0000
Credit Card Limit	96,785.91	163,643.28	0.0000
P(Interest_t > 0| Interest_t-1 >0	0.82	0.86	0.0000
N	332,470	29,753	

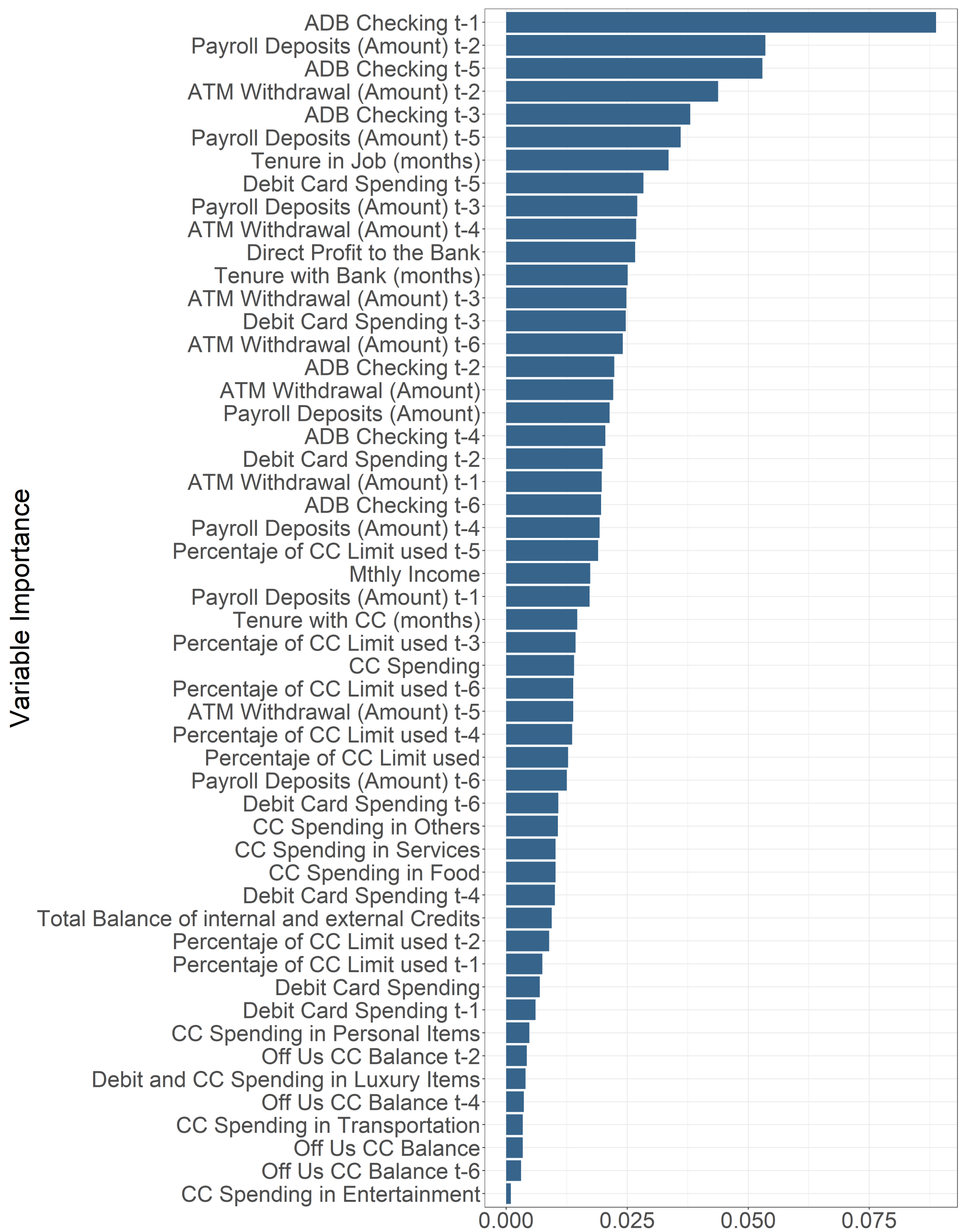
This table presents simple means of each variable for individuals that fall or not into our creditcard debt puzzle definition. We consider all invividuals who have a credit card. We say that an individual falls into the credit card debt puzzle definition if she is paying credit card interest while holding average daily balances in herchecking account that are higher than 50% of her income. Monthly income and balances are in Mexican Pesos (MXN). 1 MXN=0.047 USD. The last column presents the p-value of a t-test for differences in means with robust standard errors.

Table 5: Overall Treatment Effects of the Intervention

	(1)	(2)	(3)	(4)	(5)
	All Individuals			Individuals with a Credit Card	
	Ln Checking Acct. Balance +1	Ln Checking Acct. Balance+1	Ln Checking Acct. Balance+1	Ln Checking Acct. Balance+1	Ln Credit Card Interest+1
Any treatment	0.006*			0.014**	-0.005
	(0.004)			(0.007)	(0.004)
Msg1		0.007			
		(0.005)			
Msg2		0.008*			
		(0.005)			
Msg3		0.006			
		(0.005)			
Msg4		0.006			
		(0.005)			
Msg5		0.002			
		(0.005)			
Msg6		0.007			
		(0.005)			
Msg7		0.006			
		(0.005)			
Bi-weekly			0.006		
			(0.004)		
Weekly			0.007*		
			(0.004)		
Observations	3054503	3054503	3054503	362223	362223
Mean of Dep. Var in Control Group	17393.63	17393.63	17393.63	24331.63	213.84

