Does Saving Cause Borrowing?

Paolina Medina* and Michaela Pagel†

March 12, 2021

Abstract

We study 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. Using Machine Learning methods for causal inference, we build a score to sort individuals according to their predicted treatment effect. We then focus on the individuals in the top quartile of the distribution of predicted treatment effects who have a credit card and were paying interest at baseline. Relative to their control, this group increased their savings by 5.7% on average or 61.84 USD per month. At the same time, we can rule out increases in credit card interest larger than 1.25 USD with 95% statistical confidence. We thus estimate that for every additional dollar of savings, individuals incur less than 2 cents in additional borrowing cost. This is a direct test test of the predictions of rational co-holding models, and is an important result to evaluate policy proposals to increase savings via nudges or more forceful measures.

Keywords: savings nudges, credit card borrowing, heterogeneous treatment effects, causal forest.

*Mays Business School of Texas A&M University. E-mail: pmedina@tamu.edu
†Columbia GSB, NBER, and CEPR. E-mail: mpagel@columbia.edu

We thank seminar and conference participants at Australian National University (ANU), Shanghai Advanced Institute of Finance, SJTU (SAIF), and Lone Star Conference. We also thank Nhan Le, James Choi, and Han Xia for providing valuable comments. We thank Jose Antonio Murillo for providing support and access to the data, Isidoro Garcia for helping with the logistics of the experiment and preliminary analysis, and Sofía Terrazas for outstanding research assistance. At the time of the experiment, Medina was a full time employee of Banorte. IRB ID:IRB2018-0597.

1 Introduction

The 2019 American Household Credit Card Debt Study estimates the total revolving credit card debt owed by an average US household to be 7,104 USD, which amounts to a total of 466 billion USD. Such large high-interest debt holdings over longer periods of time are very hard to rationalize in standard economic models. For example, Laibson et al. (2003) argue that such debt holdings constitute "a debt 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. Kaplan and Violante (2014) provide an explanation for the amount of credit that we see in the US based full rationality: credit card borrowing is a response to illiquid savings in retirement accounts or other assets, i.e., in the event of transitory income shocks, individuals cannot access their illiquid savings and thus use high-interest credit instead. To date, however, limited evidence exist on whether or not individuals respond with borrowing when they are nudged or forced to save. Clearly, however, this question is of central importance for policy-makers and researchers alike to evaluate whether savings nudges or forced savings should be implemented.

In the 2001 US Survey of Consumer Finances (SCF), 27% of households reported revolving an average of 5,766 USD in credit card debt, with an APR of 14%, and simultaneously, holding an average of 7,338 USD in liquid assets, with a return of around 1% (Telyukova, 2013). This simultaneous holding of savings and consumer credit is known as the "credit card debt or co-holding puzzle." 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) provides an explanation for this behavior by arguing that households need cash for transaction purposes, and they optimally choose to finance their cash holdings with credit card debt. Instead Haliassos and Reiter (2005) argue that individuals want to constrain their impatient selves or spouses by keeping them indebted while simultaneously building their savings. As we show theoretically, a key distinctive prediction between rational co-holding models and behavioral models is that in the former case, increases in savings would cause increases in debt (since households optimally finance their liquidity needs using credit card debt). In contrast, behavioral co-holding predicts that increases in savings should not be reflected in borrowing as the two are distinct mental accounts. In this paper, we test whether indeed additional savings causes additional borrowing to distinguish between these two theories for co-holding credit card debt and savings.

We empirically evaluate and quantify whether or not savings nudges that are followed by actual increases in savings also increase high-interest unsecured borrowing in general, and for individuals that were already paying credit card interest. This question is important to understand the origins of the credit card debt puzzle and to evaluate policy proposals designed to increase savings. To do so we use a large-scale field experiment paired with comprehensive and very accurate panel data of individual credit cards and checking accounts by one of the largest banks in Mexico, Banorte. The bank ran a randomized experiment with 3,054,438 customers out of which 374,893 customers were randomly selected to be in a control group. Clients in the treatment group received ATM and SMS messages that suggested them to save and had been proven impactful in previous experiments ran by the same bank. The intervention lasted 7 weeks from September 13 to October 27, 2019.

To focus the analysis on individuals that indeed responded to the savings nudge, we use machine learning techniques to predict individual treatment effects and study the behavior of individuals with the largest predicted treatment effects. Specifically, we estimate a causal forest model as discussed in Athey and Imbens (2015), Hitsch and Misra (2018), and Athey et al. (2019). The algorithm recursively estimates heterogeneous treatment effects for different sub-populations and then predicts for each individual an estimated average treatment effect, using a rich set of pre-treatment covariates. In turn, we focus on the subsample of customers that are in the top quartile of the predicted treatment effect distribution and have a credit card. For this group of individuals, we ask whether the increased savings were accompanied by an increase in borrowing.

It is important to note that this approach does not suffer from overfitting, that could lead us to incorrectly infer large treatment effects for arbitrary sub-populations. Searching over many possible partitions and estimating treatment effects for each of them in the same sample would be problematic: after all, in any one sample observations sharing certain covariates could exhibit a larger treatment effect due to idiosyncratic shocks and not because those covariates carry a larger real treatment effect. In contrast, the predicted treatment score from the random forest is calculated using 2,000 repeated sample splits to figure out which pre-treatment covariates predict a large response to the savings treatment (holding all pre-treatment observable characteristics constant across treatment and control through explicit orthogonalization). This procedure eliminates the possibility that pre-treatment covariates predict a large treatment effect only by chance. In turn, we are confident that the the subpopulations with the largest predicted treatment effects have indeed a large response in savings, and we can thus look at borrowing as another outcome variable for the subgroup of individuals with the largest predicted treatment scores for savings.

More specifically, we estimate a causal forest with 2,000 causal trees. Each tree is built with a random subsample of the full data set. This random subsample is further split in to two samples: a splitting sample, and an estimation sample. The splitting sample is used to identify splitting rules where the estimated treatment effect differs the most. Then, the rules obtained from the splitting sample are used to calculate treatment effects in the estimation sample. To ensure unconfoundedness, treatment effects are calculated with an Augmented Inverse Probability Weighted (AIPW) estimator– the AIPW estimator ensures that characteristics of each sub-population are balanced between treatment and control groups. This process is repeated for each tree, and information from all 2,000 trees is aggregated to provide a unique prediction for each observation in the sample. The algorithm gives a consistent estimate of heterogeneous treatment effects. The multiple sampling method rules out that by chance people with some characteristics just ended up having higher savings during that period (because they would not be in all 2,000 random samples and would therefore not be consistently showing large effects but only sometimes).

We thus estimate the responses in saving and borrowing for the top quartile of predicted treatment effect individuals that have a credit card to then also look at their borrowing. For this population, the increase in savings estimate is 6.01% on a baseline savings of 31,681 MXN in their control group (1,489 USD), i.e., an increase of 1,904 MXN (89 USD).[1] On average, this group of individuals decreased their interest payments by 1.71% from a basis of 230 MXN with a standard error of 3.34%. We can thus rule out an increase in borrowing cost of more than 11 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 11/1,904 increase in borrowing or a 0.006% increase in borrowing in response to a

[1] Over our sample period, 1 MXN was corresponding to 0.047 USD on average. A rough estimate for the USD value can thus be obtained by subtracting one decimal point and dividing by 2.

1% increase in savings. For the group of individuals that also paid credit card interest at baseline, we have an increase in savings of 5.67% on a baseline value of 23,080 MXN, i.e., 1,316 MXN (62 USD). In turn, we can rule out a 6.64% increase in credit card borrowing with 95% confidence. This equals an increase of 26.68 MXN in borrowing cost. To conclude, for every 1 MXN in savings we can thus rule out a 27/1,409 or larger than 1.9 cents increase in credit card borrowing. We find however that individuals that were carrying substantial levels of credit card debt respond to the savings nudge by increasing their liquid savings.