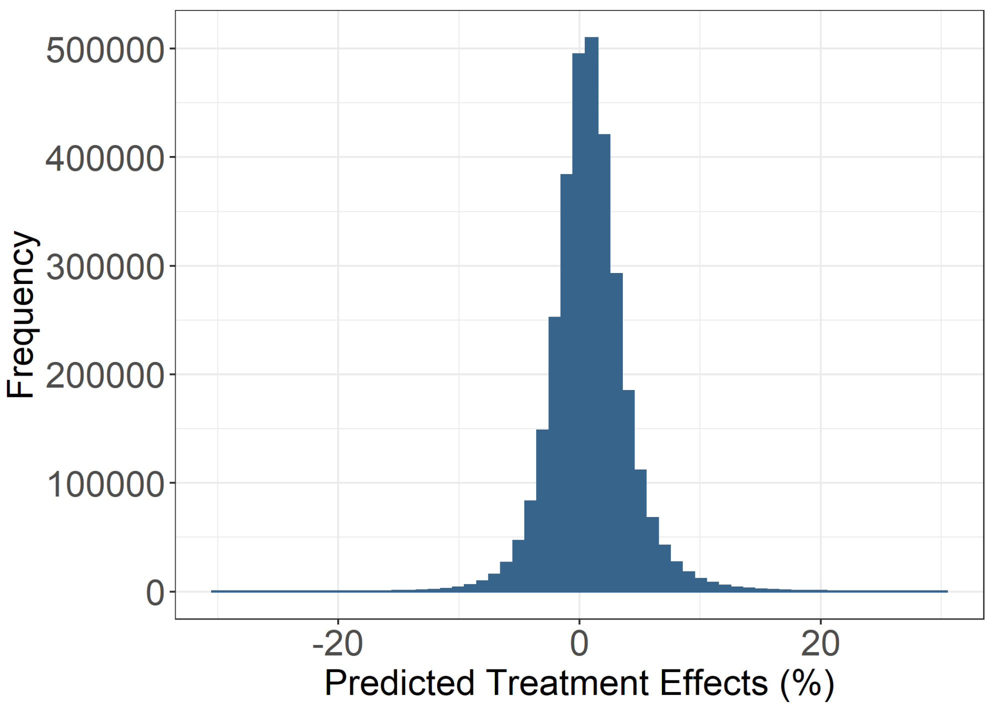
This table presents the results of estimating Equation (1) with different outcomes and with different treatment definitions. Observations are at the user level. Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1: Variable Importance: Causal Forest

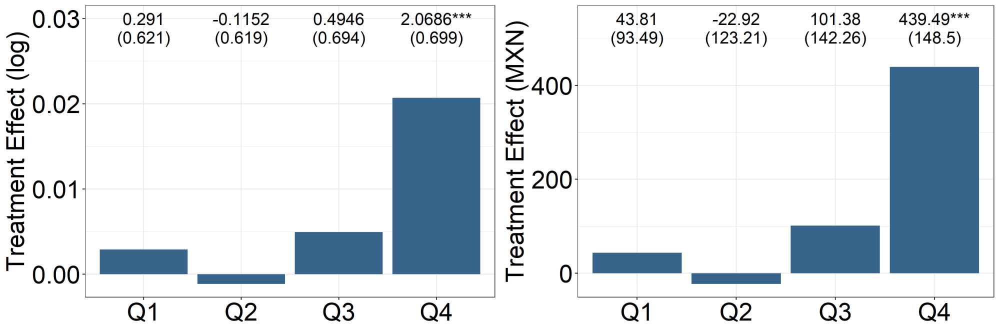


This graph shows the variable importance of the 52 variables used in the estimation of our final causal forest. Following Athey and Wager (2019), we first estimate a pilot causal forest using all available pre-treatment variables (161 variables), and re-estimate a final model using only those with variable importance larger than 1%. The resulting 52 variables are listed in the vertical axis of this Figure. Variable importance indicates how often a variable was used to select splits in the multiple trees of the causal forest. By construction, the variable importance of all variables used in a causal forests add up to one. ADB refers to average daily balances. All variables are monthly.

Figure 2: Distribution of Predicted Treatment Effects



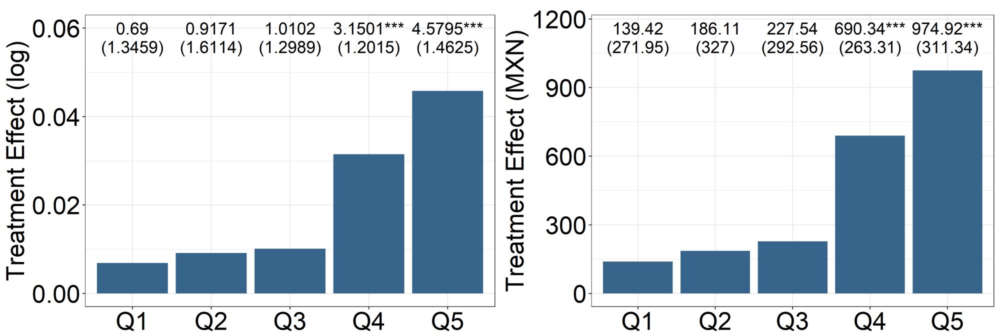
This graph shows the distribution of predicted treatment effects. We estimate a causal forest that predicts for each individual in treatment and control groups an individual treatment effect.



(a) treatment effect (% Change)

(b) treatment effect (MXN)

Figure 3: Treatment effects on checking account balances, as a function of predicted individual treatment effects by the causal forest. Individuals are split in to quartiles of treatment effects on savings, based on the score generated by the causal forest. Average treatment effects are estimated using the natural log of checking account balances, as the dependent variable. In panel b) we calculate the treatment effect in MXN by multiplying the treatment effect in % by the mean of the dependent variable in the control group during the treatment period. The sorting into quartiles is based on cross-fitted rankings over two folds.



(a) Treatment Effect (% Change)

(b) Treatment Effect (MXN)

Figure 4: Treatment effect on checking account balances, as a function of predicted individual treatment effects by the causal forest for individuals in the top quartile of the predicted treatment effect distribution. Individuals in the top quartile of the predicted treatment effect distribution are split in to quintiles of predicted treatment effects, based on the score generated by the causal forest. Average treatment effects are estimated using the natural log of checking account balances, as the dependent variable. In panel b) we calculate the treatment effect in MXN by multiplying the treatment effect in % by the mean of the dependent variable in the control group during the treatment period. The sorting into quartiles is based on cross-fitted rankings over two folds.

Table 6: Treatment Effects on Savings and Credit Card Borrowing

Dep.Var	(1)	(2)	(3)	(4)	(5)	(6)
	Ln Checking Account Balance +1	Ln Credit Card Balance (Banorte) +1	Ln Credit Card Balance (Credit Bureau) +1	Ln Credit Card Interest +1	Paid Interest {0,1}	Ln Credit Card Payments +1
Panel A: All Clients with Credit Cards						
TE	0.0614***	-0.0141	-0.0066	-0.0145	-0.0044	-0.0221
	(0.0137)	(0.0107)	(0.0060)	(0.0353)	(0.0067)	(0.0176)
Mean of Dep. Var in Control Group (MXN)	31,701.61	17,119.74	43,191.72	222.42	0.46	9,472.50
Increase in Savings (MXN)	1,947.94					
Upper Confidence Interval (MXN)		117.15	221.75	12.17	0.0087	117.75
Upper Confidence Interval (MXN) divided by incrase in savings (MXN)		0.0601	0.1138	0.0062	0.000004	0.0604
N= 126571						
Panel B: Clients who Paid Credit Card Interest at Baseline						
TE	0.0557**	-0.0120	-0.0085	-0.0191	-0.0034	-0.0286
	(0.0257)	(0.0095)	(0.0057)	(0.0422)	(0.0097)	(0.0213)
Mean of Dep. Var in Control Group (MXN)	23,244.40	22,945.46	51,401.71	410.38	0.73	7,948.76
Increase in Savings (MXN)	1,294.97					
Upper Confidence Interval (MXN)		150.39	134.82	26.08	0.0155	104.63
Upper Confidence Interval (MXN) divided by incrase in savings (MXN)		0.1161	0.1041	0.0201	0.000011	0.0808
N= 58497						