We also 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. Instead, we propose mental accounting and rules of thumb as a potential explanation, following Haliassos and Reiter (2005). The theoretical idea is the following: individuals have a spending account, i.e., their credit card, as well as an account for savings. They separate these two accounts mentally to cope with their overspending and self control problems. The reason is that individuals can maintain a rule to not touch their savings but have trouble to restrict their credit card borrowing. On their credit cards, they will spend up to some personal limit. Once they get close to that personal limit, they feel constrained and can restrict their overspending more successfully. If individuals would take their savings and repay their credit card debt, they 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 optimally decide to hold the two positions simultaneously.

To provide further evidence for this psychological mechanism behind the co-holding puzzle, we show that those individuals that co-hold, defined as holding more than 50% of their income in their checking accounts and paying credit card interest, overlap most strongly with the highest quartile of the predicted savings score, i.e., the co-holding individuals are also most susceptible to the savings nudge without increasing their credit card borrowing in response.

To evaluate rational and behavioral theories behind the credit card debt puzzle, and to understand whether or not we should induce households to save more, we nudge individuals to save more and study whether they respond with an increase in credit card borrowing. In summary, the answer is No. Our findings are consistent with individuals choosing to hold credit card debt and savings simultaneously to help them cope with limited self-control.

2 Literature review

Our paper is related to a large literature on the savings effects of automatic (as opposed to opt-in) enrollment into 401(k) savings plans. This literature generally finds that a 1% increase in default savings rates increases total savings by 0.5% to 0.8% (see, e.g., Choi et al., 2004; Chetty et al., 2014). Implicitly, this research assumes that individuals do not offset the increased savings with additional borrowing. To the best of our knowledge, the only research paper evaluating whether nudges to save increase borrowing is Beshears et al. (2019). The authors look at a natural experiment in which the US army started to automatically enroll newly hired employees into their retirement savings plan. In response, employees saved more and borrowed about 1% of their income more in secured credit such as auto loans and first-time mortgages. The measure of credit card borrowing in this paper are biannual snapshots of balances from a credit bureau. However, a biannual snapshot of credit card balances does not reveal how much high-interest unsecured debt is actually rolled over. In our study, we can instead look at the high-frequency responses in credit card borrowing using bank account transactions and balances. Additionally, we see whether individuals roll over debt in the first place and can ensure that individuals would have the ability to borrow as we observe credit limits. Similarly, the literature on savings nudges via SMS or fintech apps (Karlan et al., 2016; Gargano and Rossi, 2020; Akbaş et al., 2016; Rodríguez and Saavedra, 2015), also focuses on savings outcomes. This literature has documented effects of varying magnitude, but in general they don't study potential increases in borrowing.

Previous literature on the credit card debt puzzle began with Gross and Souleles (2002), who document the phenomenon and note that transaction demand for liquidity may contribute to it. Maki (2002) study 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. Bertaut et al. (2009) study whether households may hold liquidity and credit card debt simultaneously as a means of self-(or spouse) control. If one spouse in the household is the earner, and the other is the compulsive shopper, it is argued that the earner will choose not to pay off credit card debt in full in order to leave less of the credit line open for the shopper to spend. Several others have proposed a number of explanations related to the credit card debt puzzle, including self-control and financial literacy (Gathergood and Weber, 2014), or mental accounting (Gathergood and Olafsson, 2020).

Laibson et al. (2012) examine a related puzzle: the coexistence in household portfolios of credit card debt and retirement assets. The authors explain this behavior 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 not consume them. As mentioned, 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. However, strictly speaking, these two explanations cannot apply to the credit card debt puzzle. The key difference is that retirement assets involve a significant penalty for early withdrawal, i.e., they are not liquid in contrast to savings accounts. That said, analyzing whether liquid savings results in borrowing should provide us with a lower bound for the borrowing response to illiquid savings.

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 occur 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 relaxation of liquidity constraints. Soman and Cheema (2002) present experimental and survey evidence that consumers interpret available credit lines as indications of future earnings potential when deciding consumption expenditures.

3 Background on the Mexican credit card market

As of June 2017 in Mexico, there were 17.9 million general-purpose credit card accounts in good standing holding a positive balance, in a population of 124 million. The credit card market has expanded rapidly as in 2009 only 13 million cards were in circulation. In spite of these trends, credit card penetration in Mexico has remained small 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 card holder 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.[2] 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 percentage points above the federal short-term interest rate, which was 7.17%.

The credit card market in Mexico is fairly concentrated, similar to the US (Herkenhoff and Raveendranathan, 2020). 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 sixth 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.[3]

[2] INEGI, Encuesta Nacional de Inclusion Financiera, 2018.
[3] Refer to Banco de Mexico, multiple reports.

4 Experimental Design and Data Description

We analyze the results of a large-scale experiment to promote savings with the Mexican bank *Banorte*. The experimental pool consists of 3,054,438 customers, out of which 374,893 customers were randomly selected to be in a control group. Clients in the control group received no message. Clients in the treatment group were randomly assigned to receive 1 of 7 messages that have been proven to be effective in previous experiments nudging individuals to save. Half of the treated customers were cross-randomized to receive the messages on a weekly basis, while the other have were assigned a bi-weekly frequency.[4] The intervention lasted 7 weeks from September 13 to October 27, 2019.

For each customer in the experimental pool, we observe all information routinely collected by the bank, including balances on checking accounts and credit cards, information from the credit bureau, income and other demographic characteristics.

The treatment messages were the following:

> **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

[4] Users in the treatment group were further cross randomized across two additional dimensions: First, half of them would stop receiving the messages for two weeks, after 2 months of receiving, and then resume. Second, half of the consumers in the treatment group would receive the same message through 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.

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 in \$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 on your Banorte Account and see your wealth grow!"

Message 8: "Be preparaed for an emergency! Commit to leaving 10% more in your account. Don't withraw all your money on payday."

Table 1 shows descriptive statistics for all individuals, i.e., all 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 (635 USD) and the clients have banked with the bank for 7 years on average.[6] In turn, their checking account balance is approximately 19,384 MXN. About 30% of credit card holders pay credit card interest.

Beyond showing these descriptive statistics for all individuals we also show them separately for those individuals 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 (1,030 USD). The average individual pays with a credit card pays 169 MXN (8 USD) in interest costs per month, noting

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

[6] Over our sample period, 1 MXN was corresponding to 0.047 USD on average. A rough estimate for the USD value can thus be obtained by subtracting one decimal point and dividing by 2.

that this average includes individuals who do not pay any interest. Individuals also have substantial borrowing capacity on their cards, 102,278 MXN on average.

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. 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 this age difference is not an economically meaningful difference.

In terms of the credit card debt puzzle, Table 3 shows by deciles of savings over income the fraction of individuals that pay credit card interest and their balances on checking accounts, credit cards, and interest payments. We here restrict to only individuals who have a credit card. We can see that 20% to 30% of individuals that have a credit card pay credit card interest even when they are in the higher deciles of checking account balances. This is the population that we are concerned about: individuals with both savings and credit card debt. 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 (60 USD). Note that Banorte's average credit card interest is 35.2%, and the return on checking accounts is 0%.

We now look at all individuals rolling over credit card debt and define the savings and credit card debt puzzle population as individuals holding more than 50% of their income in their checking account 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. In turn, Table 4 compares individuals in the puzzle group, to the rest of individuals who pay credit card interest but are not in the puzzle group. 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. Both populations appear to hold debt persistently as there is a high correlation between rolling over debt in any given month and doing so in the previous month. While credit cards in the first place and co-holding are not as common in our overall population relative to the US, the size of our experiment will provide sufficient statistical power to analyze this subpopulation.