This table shows treatment effects on a selection of variables related to saving and borrowing behavior. Column 1 shows the treatment effect on ln(Checking Account Balances+1). Columns 2 and 3 show the treatment effect on ln(Credit Card Balances+1) considering only credit cards held at Banorte and all credit cards reported to the credit bureau respectively. Columns 4 and 5 show the treatment effect on ln(Credit Card Interest+1) and a binary variable indicating if an individual is paying interest on her credit card, respectively. Column 6 shows the treatment effect on ln(Credit Card payments+1). In all cases we consider individuals in the top quartile of the predicted treatment effect distribution stemming from the causal forest. Panel A considers all individuals who have a credit card. Panel B considers only individuals who have a credit card and incurred interest at baseline (in the 6 months previous to the intervention). Average treatment effects are calculated with the AIPW method. The increase in savings, expressed in MXN, is calculated by multiplying the TE and the Mean of Checking Account Balances in the Control Group. Upper confidence intervals, expressed in MXN, are calculated as (point estimate + 1.96*Standard Error)*Mean of Dep. Var in Control Group. [1]The upper confidence interval for the probability of incurring credit card interest during the treatment period is expressed in percentage points and not in MXN (point estimate + 1.96*Standard Error). The number of observations is the same across all columns in the same panel. Robust standard errors in parentheses. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Table 7: Treatment Effects On Deposits, ATM Withdrawals, and Spending

	(1)	(2)	(3)
Dep.Var.	Ln Deposits +1	Ln ATM Withdrawals +1	Ln Spending with Credit or Debit Card +1
Panel A: Clients With Credit Card			
TE	-0.0086	-0.0511***	-0.0467***
	(0.0098)	(0.0101)	(0.0107)
Mean of Dep. Var	28184.53	12634.46	15615.62
N=126571			
Panel B: Clients With Credit Card Who Paid Interest At Baseline			
TE	-0.0063	-0.0712***	-0.0394***
	(0.0099)	(0.0167)	(0.0107)
Mean of Dep. Var	23199.13	14008.18	21063.06
N=58947			

This table considers all individuals with credit cards in the top quartile of the distribution of the predicted treatment effect distribution stemming from the causal forest. Deposits, withdrawals, credit card spending, and debit card spending are all monthly. Spending with Credit or Debit Card is defined as the sum of debit or credit card store or online purchases. The number of observations is the same across all columns in the same panel. Interest payments are in Mexican Pesos (MXN). 1 MXN = 0.107 USD PPP. Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 8: Treatment Effects on Savings and Credit Card Borrowing for Individuals for whom Banorte is their Main Bank

Dep.Var	(1)	(2)	(3)	(4)	(5)
	Ln Checking Account Balance +1	Ln Credit Card Balance (Banorte) +1	Ln Credit Card Interest +1	Paid Interest {0,1}	Ln Credit Card Payments +1
Panel A: All Clients with Credit Cards					
TE	0.0607***	-0.0107	-0.0028	-0.0016	-0.0112
	(0.02)	(0.01)	(0.04)	(0.01)	(0.02)
Mean of Dep. Var in Control Group (MXN)	34,395.46	12,884.18	226.60	0.70	10,314.65
Increase in Savings (MXN)	2,087.37				
Upper Confidence Interval (MXN)		208.74	16.16	0.02	209.96
Upper Confidence Interval (MXN) divided by incrase in savings (MXN)		0.1000	0.0077	0.000010	0.1006
N= 89899					
Panel B: Clients who Paid Credit Card Interest at Baseline					
TE	0.0526**	-0.0097	-0.0191	-0.0014	-0.0096
	(0.02)	(0.01)	(0.05)	(0.01)	(0.03)
Mean of Dep. Var in Control Group (MXN)	28,271.85	19,272.32	399.34	0.69	8,888.42
Increase in Savings (MXN)	1,487.98				
Upper Confidence Interval (MXN)		147.26	31.51	0.04	391.59
Upper Confidence Interval (MXN) divided by incrase in savings (MXN)		0.0990	0.0212	0.000030	0.2632
N= 41223					

This table shows treatment effects on a selection of variables related to saving and borrowing behavior for clients for whom Banorte is their main Bank. Column 1 shows the treatment effect on ln(Checking Account Balances+1). Column 2 shows the treatment effect on ln(Credit Card Balances+1) considering all credit cards held at Banorte. Columns 3 and 4 show the treatment effect on ln(Credit Card Interest+1) and a binary variable indicating if an individual is paying interest on her credit card, respectively. Column 6 shows the treatment effect on ln(Credit Card payments+1). In all cases we consider individuals in the top quartile of the predicted treatment effect distribution stemming from the causal forest. Panel A considers all individuals who have a credit card. Panel B considers only individuals who have a credit card and incurred interest at baseline (in the 6 months previous to the intervention). Average treatment effects are calculated with the AIPW method. The increase in savings, expressed in MXN, is calculated by multiplying the TE and the Mean of Checking Account Balances in the Control Group. Upper confidence intervals, expressed in MXN, are calculated as (point estimate + 1.96*Standard Error)*Mean of Dep. Var in Control Group. [1]The upper confidence interval for the probability of incurring credit card interest during the treatment period is expressed in percentage points and not in MXN (point estimate + 1.96*Standard Error). The number of observations is the same across all columns in the same panel. Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Table 9: Heterogeneous Treatment Effects by Experimental Strata