5 Methodology

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

Then, we use machine learning techniques to predict individual treatment effects. Specifically, we estimate a causal forest as discussed in Athey and Imbens (2016), Hitsch and Misra (2018), and Athey et al. (2019).

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. Thus, the interaction coefficients identifies the incremental effect of the treatment on individuals above the median age. If there are several potential explanatory variables, the dimensionality of the model grows significantly, since one would need to interact all variables of interest with each other and with the treatment. Then researchers run the risk of overfitting or capturing heterogeneous treatment effects by chance, i.e., 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 or multiple hypothesis testing problems. This method is tailored for efficiently predicting causal effects of a treatment for a rich set of different sub-populations through three distinctive features that will be discussed below: sample splitting, orthogonalization, and an optimization method designed to capture treatment effect heterogeneity.

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 X's. 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 sample, i.e., 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 last set of leaves, the tree provides out-of-sample predictions by figuring out in which terminal leaf a certain observation falls in, 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 the n predictions, obtainined 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 average 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. And thus define a new criterion for sample splitting specifically designed to identify treatment effect heterogeneity. They further 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 used to calculate treatment effects within each leaf. In addition, Athey et al. (2019) argue for the importance of orthogonalization, i.e., the treatment effect estimation in the estimation sample has to balance covariates between treatment and control groups.

Thus, causal forests are different than off-the-shelve machine learning methods in three ways: 1) they estimate treatment effects with a repeated split sample method (referred to by Athey and Imbens (2016) as "honest estimation"), 2) they use a splitting rule for the trees that aims to directly find sub-populations with different treatment effects, instead of predicting levels of the outcome of interest in treatment and control groups separately, and 3) they use orthogonalization methods to ensure covariate balance across multiple subpopulations.

First, in addition to dividing data in training and validation 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 average treatment effect on a subpopulation 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.[7]

We use the generalized random forest 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), 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.

[7]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.

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).

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. [8]

   (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, i.e. we estimate 2,000 trees.

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

[8] 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.

(a) Each test observation is pushed down each tree to determine what leaf it falls in. Given this information, a list with neighboring observation 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 average 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.

In turn, we can look at the frequency distribution of those individual treatment effects and identify the sub-population with the largest predicted treatment effects on savings. For them, we will study the borrowing consequences of saving by looking at average treatment effects on savings and on credit card outcomes.

6 Results

6.1 Aggregate Effects of the Intervention

We study the treatment effect of the intervention on savings for the entire experimental pool, as well as the treatment effect on savings 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 treatment and control groups.

Table 5 shows the average treatment effects across all treatments, by treatment message and by treatment frequency. Column 1 shows that on average there is a 0.6% increase in savings, from a basis of 21,867 MXN. Column 2 shows that if we break out the effect by treatment message, we can see that only Message 2 has a positive and small treatment effect. Column 3 shows that only the treatment with weekly messages has a positive treatment effect. The treatment that skips one week and send messages on a bi-weekly basis does not lead to a significant effect. However, all treatment messages and frequencies have similar coefficients, and they are not statistically different from each other.

Columns 4 and 5 show the average treatment effect for individuals who have a credit card. We pool all treatments into one single dummy variable that takes the value of one if a given individual was assigned to any of the treatments. Here we find a 1.4% increase in savings, from a basis of 24,331 MXN, which represents an increase in savings of 340 MXN. While this is a small increase in savings, we nevertheless explore if there is any increase in credit card interest payments and do not find a significant effect.

Note that these average treatment effects are intention-to-treat (ITT) effects because individuals may or may not have seen the messages and then choose how much to respond. The fact that we find a positive and significant effect in a randomized setting effect implies that at least some individuals saw the message and their behavior was affected by it.

Consistent with previous literature on saving nudges via SMS, the impact is relatively small (Karlan et al., 2016). The fact that credit card holders have a stronger effect on savings suggests that there may be some sub-populations with a stronger response than others. We thus study treatment effects heterogeneities in the following section.

6.2 Heterogeneous Effects and Subpopulation 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 strong heterogeneities, with some individuals responding strongly, while others remain unaffected. Our setting allows us to characterize sub-populations who indeed respond to saving nudges, and provides insights of how to perform targeted interventions. Second, any meaningful test of the effect of saving-nudges on borrowing requires nudges to first have a strong effect on nudges. Testing the impact of saving nudges on borrowing, where there is a small effect on savings would be of limited use. In contrast, testing the effect of saving nudges on borrowing for individuals who indeed responded to the saving nudge and experience meaningful increases on saving is relevant both for policy, and for testing theoretical explanations of the credit card debt puzzle. Our heterogeneity analysis allows us to study the borrowing response of individuals who indeed increase their savings significantly as a result of a savings nudge. To identify individuals with the highest response to the treatment we use a causal forest. Section 7 contrast the results of the causal forest with a manual exploration based on choosing the strata (or strata blocks) with the largest observed average treatment effects.

A causal forest produces individual predictions of treatment effects for each observation in the sample (both treatment and control groups). Following Athey and Wager (2019) we first train a pilot causal forest with 2,000 trees using all 161 pre-treatment variables available for the analysis. This variables include past financial behavior (for example, for checking and credit card balances and interest we include 6 monthly lags), demographic variables, and a number of geographic dummies. We then train a second forest only on the 52 variables with the higher importance, i.e. those who saw the largest number of splits in the first estimation. For this second causal forest estimation, Figure 2 shows the 27 variables with the highest variable importance, and Figure 1 shows the distribution of the predicted treatment effects at the individual level, listing the 52 final pre-treatment variables in the caption. This will be the basis for our subsequent analysis.

Figure 1 shows the distribution of pre-treatment covariates. We use individual predictions as a "score" value that ranks observations according to their predicted treatment effects (Chernozhukov et al., 2018), and we split the sample of users according to each individual's score value into quartiles and calculate treatment effects on savings. Figures 4 and 3 show how the treatment effects on savings is larger for individuals with larger scores, suggesting that predicted treatment effects are a valid score for actual treatment effects. We can see that the top 5% of individuals in the sample have a treatment effect of 5.33% or 1,162.5 pesos. Appendix 1 provides a formal test for the validity of individual treatment effects as a score for actual treatment effects (Chernozhukov et al., 2018).

Table 6 compares the baseline characteristics of individuals in the top and bottom quartiles of the distribution of predicted treatment effects. Compared to individuals in the bottom quartile of the distribution of predicted treatment effects, individuals with the highest predicted response are about one year older, have higher income, larger tenure with the bank, larger checking account balances, as well as larger credit card balances and credit card limits.

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 in the top quartile of predicted treatment effects, we are capturing a relavant fraction of the puzzle population. This also speaks to the idea that co-holding is a psychological mechanism to exercise self control, which also makes individuals more susceptible to savings nudges.

We now focus on individuals in the top quartile of predicted treatment effects who have a credit card. For them, we calculate the treatment effect on borrowing and saving. 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. We do not search 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" or overfitting i.e., we would pick a group of individuals that in one single sample displayed large savings in response to the treatment even when something else might be going on with this sample thus resulting on a treatment effect that is larger than the real treatment effect. Instead, our predictions are based on 2,000 causal trees, each trained with a different sample which is further split in to a splitting sample and an estimation sample. Heuristically, individuals in the top quartile of predicted treatment effects are those who consistently showed a high treatment effects across the multiple training samples.

Furthermore, since the top quartile of predicted treatment effects is an arbitrary sample cut, from the perspective of the experimental design, there may not be covariate balance. Therefore, instead of calculating average treatment effects with a simple regression of treatment status on the outcome, we adjusts our treatment effect estimates by treatment propensity or covariate imbalance using a variation of the adjusted inverse probabilty 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 is true on average, perfect covariate balance is not necessarily present across all partitions of the sample. AIPW effectively controls for these imabalances.

Table 7 shows the average treatment effect on borrowing an saving for individuals who are in the top quartile of predicted treatment effect and who have a credit card . Panel A considers all individuals who have a credit card, while Panel B focuses on the subset of individuals who are 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 that have a credit card. Here, the increase in savings estimate is 6.01% on a baseline savings of 31,681 MXN, i.e., 1,904 MXN. On average, this group of individuals decreased their credit card balances by 1.55% from a basis of 17,097 MXN and a standard error of 1.16% as can be seen in column (2). We can thus rule out an increase in borrowing of more than 124 MXN with 95% statistical confidence. Similarly, in column (4) we can see that interest payments decreased by 1.71% from a basis of 230 MXN with a standard error of 3.34% as can be seen in Column (3). We can thus rule out an increase in borrowing costs of more than 11 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 124/1,904 or 11/1,904 increase in borrowing or borrowing cost respectively. In other words, we can rule out a 0.06% increase in borrowing or 0.01% increase in borrowing cost in response to a 1% 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% increase 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 months whereas for Banorte credit cards we use the average daily balances. These results indicate that individuals do not borrow using other cards instead of their Banorte credit card. Furthermore, comparing our Banorte results for credit balances versus actual rolled-over debt, we conclude that the balances are indicative of the actual rolled-over debt.

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 0.68 percentage points on a baseline probability of 42%. Thus, for every 1 MXN in savings, the increase in the likelihood to borrow is only 0.0068/1,904 or 0.0000036 percentage points.

Finally, in Column (6) we report results for credit card payments, i.e., whe individuals repay their outstanding credit card balances or rolled over credit card debt. Here, we also document a very small and tightly estimated treatment effect.

We now turn to the results in Panel B of Table **??**, which correspond to individuals that pay credit card interest at baseline and are in the top quartile of predicted treatment effects. For this group, we have an increase in savings of 5.67% on a baseline of 23,194 MXN, i.e., 1,315 MXN. In turn, we can rule out an increase of 133.97 MXN in credit card borrowing or 26.68 MXN in borrowing cost. To conclude, for every 1 MXN in savings we can thus rule out increases larger than 10 cents (134/1,315) or 2 cents (27/1,315) in credit card borrowing and borrowing costs respectively.

Table 8 shows the increases in savings and borrowing for five quintiles of the treatment effect score for the group of individuals that have a credit card. To be clear, the table conditions on the top quartile of predicted treatment effect for savings and then further splits the sample into quintiles. Additionally, the table shows the respective increase in borrowing costs and the likelihood to borrow. As we can see, for all predicted treatment effect quintiles, the increases in borrowing are very small. Table 9 shows the same for individuals with a credit card that pay interest at baseline.

Figure 6 shows in a graph the treatment effect on interest charges for consumers with credit cards and separately for those consumers who pay interest at baseline. We can see that the negative effect is concentrated in the first quintile of predicted savings effect but all quintiles' estimates are insignificant and small.

We want to know whether individuals increased their saving without increasing their borrowing by decreasing their spending or increasing their income. Table 10 shows the treatment effects on deposits, ATM withdrawals, and spending for the top quartile of predicted savings scores. We can see that the treatment effect appears to work through a 6.0%

decrease in monthly ATM withdrawals and a slightly smaller but still significant 4.2% decrease in card spending. This is true for all individuals with a credit card and also the subset of those paying credit card interest. We thus conclude that a decrease in spending, in particular, discretionary spending that may be financed by cash, was responsible for the increase in savings.

Finally, we also replicate the analysis for individuals for whom Banorte is likely to be their main bank. We say that Banorte is likely to be the main bank of a given individual when the following three conditions are satisfied: she receives her payroll on a Banorte payroll account (defined as such by the regulation), she has a credit card with Banorte, and she has not credits (of any type) outside of Banorte, according to credit bureau records. Table 11 shows borrowing and saving results for this group. Panel A shows results for all clients in the top quartile of predicted treatment effects 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 10 cents in credit card balances, or 1 cent in borrowing cost, for every additional peso saved as a result of the saving nude. 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 11 cents in credit card balances, or 2 cents in borrowing cost for every additional peso saved.

7 Analysis of Other Methods to Identify Subpopulations with Large Treatment Effects

Our preferred method to study heterogeneous treatment effects is a causal forest (Athey et al., 2019). This method, based on orthogonalization and sample splitting, allows us to derive valid inferences for the treatment effect of the intervention across different subpopulations, and to identify the subpopulation with the largest treatment effect without concerns for overfitting. We contrast this method with an exploration of heterogeneous treatment effects based on randomization strata. A standard way to study heterogeneous treatment effects is to split the sample based on strata from the experimental design.

Table 12 shows average treatment effects on savings across experimental strata. We find limited heterogeneity across the subpopulations pre-selected for heterogeneity analysis before the experiment was ran. 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).

This is a useful approach to estimate how a treatment affects a subpopulation of interest that is specified before the experiment takes place. However, this method is inappropriate, when trying to identify the subpopulation with the largest treatment effects. Comparing treatment effects across experimental strata is inefficient when searching for the group with largest effects because it is based on very coarse partitions on the covariate space. Furthermore, attempts to perform more granular partitions without adjusting for overfitting (as causal forst do) would lead to substantial bias. In the following, we exemplify these pitfalls.

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. For them, Table 13 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. Here we see that on average individuals in the top quartile of pre-treatment chicking account balances have a large a significant response to the savings nudge, individuals with a credit card who have pre-treatment checking account balances in the top quartile do not show a statistically significant increase in savings.

To identify a subpopulation with a larger treatment effect, one could further split the sample of individuals in the top quartile of pre-treatment checking account balances, for example overlaying strata dimensiones and ultimately calculate treatment effects for each strata block. 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 have 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 over fitting.

Specifically, we calculate split the sample into 6,104 non-empty mutually exclusive groups defined by the interaction of all experimental strata. For each group we calculate average treatment effects, and we assign to each observation in the group the average treatment effect of its group. We then split the sample into quartiles based on the average treatment effect assigned to each observation. The top quartile corresponds to the 25% of observations which belong to strata blocks with the highest observed average 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 14. Column 1 shows the number of observations included in this section of the analysis. Columns 2 to 4 show the treatment effect for individuals in strata blocks with the largest observed average 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. Similarly, this 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. This estimates, which are free of over fitting bias, are significantly smaller than in columns 2-4. The large overestimation we find, is consistent with the discussion of Abadie et al. (2018) who also finds that sample splitting reduces substantial bias in the context of endogenous stratification.

We then compare the overlap between observations assigned to quartiles of predicted individuals treatment effects calculated with the causal forest, and observations assigned to quartiles of oberserved average treatment effects, calculated for each strata block. In Table 15, the rows represent quartiles based on observed average 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 average treatment effects.

We thus conclude that causal forests, or more general, double machine learning algorithm, are the appropriate methods to identify sub-populations with the largest treatment effects. And we use as our prefered method for identifying responsive subpopulations in this analysis.

8 A Toy Model 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 behavioral preferences 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 consume in periods 1 and 2, $c_{1,2}$, and in period 1 they may borrow to consume b_1 because they must hold a certain amount of cash for transaction purposes $x_1 - c_1 > x$. 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:

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

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

In conclusion, if we increase the amount of cash held for transaction-convenience reasons, i.e., x, by encouraging individuals to save, we increase borrowing b_1 in the rational model.