	Dep. Var: Ln (Checking Account Balances +1)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Any Treatment	-0.006	0.009	0.013*	0.006	0.002	0.008*	0.006	0.007*	0.005
	(0.007)	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Any Treatment*Group$_1$	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted	Omitted
Any Treatment*Group$_2$	0.012	0.001	-0.013	0.001	0.002	-0.010	0.000	-0.003	0.009
	(0.01)	(0.01)	(0.01)	(0.007)	(0.007)	(0.009)	(0.010)	(0.010)	(0.007)
Any Treatment*Group$_3$	0.010	0.014	-0.002			-0.001			
	(0.01)	(0.01)	(0.01)			(0.009)			
Any Treatment*Group$_4$	0.024**	0.002	-0.013						
	(0.01)	(0.01)	(0.01)						
Group Definition	Quartiles of Checking Acct. Balance	Quartiles of Income	Quartiles of Age	Median of Tenure with Banorte	Median of ATM Withrawals	Median of Debit Card Transactions	Is Digital?	Main Bank?	Has Credit Card?
Observations	3054503	3054503	3054503	3054503	3054503	3054503	3054503	3054503	3054503

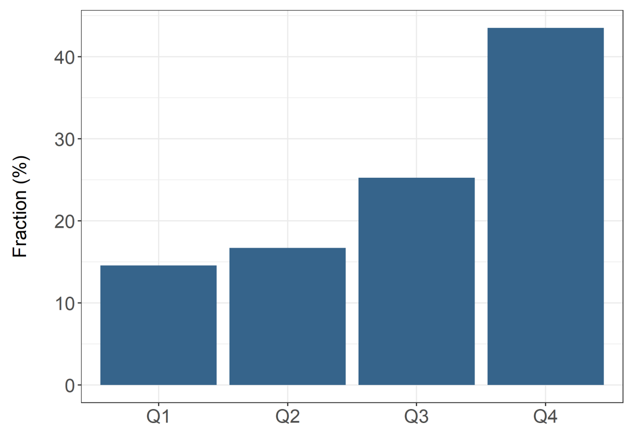
This table presents heterogeneous treatment effects by experimental strata. Treatment effects are estimated in each column with the following OLS regression: $y_i = \alpha_s + Treatment_i + Group_{ij} + Treatment * Group_{ij}$ where $alpha_s$ represents strata fixed effects and $Group_{ij}$ is a dummy variable that takes the value of 1 when individual i belongs to Group j. In each column the groups are defined over a different variable which in turn defines the experimental strata. In all cases we consider all 3.1 million observations at the user level. Robust standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Table 10: Treatment Effects for Users in Groups with the Highest Observed Treatment Effect and for Users with the Highest Individual Treatment Effects Predicted by the Causal Forest

	Top Quartile of Individuals Observed Treatment Effects				Top Quartile of Individuals Individual Treatment Effects Predicted by Causal Forest			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep.Var.	N	Ln Checking Account Balance +1	Ln Credit Card Interest +1	Ln Credit Card Balance (Banorte) +1	N	Ln Checking Account Balance +1	Ln Credit Card Interest +1	Ln Credit Card Balance (Banorte) +1
Panel A: All Clientes	763,511							
ATE		0.2401***	-0.0197***	-0.0142***	763,625	0.0207***	-0.0031	-0.0028
		(0.0072)	(0.0037)	(0.0048)		(0.0070)	(0.0059)	(0.0057)
Mean of Dep. Var (MXN)		18283.47	66.66	4161.45		21245.03		
Panel B: Clients with Credit Card	126,468				126,571			
ATE		0.4403***	-0.0991***	-0.1089***		0.0614***	-0.0145	-0.0141
		(0.0148)	(0.0095)	(0.0083)		(0.0137)	(0.0353)	(0.0107)
Mean of Dep. Var (MXN)								
		21623.82	241.41	15077.12		31701.61	222.42	17119.74
Panel C: Clients with Credit Card Who Paid Interest at Baseline	61,204				58,947			
ATE		0.5167***	-0.1109***	-0.1946***		0.0557**	-0.0191	-0.0120
		(0.0114)	(0.0094)	(0.0092)		(0.0257)	(0.0422)	(0.0095)
Mean of Dep. Var (MXN)		14994.75	410.8639	19585.27		23244.40	410.38	22945.46

This table shows treatment effects on a selection of variables related to saving and borrowing behavior, for clients in groups with the highest ex-post observed treatment effects or for clients with the highest individual treatment effects predicted by the causal forest. For columns 1 to 4, we split the sample into 6,104 mutually exclusive groups defined by the interaction of all experimental strata. For each group we calculate treatment effects. We then assign to each observation in the group the treatment effect of its group. We then split the sample into quartiles based on the treatment effect assigned to each observation. The top quartile corresponds to the 25% of observations, which belong to strata blocks with the highest observed treatment effect. For them, we calculate treatment effects on checking account balances, credit card interest, and credit card balances regressing the corresponding outcome variable on a treatment indicator and strata-block fixed effects. We do the same in columns 5 to 8 but for the top quartile of individuals with the highest individual treatment effects as predicted by the causal forest. Robust standard errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Figure 5: Distribution of the Puzzle Group by Quartiles of Predicted Treatment Effects



This graph presents the distribution of individuals in the puzzle group, across quartiles of predicted treatment effects. The puzzle group is defined as the set of individuals who carry checking account balances of at least 50% of their income, and also pay credit card interest. Predicted treatment effects are calculated with the causal forest.

A Appendix

A.1 Causal Forests and The Generalized Random Forest Algorithm

Causal forests are based on causal trees, and their relation is analogous to the relation between widely known random forests and regression trees. Regression trees predict an individual outcome Y_i using the mean Y of observations that share similar covariates, X. To define what counts as similar, regression trees partition the covariate space into disjoint groups of observations called 'leaves.' Within each leaf, all observations share values (or belong to the same value interval) of certain Xs. A tree starts with a training sample that is treated first as a single group and then recursively partitioned. For each value $X_j = x$, the algorithm forms candidate splits, placing all observations with $X_j \leq x$ in a left leaf and all observations with $X_j > x$ in a right leaf. The split is implemented if it minimizes a certain loss criterion, such as mean squared error ($\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$). This criterion is evaluated in the sample, that is, the same observations used to define where to split are also used to calculate the mean value of the outcome in each leaf. The algorithm then repeats the process for each of the two new leaves and so on until it reaches a stopping rule. Using the final set of leaves, the tree provides out-of-sample predictions by figuring out in which terminal leaf a certain observation falls based on its covariate values and assigning a predicted value equal to the average value of all observations in that leaf in the training sample.

Random forests are an ensemble of n trees in which n random subsamples of the data are taken and each subsample is used to train a causal tree. Predictions for each observation in a test sample (which could be the full original dataset) are defined as the average across n predictions, obtained by pushing that one observation down each of the n trees.