Self-control model:

We have 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, they consider that money locked away for future consumption. Therefore, the money gets subtracted from x_1, the state variable the agent then uses as his or her decision variable. As an alternative interpretation, we can think of an amount of money, x, that one spouses hides from the other, or that the saver self is successfully able to constrain the spender self by locking away x without that entering the saver selfs decision problem. Additionally, we assume that the agent is impatient: i.e., discounts future utility by an additional factor β.

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

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

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

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

and if $x > 0$ then $c_1^* = \dfrac{1}{\beta\delta + 1}(x_1 - x)$ and $b_1 = 0$ (independent of x).

In conclusion, 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.

9 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 their 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. This is an important result to evaluate policy proposals to increase savings via nudges or more forceful measures.

Our results help us to understand the mechanism behind the so-called credit card debt puzzle, i.e., when individuals hold credit card debt and savings simultaneously. We find that individuals who paid credit card interest at baseline also responded to the savings nudge with significant increases in liquid savings. Nevertheless, this increase in savings was not accompanied by meaninguful increases in credit card debt. A null increase in savings is inconsistent with the predictions of rational explanations of the credit card debt puzzle based on liquidity needs. We argue that this result is consistent with the idea that individuals hold savings and credit card debt simultaneously to deal with self control problems via mental accounting, i.e., they maintain a rule to not touch their savings but are simultaneously indebted due to overspending.

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.

Athey, S. and G. W. Imbens (2015). Machine learning methods for estimating heterogeneous causal effects. *stat 1050*(5), 1–26.

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*.

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 W. L. Skimmyhorn (2019). Borrowing to save? the impact of automatic enrollment on debt. 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.

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.

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.

Herkenhoff, K. F. and G. Raveendranathan (2020). Who bears the welfare costs of monopoly? the case of the credit card industry. Technical report, National Bureau of Economic Research.

Hitsch, G. J. and S. Misra (2018). Heterogeneous treatment effects and optimal targeting policy evaluation. *Available at SSRN 3111957*.

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.

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).

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.

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

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

Income, balances, and interest payments are in Mexican Pesos (MXN). 1 MXN = 0.047 USD. For each individual, we consider information from the 6 months previous to the intervention.

Table 2: Covariate Balance

Variable	Control	Treatment	Difference
Age (Years)	44.73	44.72	-0.01
			(0.01)
Monthly Income ($)	13,506.49	13,498.98	-7.51
			(19.71)
Tenure (months)	81.75	81.66	-0.08
			(0.1)
Checking Acct. Balance ($)	19,322.25	19,392.22	69.98
			(76.91)
Credit Card Balance ($)	3,858.71	3,882.64	23.94
			(25.76)
Credit Card Limit ($)	17,203.11	17,164.27	-38.84
			(101.91)

Income and balances, are in Mexican Pesos (MXN). 1 MXN = 0.047 USD. We test for covariate balance estimating Equation 1 with different dependent variables. Columns 1 and 2 present the average value of each dependent variable for Treatment and Control groups, adjusting to reflect only differences within strata, and to reflect the average in the experimental pool. The adjusted average for the Control group is defined as the α such that $\bar{y} = \alpha + \beta\bar{x}$, where β is the coefficient of the treatment indicator estimated within strata using Equation 1. The adjusted average for the Treatment group is defined as $\alpha + \beta$. Column 3 shows the coefficient of the treatment indicator estimated within strata i.e. β. The p-value of an F-test from regressing the treatment indicator on all of the covariates with strata fixed effects is 0.1519.

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

Balances and interest payments are in Mexican Pesos (MXN). 1 MXN = 0.047 USD.

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

Variable	No-Puzzle (Less than 50%)	Puzzle (50% or more)	Difference
Age (Years)	42.72	48.03	5.32
			(0.08)
Monthly Income ($)	19,602.03	21,339.81	1737.78
			(112.84)
Tenure (months)	100.89	134.53	33.64
			(0.44)
Checking Acct. Balance ($)	29,243.58	65,127.67	35884.1
			(423.32)
Credit Card Balance ($)	19,855.37	44,921.26	25065.89
			(205.6)
Credit Card Limit ($)	96,785.91	,163,643.28	66857.37
			(823.46)
$P(Interest_t > 0 \| Interest_{t-1} > 0)$	0.82	0.86	0.03
			(0.0014)

Income and balances, are in Mexican Pesos (MXN). 1 MXN = 0.047 USD. The probability of incurring credit card interest, conditional on incurring credit card interest on the previous period is calculated with monthly information corresponding to the 6 months previous to the intervention, and with standard errors clustered at the user level.

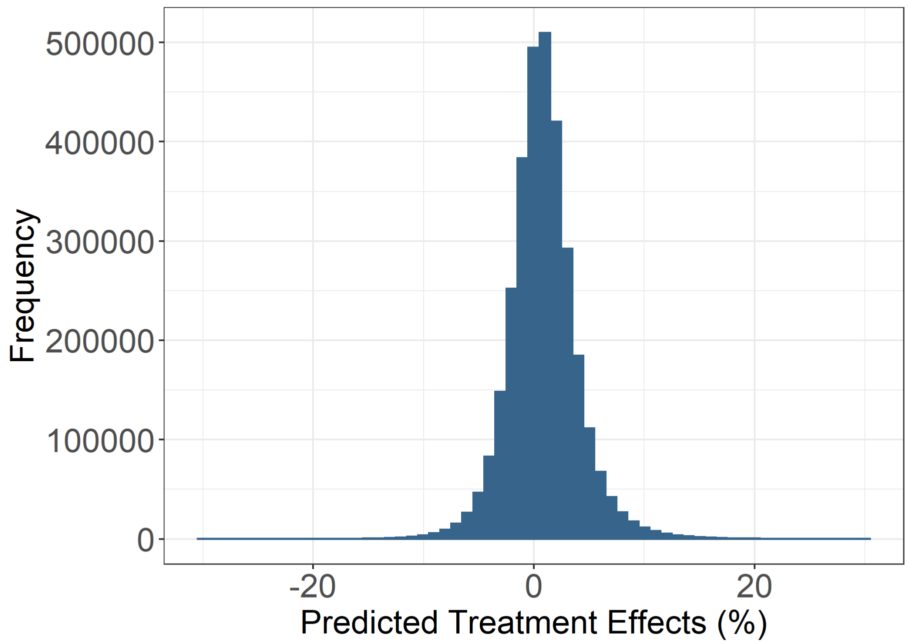
Table 5: Overall Treatment Effect of the Intervention

	All Individuals			Individuals with a Credit Card	
	(1)	(2)	(3)	(4)	(5)
	Log of Checking Acct. Balance +1	Log of Checking Acct. Balance +1	Log of Checking Acct. Balance +1	Log of Checking Acct. Balance +1	Log of 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

Robust Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Distribution of Predicted Treatment Effects.