In contrast to regular random forests that predict individual outcomes Y_i, causal forests want to predict conditional treatment effects ($E[Y_1 - Y_0 | X = x]$ in a potential outcomes framework), to measure how causal effects vary for different sub-populations. Standard loss criteria such as goodness-of-fit measures are not available because we do not observe the treatment effect $Y_1 - Y_0$ for any one individual. Athey and Imbens (2016) show that maximizing the expected mean squared error of predicted treatment effects instead of the infeasible mean squared error itself is basically equivalent to maximizing the variance of treatment effects across leaves. Thus, this defines a new criterion for sample splitting that is specifically designed to identify treatment effect heterogeneity. They also show that, to reduce overfitting bias, the training sample should be further split into a splitting and an estimation sample so that the observations used to choose where to create new leaves are not the same ones used to calculate treatment effects within each leaf. In addition, Athey et al. (2019) argue for the importance of orthogonalization: in other words, the treatment effect estimation in the next sample (the estimation sample) has to balance covariates between the treatment and control groups. Causal forests are different from off-the-shelf machine learning methods in three ways:

First, in addition to dividing the data into training and test samples, causal forests divide the training data further in two sub-samples: a splitting sample and an estimation sample. The splitting sample is used to grow trees (2,000 in our case) and the estimation sample is used to estimate the treatment effects. This honesty is crucial for causal forests to attain consistent estimation of treatment effects, and similar strategies are implemented in other recently developed methods for causal inference with machine learning (Chernozhukov et al., 2018).

Second, causal forests use a splitting rule that tackles treatment effect heterogeneity directly. This is, each tree splits into two children nodes where heterogeneity in treatment effects is maximized. Thus, causal forests are tailored to find sub-populations with different treatment effects.

Finally, causal forests calculate treatment effects ensuring that the treatment indicator is orthogonal to all covariates for all observations. The algorithm computes estimates of propensity scores and outcomes for treatment and control group by training separate regression forests. Then the algorithm performs sample splits to identify heterogeneous treatment effects on residual treatments and outcomes. To calculate the treatment effect on a sub-population of interest, the algorithm plugs the individual predictions of the causal forest into an Augmented Inverse Probability Weighting Estimate (AIPW) that combines models of outcome regressions with models of treatment propensity to estimate causal effects.[12]

We use the generalized random forest (grf) package in R, to estimate our causal forests. This package allows for estimation of causal forests, but also allows for estimation of other forest-based methods for causal inference. To do so efficiently, this package involves an approximate gradient based loss criterion (instead of the exact loss criterion described above to build intuition), aggregates the results of the n trees with one single weighted estimation of treatment effect, instead of averaging n estimations of treatment effects. The mechanics of the algorithm is as follows:

1. The first step is to compute estimates of propensity scores for the treatment and marginal outcomes conditional on covariates, by training separate regression forests and performing predictions (fitted values) for each observation. These predictions are used to calculate residuals, which will be referred to as orthogonalized outcomes and orthogonalized treatment status.

2. For each tree, a random subsample with 50% of the database is drawn (training sample).

3. The trainning sample is further split into a splitting subsample and an estimation subsample (50-50 by default).

[12]This estimator is locally efficient and is known as a “doubly robust estimator" since it is consistent whenever the model of treatment propensity *or* the model of expected outcomes are correctly specified.

4. A single initial root node is created for the splitting sample, and child nodes are split recursively to form a tree. Each node is split using the following algorithm:

   (a) A random subset of variables are selected as candidates to split on. [13]

   (b) For each of these variables, we look at all of its possible values and consider splitting into two child nodes based on a measure of goodness of split, determined to maximize the heterogeneity in treatment effect estimates across nodes.

   (c) All observations with values for the split variable that are less than or equal to the split value are placed in a new left child, and all examples with values greater than the split value are placed in a right child node.

5. The estimation sample is used to populate the leaf nodes of the tree. Each observation is 'pushed down' the tree, and assigned to the leaf in which it falls.

6. Steps 2 to 5 are repeated 2,000 times, that is we estimate 2,000 trees.

7. Treatment effects are predicted for each observation on a test dataset (potentially the full dataset) as follows:

   (a) Each test observation is pushed down each tree to determine what leaf it falls in. Given this information, a list with neighboring observations in each tree leaf is created (the neighbors come from the estimation sample of each tree). Each neighbor observation is weighted by how many times it fell in the same leaf as the test observation.

   (b) Treatment effects are calculated using orthogonalized outcomes and treatment status of the neighbor observations.

8. In addition to personalized treatment effects, the package allows for estimation of treatment effects across all observations in a dataset, or arbitrary subsamples of it. This is done with an AIPW estimator, that ensures balance across all covariates in the group, using the treatment propensities estimated in step 1.

A.1.1 Calibration Test

We formally test for whether heterogeneity in individual predictions is associated with heterogeneity in treatment effects using the "calibration test" based on Chernozhukov et al. (2018), as implemented in the grf package of R, and described in Athey and Wager (2019). This test seeks to

[13] By default $min\{sqrt(p) + 20, p\}$ variables are sampled, where p is the total number of variables in the dataset. In our analysis, $p = 161$ the first time we run the algorithm, and $p = 52$ the second time we run the algorithm, and we use 32 or 27 candidate variables in each split.

fit conditional treatment effects as a linear function of the causal estimates of the causal forest and computes the best linear fit of the treatment effects using the forest prediction as well as the mean forest prediction as the sole two regressors. A coefficient of 1 for 'mean.forest.prediction' suggests that the mean forest prediction is correct, The p-value of the 'differential.forest.prediction' coefficient acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity. Table A1 shows the results of the calibration test. We find that the coefficient measuring the ability of the forest to predict heterogeneities in treatment effects is positive and significant. We conclude that the individual level treatment effect predictions are a valid linear predictor for heterogenous treatment effects: larger predicted treatment effects (score value) indeed result in larger treatment effects.

Table A1: Calibration Test for Evaluation Of The Quality Of The Causal Forest

	estimate	std.error	t-statistic	p.value
mean.forest.prediction	1.0286	0.3732	2.7564	0.0029
differential.forest.prediction	0.3470	0.1280	2.7132	0.0033

This table presents the result of a calibration test, based on Chernozhukov et al. (2018), as implemented by the grf package in R. This test computes the best linear fit of the target estimand using the forest prediction as well as the mean forest prediction as the sole two regressors. A coefficient of 1 for 'mean.forest.prediction' suggests that the mean forest prediction is correct. The p-value of the 'differential.forest.prediction' coefficient also acts as an omnibus test for the presence of heterogeneity: If the coefficient is significantly greater than 0, then we can reject the null of no heterogeneity.

To understand the differences between individuals who respond to nudges and those who don't, we can compare the descriptive statistics of individuals in the top and bottom quartiles of the predicted treatment effect distribution. Note that, by design, we would not expect these to be balanced. The comparison can be found in Table A2.

To make sure our results are not driven by the selection of a specific functional form (log+1), Table A3 presents the impact of the intervention of borrowing and saving outcomes measured in Mexican pesos.

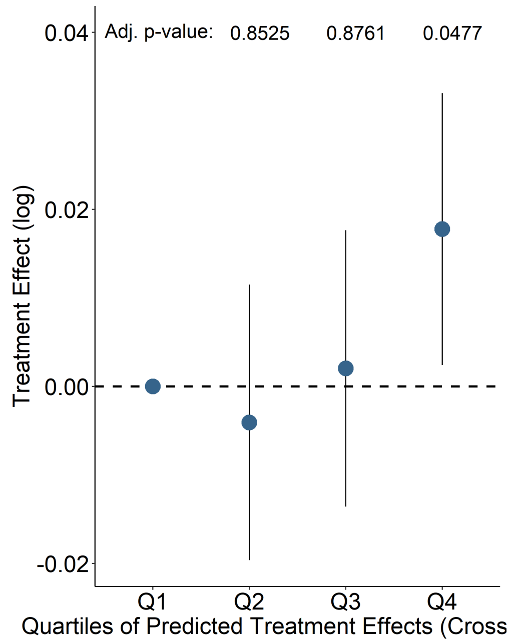


Figure A6: This Figure shows the differences in treatment effects across quintiles of predicted treatment effects, for observations in the top quartile of predicted treatment effects. P-values are adjusted for for multiple hypothesis testing with correction of Romano and Wolf (2005)

.

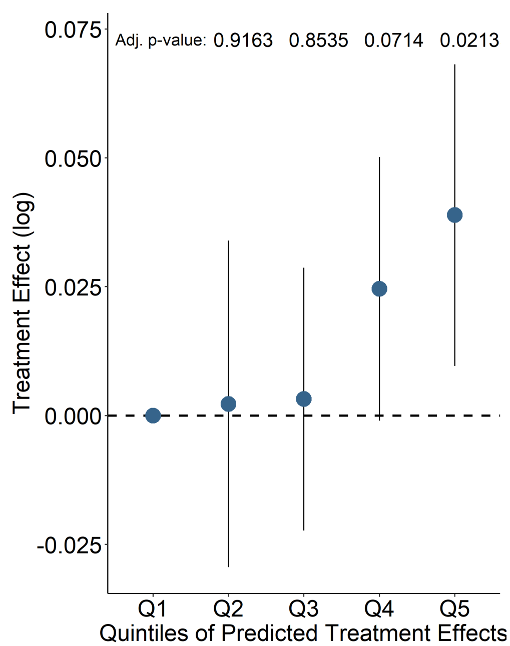


Figure A7: This Figure shows the differences in treatment effects across quintiles of predicted treatment effects, for observations in the top quartile of predicted treatment effects. P-values are adjusted for for multiple hypothesis testing with the correction of Romano and Wolf (2005).

Table A2: Differences Between Top and Bottom Quartiles of the Distribution of Predicted Treatment Effects

	Bottom 25%	Top 25%	P-value of Difference
Age (Years)	44.18	46.35	0.0054
Monthly Income	14,118.44	15,109.87	0.0000
Tenure (Months)	74.60	88.69	0.0000
Checking Account Balance	16,017.05	21,338.30	0.0000
Credit Card Balance	2,435.53	6,038.65	0.0000
Credit Card Limit	10,812.16	29,933.66	0.0000

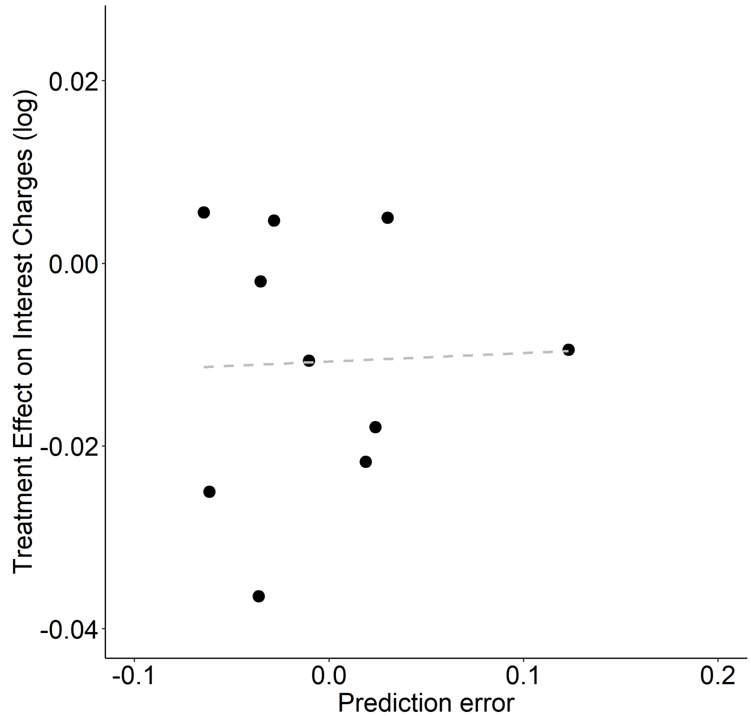
This table presents simple means of each variable for individuals that fall or not into the top and bottom quartiles of the distribution of predicted treatment effects. 1 MXN=0.107 USD PPP. The last column presents the p-value of a t-test for differences in means with robust standard errors.