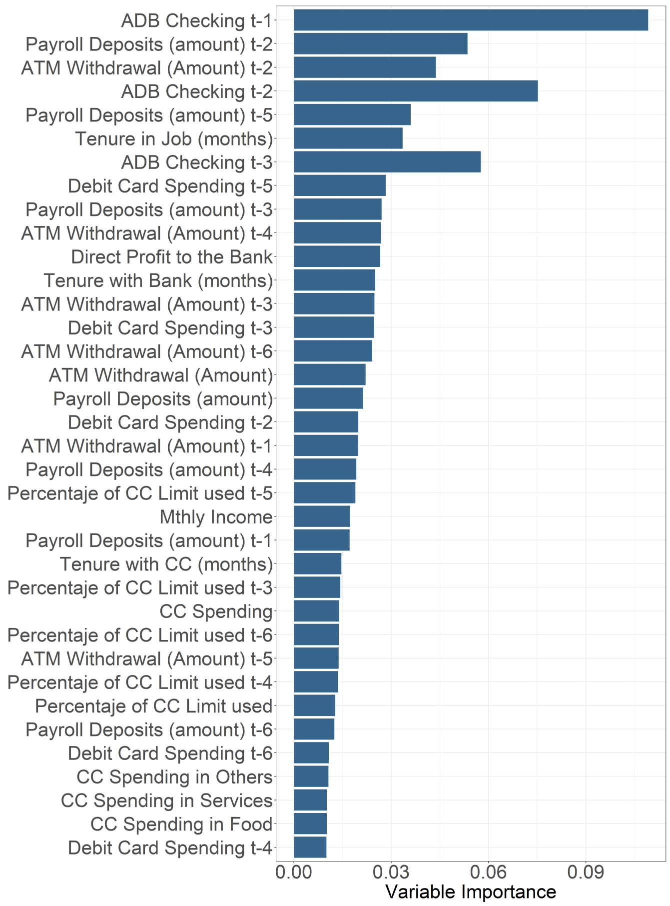
Distribution of Predicted Treatment Effects. We estimate a causal forest that predicts for each individual in treatment and control groups an individual treatment effect. We first estimate the causal forest using 161 pre-treatment variables and then restrict to the 52 most important ones in the second estimation (results shown here). The 52 variables are: ADB Checking t-1, Payroll Deposits (amount) t-2, ADB Checking t-2, ATM Withdrawal (Amount) t-2, ADB Checking t-3, Payroll Deposits (amount) t-5, Tenure in Job (months), Debit Card Spending t-5, Payroll Deposits (amount) t-3, ATM Withdrawal (Amount) t-4, Direct Profit to the Bank, Tenure with Bank (months), ATM Withdrawal (Amount) t-3, Debit Card Spending t-3, ATM Withdrawal (Amount) t-6, ADB Checking t-2, ATM Withdrawal (Amount), Payroll Deposits (amount), ADB Checking t-1, Debit Card Spending t-2, ATM Withdrawal (Amount) t-1, ADB Checking t-3, Payroll Deposits (amount) t-4, Percentage of CC Limit used t-5, Mthly Income, Payroll Deposits (amount) t-1, Tenure with CC (months), Percentage of CC Limit used t-3, CC Spending, Percentage of CC Limit used t-6, ATM Withdrawal (Amount) t-5, Percentage of CC Limit used t-4, Percentage of CC Limit used, Payroll Deposits (amount) t-6, Debit Card Spending t-6, CC Spending in Others, CC Spending in Services, CC Spending in Food, Debit Card Spending t-4, Total Balance of internal and external Credits, Percentage of CC Limit used t-2, Percentage of CC Limit used t-1, Debit Card Spending, Debit Card Spending t-1, CC Spending in Personal Items, Non-Banorte CC Balance t-2, Debit and CC Spending in Luxury Items, Non-Banorte CC Balance t-4, CC Spending in Transportation, Non-Banorte CC Balance, Non-Banorte CC Balance t-6, and CC Spending in Entertainment. ADB refers to average daily balances, all variables are monthly. Stratifying variables are also included to predict probability of treatment in all cases.

Table 6: Differences Between Top and Bottom Quartiles of the Distribution of Predicted Treatment Effects

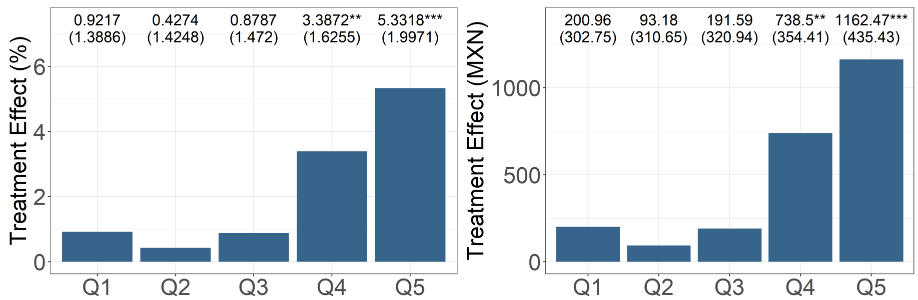
Variable	Bottom 25%	Top 25%	Difference
Age (Years)	43.92	45.28	1.37
			(0.03)
Monthly Income ($)	12,924.95	14,655.87	1730.93
			(23.45)
Tenure (months)	73.95	87.14	13.19
			(0.12)
Checking Acct. Balance ($)	15,791.01	21,340.95	5549.94
			(84.40)
Credit Card Balance ($)	2,688.76	6,391.2	3702.43
			(29.36)
Credit Card Limit ($)	10,402.82	28,641.07	18238.25
			(117.17)

Income and balances, are in Mexican Pesos (MXN). 1 MXN = 0.047 USD. This Table presetopnts simple means for individuals in the top and bottom 25% of the distribution of predicted topment effects on the log of checking account balances.

Figure 2: Variable Importance: Causal Forest

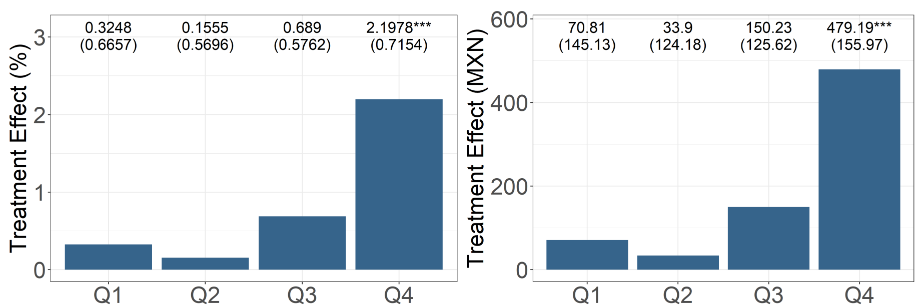


This graph shows the variable importance of the 27 most important variables used in the estimation of the causal forest. Variable importance indicates how often was the given variable used to select splits in the multiple trees of the causal forest. We first estimate the causal forest using 161 pre-treatment variables and then restrict to the 52 most important ones in the second estimation (of which the 27 most important ones are shown here). The 52 variables are listed in the caption of Figure 1. ADB refers to average daily balances, all variables are monthly.



(a) Treatment Effect (% Change) (b) Treatment Effect (MXN)

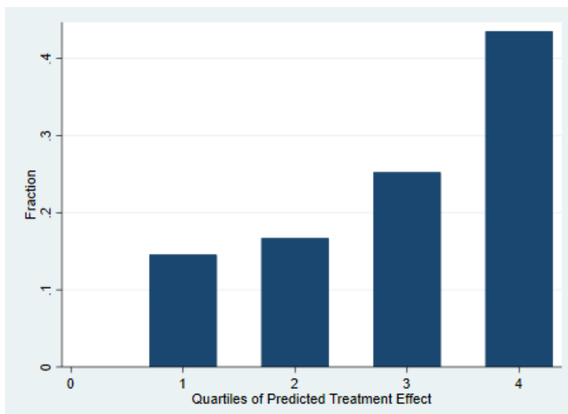
Figure 3: Treatment effect on checking account balances, as a function of individual treatment effects. Individuals in the top quartile of the predicted treatment effect distribution are split in to quintiles of predicted treatment effects, based on the score generated by the causal forest.



(a) Average Treatment Effect (% Change) (b) Average Treatment Effect (MXN)

Figure 4: Treatment effect on checking account balances, as a function predicted treatment effects for each individual. Individuals are split in to Quartiles of treatment effects on savings, based on the score generated by the causal forest.

Figure 5: Distribution of the Puzzle Group, by Quartiles of Predicted Treatment Effect.