Table A3: Treatment Effects on Savings and Credit Card Borrowing (MXN)

Dep.Var	(1)	(2)	(3)	(4)	(5)
	Checking Account Balance	Credit Card Balance (Banorte)	Credit Card Balance (Credit Bureau)	Credit Card Interest	Credit Card Payments
Panel A: All Clients with Credit Cards					
TE	2,109.66***	-140.32	-229.71	-3.39	151.76
	(727.47)	(190.94)	(303.22)	(9.46)	(187.36)
Mean of Dep. Var in Control Group (MXN)	31,681.46	17,097.99	43,136.75	230.39	9,500.24
Upper Confidence Interval (MXN)		233.92	364.60	15.15	518.99
Upper Confidence Interval (MXN) divided by incrase in savings (MXN)		0.1109	0.1728	0.0072	0.2460
N= 126571					
Panel B: Clients who Paid Credit Card Interest at Baseline					
TE	1,790.54***	-214.46	-311.77	-6.13	-87.37
	(594.86)	(289.81)	(439.11)	(19.77)	(260.83)
Mean of Dep. Var in Control Group (MXN)	23,194.21	23,080.11	51,491.24	413.31	8,012.99
Upper Confidence Interval (MXN)		353.57	548.89	32.62	423.86
Upper Confidence Interval (MXN) divided by incrase in savings (MXN)		0.1975	0.3065	0.0182	0.2367
N= 58947					

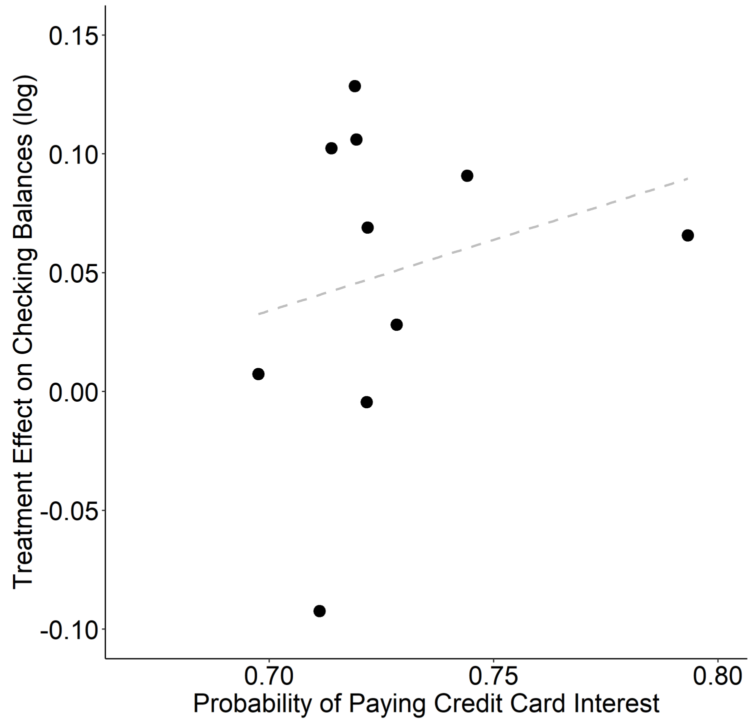
This table shows treatment effects on a selection of variables related to saving and borrowing behavior. All dependent variables are expressed in MXN pesos. Column 1 shows the treatment effect on Checking Account Balances. Columns 2 and 3 show the treatment effect on Credit Card Balances considering only credit cards held at Banorte and all credit cards reported to the credit bureau respectively. Column 4 show the treatment effect on Credit Card Interest. Column 6 shows the treatment effect on Credit Card payments. In all cases we consider individuals in the top quartile of the predicted treatment effect distribution stemming from the causal forest. Panel A considers all individuals who have a credit card. Panel B considers only individuals who have a credit card and incurred interest at baseline (in the 6 months previous to the intervention). Average treatment effects are calculated with the AIPW method. The increase in savings, expressed in MXN, is calculated by multiplying the TE and the Mean of Checking Account Balances in the Control Group. Upper confidence intervals, expressed in MXN, are calculated as point estimate + 1.96*Standard Error. The number of observations is the same across all columns in the same panel. Robust standard errors in parentheses. *p<0.1; **p<0.05; ***p<0.01.

Figure A8: Correlation between Prediction Errors and Treatment Effects on Borrowing



This graph shows the correlation between prediction errors and treatment effects on credit card interest. Prediction errors are defined as the difference between the simple average of individual-level predicted treatment effects, and the actual average treatment effect of observations in each group, as estimated with the AIPW method. Predicted treatment effects are winsorized at the 1 and 99 percentiles. The analysis considers observations in the top 25% of predicted treatment effects, which are further split into deciles. Observations are ranked with a cross-fitted procedure over two folds.

Figure A9: Correlation between the Fraction of Individuals Paying Credit Card Interest and the Treatment Effect of the Intervention on Checking Account Balances



This graph shows the correlation between the fraction of individuals paying credit card interest during the treatment period and the treatment effect of the intervention on checking account balances. Treatment effects are calculated with the AIPW method. The analysis considers observations in the top 25% of predicted treatment effects, which are further split into deciles. Observations are ranked with a cross-fitted procedure over two folds.

A.2 Saving and Borrowing by Treatment Message

To explore the relation between borrowing and savings across each of the seven messages included in the experiment, we focus on the 126,458 individuals in the top quartile of predicted treatment effects who had a credit card. For them, we calculate the treatment effect on saving and borrowing of receiving each specific treatment message.

Table A4: Treatment Effects on Saving and Credit Card Borrowing: Individuals in the Top Quartile of Predicted Treatment Effects who Have a Credit Card

	(1)	(2)	(3)	(4)	(5)	(6)
	Ln Checking Account Balance +1	Increase in Savings (MXN)	Ln Credit Card Interest +1	Upper Confidence Interval of Credit Card Interest (MXN)	Upper Confidence Interval for Interest Charges Divided by Increase in Savings	N
All messages	0.0601*** (0.0177)	1904.37	-0.0171 (0.0336)	11.12	0.006	126458
Msg 1	0.0265 (0.0228)	839.56	-0.0055 (0.0336)	13.90	0.017	38802
Msg 2	0.1170*** (0.0228)	3705.46	-0.0183 (0.0336)	10.96	0.003	38775
Msg 3	0.0413* (0.0228)	1306.86	-0.0142 (0.0336)	11.90	0.009	38822
Msg 4	0.0979*** (0.0229)	3102.57	-0.0256 (0.0339)	9.41	0.003	38700
Msg 5	0.0623*** (0.0237)	1974.71	-0.0348 (0.0350)	7.79	0.004	38803
Msg 6	0.0338 (0.0253)	1069.25	-0.0291 (0.0374)	10.20	0.010	38752
Msg 7	0.042 (0.0298)	1330.94	0.008 (0.0440)	21.72	0.016	38590

This table shows treatment effects of each individual message on a selection of variables related to saving and borrowing behavior. Column 1 shows the treatment effect on ln(Checking Account Balances+1). Column 3 shows the treatment effect on ln(Credit Card Interest+1). In all cases we consider individuals in the top quartile of the predicted treatment effect distribution stemming from the causal forest who had a credit card. Treatment effects are calculated with the AIPW method for each message separately. The increase in savings, expressed in MXN, is calculated by multiplying the TE and the Mean of Checking Account Balances in the Control Group (31,681.46 MXN). Upper confidence intervals, expressed in MXN, are calculated as (point estimate + 1.96*Standard Error)*Mean of Dep. Var in Control Group. The Mean of Dep. Var in Control Group for credit card interest is 213.39 MXN. Standard errors in parenthesis. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

A.3 Comparison of Sorting Methods for Heterogeneous Treatment Effects

We first estimate the treatment effect on saving and borrowing outcomes for individuals in the top quartile of pre-treatment checking account balances who have a credit card.