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.

Table 7: The Treatment Effect on Savings and on Credit Card Borrowing

	(1)	(2)	(3)	(4)	(5)	(6)
Dep.Var.	Ln Checking Account Balance	Ln Credit Card Balance (Banorte)	Ln Credit Card Balance (Credit Bureau)	Ln Credit Card Interest	Paid Interest {0,1}	Ln Credit Card Payments
Panel A: All Clients with Credit Cards						
ATE	0.0601***	-0.0155	-0.0077	-0.0171	-0.0037	-0.0159
	(0.0177)	(0.0116)	(0.0062)	(0.0334)	(0.0054)	(0.0150)
Mean Dep. Var in Control Group (MXN)	31681.46	17097.99	43136.75	230.39	0.42	9500.24
Increase in Savings (MXN)	1904.37					
Upper Confidence Interval (MXN)[1]		123.54	195.50	11.12	0.0068	127.79
Upper Confidence Interval (MXN)[1] divided by increase in Savings (MXN)		0.06	0.10	0.01	0.0000036	0.07
N= 126458						
Panel B: Clients who Paid Credit Card Interests at Baseline						
ATE	0.0567**	-0.0102	-0.0091	-0.0242	-0.004	-0.0133
	(0.0251)	(0.0082)	(0.0072)	(0.0453)	(0.007)	(0.0202)
Mean Dep. Var in Control Group (MXN)	23194.21	23080.11	51491.24	413.31	0.71	8012.99
Increase in Savings (MXN)	1315.58					
Upper Confidence Interval (MXN)[1]		133.97	262.18	26.68	0.0097	210.99
Upper Confidence Interval (MXN)[1] divided by increase in Savings (MXN)		0.10	0.20	0.02	0.0000074	0.16
N= 58485						

This Table shows average 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) considering only credit cards held at Banorte, and all credit cards reported to the credit bureau respectively. Columns 4 and 5 shows 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 ln(Credit Card payments). In all cases we consider individuals in the top quartile of the predicted savings effect. 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. Average Treatment Effects are calculated with the Augmented Inverse Probability Weighted method. Treatment propensities come from estimating Causal Forests on the corresponding dependent variables. The increase in savings expressed in MXN, calculated by multiplying the ATE and the Mean of Checking account Balances in the Control Group. Upper confidence intervals expressed in MXN are calculated as (point estimate + 1.96*Estandar Error)*Mean of Dep. Var in Control Group. \[1] The upper confidence interval for the probability of incurring credit card interests during the treatment period is expressed in percentage points and not in MXN (point estimate + 1.96*Standard Error) *p<0.1; **p<0.05; ***p<0.01.

Table 8: Treatment Effects by Quintile of Saving Score for Individuals with Credit Cards

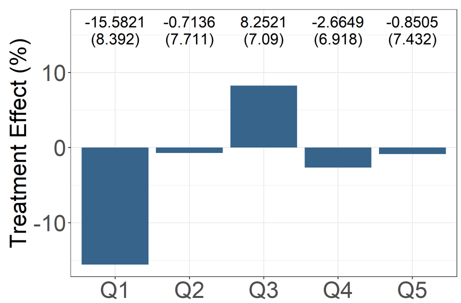
	Q1	Q2	Q3	Q4	Q5
Panel A: Treatment Effect on Checking Account Balances					
ATE Ln Checking Account Balance	0.09***	0.09***	0.05*	0.02	0.06*
	(0.0379)	(0.039)	(0.0357)	(0.0346)	(0.0478)
Mean Checking Account Balance in Control Group (MXN)	30112	28471	32456	36392	30001
Panel B: Treatment Effect on Credit Card Balances					
ATE Ln Credit Card Debt Balance	-0.0179	-0.00834	-0.1053***	0.0072	0.0032
	(0.0159)	(0.0081)	(0.0350)	(0.0081)	(0.0036)
Mean Checking Account Balance in Control Group (MXN)	50169.96	38223.04	43398.37	34334.49	55121.73
Panel C: Treatment Effect on Credit Card Interest					
ATE Ln Credit Card interest)	-0.16	-0.01	0.08	-0.03	-0.01
	(0.0839)	(0.0771)	(0.0709)	(0.0692)	(0.0743)
Mean Credit Card Interest in Control Group (MXN)	200.6	214.5	222.7	233.2	272.9
Panel D: Treatment Effect on Probability of Incurring Credit Card Interest					
ATE Probability of Incurring Credit Card Interest	-0.0213	0.0032	0.0081	-0.0099	-0.0008
	(0.0139)	(0.0127)	(0.0115)	(0.0109)	(0.0115)
Fraction Incurring Credit Card Interest in Control Group	0.3826	0.3970	0.3963	0.4060	0.4882

Interest payments are in Mexican Pesos (MXN). 1 MXN = 0.047 USD. This table considers individuals in the top quartile of the distribution of the predicted savings effects. We further split them into quintiles and report average treatment effects on savings, interest payments and probability of paying interests for individuals in each of the quintiles who have at least one credit card.

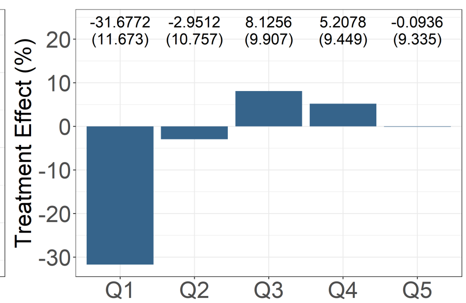
Table 9: Treatment Effects by Quintile of Saving Score for Individuals with Credit Cards who Paid Credit Card Interest at Baseline

	Q1	Q2	Q3	Q4	Q5
Panel A: Treatment Effect on Checking Account Balances					
ATE Ln Checking Account Balance	0.1**	0.14***	0.02	-0.01	0.06
	(0.052)	(0.0568)	(0.051)	(0.0493)	(0.0658)
Mean Checking Account Balance in Control Group (MXN)	22934	22375	25050	26323	19473
Panel B: Treatment Effect on Credit Card Balances					
ATE Ln Credit Card Debt Balance	-0.0116	-0.0142	0.0003	-0.0606***	-0.0161
	(0.0105)	(0.0114)	(0.0084)	(0.0268)	(0.0122)
Mean Checking Account Balance in Control Group (MXN)	63517.78	48032.82	41684.96	52989.8	63553.46
Panel C: Treatment Effect on Credit Card Interest					
ATE Ln Credit Card interest	-0.32	-0.03	0.08	0.05	0.00
	(0.1167)	(0.1076)	(0.0991)	(0.0945)	(0.0934)
Mean Credit Card Interest in Control Group (MXN)	387.8	396.4	411.1	418.7	440.0
Panel D: Treatment Effect on Probability of Incurring Credit Card Interest					
ATE Probability of Incurring Credit Card Interest	-0.0388	0.007	0.0066	-0.0013	-0.0025
	(0.0186)	(0.0169)	(0.0155)	(0.0145)	(0.0138)
Fraction Incurring Credit Card Interest in Control Group	0.6845	0.6886	0.6909	0.6977	0.7581

Interest payments are in Mexican Pesos (MXN). 1 MXN = 0.047 USD. This table considers individuals in the top quartile of the distribution of the predicted savings effects. We further split them into quintiles and report average treatment effects on savings, interest payments and probability of paying interests for individuals in each of the quintiles who have at least one credit card and paid credit card interest at baseline.



(a) Individuals without Interest Charges



(b) Individuals with Interest Charges

Figure 6: Treatment effect on credit card interest charges for individuals in the top quartile of the predicted savings effect who do or do not pay interest at baseline split in to quintiles of predicted treatment effects on savings, based on the score generated by the causal forest.

Table 10: Treatment Effects On Deposits, ATM Withdrawals and Spending

	(1)	(2)	(3)	(4)
Dep.Var.	Ln Deposits	Ln ATM Withrawals	Ln Spending with Credit Card	Ln Spending with Debit Card
Panel A: Clients With Credit Card				
ATE	-0.0083	-0.0602***	0.0122	0.0537***
	(0.0091)	(0.0090)	(0.0079)	(0.0074)
Mean of Dep. Var.	28,271.71	12,733.68	6,261.15	9,527.28
Panel B: Clients With Credit Card Who Paid Interest At Baseline				
ATE	-0.0071	-0.0737***	0.0111	0.0601***
	(0.0097)	(0.0094)	(0.0085)	(0.0078)
Mean of Dep. Var.	22,987.99	13,997.47	11,960.88	9,023.12

Interest payments are in Mexican Pesos (MXN). 1 MXN = 0.047 USD. This table considers all individuals with credit cards in the top quartile of the distribution of predicted treatment effects on savings. Deposits, withdrawals, credit card spending and debit card spending are all monthly. Card Spending is defined as the sum of debit or credit card store purchases, respectively.

Table 11: The Treatment Effect on Savings and on Credit Card Borrowing for whom Banorte is their main bank

	(1)	(2)	(3)	(4)	(5)
Dep.Var.	Ln Checking Account Balance	Ln Credit Card Balance (Banorte)	Ln Credit Card Interest	Paid Interest {0,1}	Ln Credit Card Payments
			Panel A: All Clients with Credit Cards		
ATE	0.0568***	-0.0106	-0.0029	-0.0021	-0.0108
	(0.0181)	(0.0128)	(0.0371)	(0.0059)	(0.0170)
Mean Dep. Var in Control Group (MXN)	34391.41	12889.39	213.8667	0.3539553	10312.63
Increase in Savings (MXN)	1953.43				
Upper Confidence Interval (MXN)[1]		186.74	14.93	0.0095	232.24
Upper Confidence Interval (MXN)[1] divided by increase in Savings (MXN)		0.10	0.01	0.0000048	0.12
N=89904					
			Panel B: Clients who Paid Credit Card Interests at Baseline		
ATE	0.0531**	-0.0091	-0.0197	-0.0015	-0.0093
	(0.0226)	(0.0090)	(0.0498)	(0.0077)	(0.0228)
Mean Dep. Var in Control Group (MXN)	28281.41	19264.42	434.08	0.68	8897.35
Increase in Savings (MXN)	1501.74				
Upper Confidence Interval (MXN)[1]		164.13	33.82	0.01	314.77
Upper Confidence Interval (MXN)[1] divided by increase in Savings (MXN)		0.11	0.02	0.0000061	0.21
N=41226					

This Table shows average 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) considering only credit cards held at Banorte, and all credit cards reported to the credit bureau respectively. Columns 4 and 5 shows 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 ln(Credit Card payments). In all cases we consider individuals in the top quartile of the predicted savings effect and for whom Banorte is their main bank. We say that Banorte is the main bank for individuals who receive their payroll at Banorte and who do not have credits with other banks according to credit bureau records. 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. Average Treatment Effects are calculated with the Augmented Inverse Probability Weighted method. Treatment propensities come from estimating Causal Forests on the corresponding dependent variables. The increase in savings expressed in MXN, calculated by multiplying the ATE and the Mean of Checking account Balances in the Control Group. Upper confidence intervals expressed in MXN are calculated as (point estimate + 1.96*Estandar Error)*Mean of Dep. Var in Control Group. \[1] The upper confidence interval for the probability of incurring credit card interests during the treatment period is expressed in percentage points and not in MXN (point estimate + 1.96*Standard Error). *$p<0.1$; **$p<0.05$; ***$p<0.01$.

Table 12: 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	
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?

Treatment effects are estimated in each column with a standard OLS as $y_i = \alpha_s + Treatment_i + Group_{ij} + Treatment * Group_{ij}$ 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 turns defines experimental strata. In all cases we consider 3.1 million observations. Robust Standard Errors in parenthesis. *p<0.1; **p<0.05; ***p<0.01.

Table 13: 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.

Table 14: Average Treatment Effects for users in groups with the highest observed average treatment effect and for users with the highest individual treatment effects predicted by the causal forest

	Observed Average Treatment Effects				Individual Treatment Effects predicted by Causal Forest			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep.Var.	N	Ln Checking Account Balance	Ln Credit Card Interest	Ln Credit Card Balance (Banorte)	N	Ln Checking Account Balance	Ln Credit Card Interest	Ln Credit Card Balance (Banorte)
Panel A: All Clientes	763,511							
ATE		0.2401***	-0.0197***	-0.0142***	763,625	0.0220***	-0.0023	-0.0019
		(0.0072)	(0.0037)	(0.0048)		(0.0072)	(0.0048)	(0.0041)
Mean of dep var (MXN)		18283.47	66.66463	4161.451		21872.15		
Panel B: Clients with Credit Card	126,468				126,458			
ATE		0.4403***	-0.0991***	-0.1089***		0.0601***	-0.0171	-0.0155
		(0.0148)	(0.0095)	(0.0083)		(0.0177)	(0.0334)	(0.0116)
Mean of dep var (MXN)								
		21623.82	241.41	15077.12		31681.46	230.39	17097.99
Panel C: Clients with Credit Card who paid interest at baseline	61,204				58,485			
ATE		0.5167***	-0.1109***	-0.1946***		0.0567**	-0.0242	-0.0102
		(0.0114)	(0.0094)	(0.0092)		(0.0251)	(0.0453)	(0.0082)
Mean of dep var (MXN)		14994.75	410.8639	19585.27		23194.21	413.31	23080.11

This Table shows average treatment effects on a selection of variables related to saving and borrowing behavior, for clients in groups with the highest observed average treatment effects or for clients with the highest individual treatment effects predicted by the causal forest. For columns 1 to 3 we split the sample into 6,104 mutually exclusive groups defined by the interaction of all experimental strata. For each group we calculate average treatment effects, and we assign to each observation in the group the average treatment effect of its group. We then split the sample into quartiles based on the average treatment effect assigned to each observation. The top quartile corresponds to the 25% of observations which belong to strata blocks with the highest observed average 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-blocks fixed effects. For columns 3 to 6 we use the individual treatment effect predictions from the causal forest on the entire sample. We split the sample into quartiles and calculate, for the top quartile, average treatment effects with the AIPW method included in the grf R package. *p<0.1; **p<0.05; ***p<0.01.

Table 15: Distribution of observations according to the average treatment effect of strata blocks and predicted treatment effects at the individual level

Rows: Sorting Based on Observed Average Treatment Effects

Columns: Sorting Based on Predicted Individual Treatment Effects

	1	2	3	4	Total
1	186,989	188,445	192,417	198,695	766,546
2	202,611	191,334	185,986	181,425	761,356
3	191,371	197,792	192,344	181,583	763,090
4	182,655	186,055	192,879	201,922	763,511
Total	763,626	763,626	763,626	763,625	3,054,503

This Table shows the distribution of observations according to the observed average 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 average 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 average treatment effects, and we assign to each observation in the group the average treatment effect of its group. We then split the sample into quartiles based on the average 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 average treatment effects, and viceversa.

Internet Appendix

Appendix 1   Calibration test

We formally test for whether heterogeneity in individual predictions is associated with heterogeneity in treatment effects using the “calibration test" described in Athey and Wager (2019), motivated by Chernozhukov et al. (2018). This tests seeks to fit conditional average treatment effects as a linear function of the causal estimates of the causal forest. This test 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 16 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 16: 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 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.