Table A5: Treatment Effects on Saving and Borrowing for Individuals in the Top Quartile of Pre-Treatment Checking Account Balances Who Have a Credit Card

	(1) Ln (Checking Account Balance +1)	(2) Ln (Credit Card Interest +1)
Any Treatment	0.014	-0.012
	(0.009)	(0.008)
N	118,706	118,706
Mean of dependent variable (MXN)	67791.11	184.23

Treatment effects are estimated with equation 1. We consider observations in the top quartile of pre-treatment checking account balances, who have a credit card. $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$.

Next, we compare the overlap between observations assigned to quartiles of predicted individual treatment effects calculated with the causal forest, and observations assigned to quartiles of the observed treatment effects, calculated for each strata block. In Table A6, the rows represent quartiles based on observed treatment effect for each strata block. The columns represent quartiles of individual treatment effects predicted by the causal forest. A perfect overlap would have all observations across the diagonal. We can see that is not the case: out of the 763,625 observations assigned by the causal forest to the top quartile of predicted treatment effects, only 201,992 are in strata blocks on the top quartile of observed treatment effects.

Table A6: Distribution of Observations According to the Treatment Effect of Strata Blocks and Predicted Treatment Effects at the Individual Level

Rows: Sorting Based on Observed treatment effects

Columns: Sorting Based on Predicted Individual Treatment Effects

	1	2	3	4	Total
1	186854	184315	191453	203924	766546
2	201534	175485	185114	199223	761356
3	193851	199564	202513	167162	763090
4	181387	204262	184546	193316	763511
Total	763,626	763,626	763,626	763,625	3,054,503

This table shows the distribution of observations according to the observed treatment effect of their strata blocks, and their individual predicted treatment effect, as returned by the causal forest. The rows represent quartiles based on observed treatment effect for each strata block. For them we split the sample into 6,104 mutually exclusive groups defined by the interaction of all experimental strata. For each group we calculate treatment effects, and we assign to each observation in the group the treatment effect of its group. We then split the sample into quartiles based on the treatment effect assigned to each observation. The columns represent quartiles of individual treatment effects as predicted by the causal forest. For each observation, the causal forest returns a predicted treatment effect, which we split into quartiles. The across rows and columns adds up to the 3,054,503 observations included in the analysis. We can see that there is poor overlap with these two sorting methods. For example, the predictions of the top quartile according to the causal forest are split across strata groups in all four quartiles of observed treatment effects, and viceversa.

A.4 Two Toy Models Illustrating the Predictions of Rational versus Behavioral Theories of the Co-Holding Puzzle

We now outline two toy models to rationalize the co-holding puzzle. The first is based on Telyukova (2013) and Kaplan and Violante (2014) and rationalizes co-holding with transaction convenience constraints. The second model rationalizes co-holding with mental accounting and self-control problems and is based on the theories in Laibson et al. (2007), Haliassos and Reiter (2005), and Bertaut et al. (2009).

Transaction-convenience model:

We assume a simple model with two periods, one consumption good, and log utility. Individuals receive an endowment x_1 in period 1 and consume $c_{1,2}$ in periods 1 and 2. In addition, they must hold a certain amount of cash x for transaction purposes $x_1 - c_1 > x$, and they may borrow b_1 in period 1 for additional consumption. Additionally, we assume that the agent discounts future utility by a factor δ.

$$max\{log(c_1 + b_1) + \delta log(x_1 - c_1 - (1 + r)b_1)\}$$

subject to $x_1 - c_1 > x$ and $b_1 < b$.

Suppose $r = 0$ and $b = \infty$, then the optimal solution for c_1^* is:

$$c_1^* = \frac{1}{\delta + 1}x_1 \text{ and } b_1 = 0 \text{ if } x_1 - c_1^* \geqq x$$

$$\text{and if } x_1 - c_1^* < x \text{ then } c_1^* = \frac{1}{\delta + 1}x_1 \text{ and } b_1 = c_1^* + x - x_1.$$

In this model, the amount of cash x held for transaction-convenience reasons is set aside for consumption in circumstances in which cash is required. Individuals take into account that this cash is available for future consumption and individuals thus borrow against that amount. From the equations, it is clear that if we increase the amount of cash x held for transaction-convenience reasons, that is, by encouraging individuals to save, we increase borrowing b_1 in the rational model.

We note that the assumption of r=0 is only for simplicity, but is not required to lead to the result that savings increases borrowing. The intuition carries forward with r>0, and the results are available from the authors upon request.

Self-control model:

We start from the same setting as in the transaction-convenience model but instead of having a transaction-convenience constraint, we assume that when individuals hold a certain amount of cash dedicated for savings, x this amount goes to a separate (non-fungible) mental account. Therefore, x gets subtracted from the original endowment x_1 available for consumption, and does not enter the consumption decision of the agent more than as an exogenous constraint in the available resources. The role of mental accounting thus, is to remove a certain amount of money labeled as savings from the optimization problem, in a manner equivalent to a wealth shock. As an alternative interpretation, we can think of an amount of money, x, that one spouses hides from the other, or that the planner-self is successfully able to remove from spender-self decision problem. As before, the role of the spouse is to remove a certain amount of money labeled as savings from the optimization problem. In addition, we assume that the agent is impatient: that is, discounts future utility by an additional factor β.

$$max\{log(c_1 + b_1) + \beta\delta log(x_1 - x - c_1 - (1 + r)b_1)\}$$

subject to $b_1 < b$. Suppose $r = 0$ and $b = \infty$, then the optimal solution for c_1^* is:

$$c_1^* = \frac{1}{\beta\delta + 1}(x_1 - x) \text{ and } b_1 = 0 \text{ (independent of } x).$$

From the equation above we can see that, if we increase the amount of money that the saver self/spouse hides from the spender self/spouse, x, we decrease c_1 but nothing happens to borrowing b_1.

As before, we note that the assumption of r=0 is only for simplicity, but is not required to lead to the result that savings increases borrowing. The intuition carries forward with r>0, and the results are available from the authors upon request.