# The controlled choice design and private paternalism in pawnshop borrowing*

Craig McIntosh Isaac Meza Joyce Sadka Enrique Seira
Francis J. DiTraglia †

This draft: September 12, 2023

**Abstract**

In the context of pawnbroker lending, we show that forcing people into commitment contracts with a regular repayment structure decreases their (fee-including) financial cost by 20%, increases the likelihood of recovering their pawn by 15%, and increases the likelihood of repeat business by 20%. Using a multi-armed RCT that includes both a voluntary and a forced arm, we illustrate how to point-identify both the Treatment on the Untreated and the Treatment on the Treated. We find modest selection on gains but nonetheless significant positive benefits of commitment even for those who would not have chosen it.

**Keywords:** Private paternalism, choice, heterogeneous treatment effects, commitment, overconfidence.

**JEL codes:** G41, C93, O16, G21

*We want to thank Mauricio Romero and Anett John for advice and encouragement. Ricardo Olivares, Gerardo Melendez, and Alonso de Gortari provided excellent research assistance and Erick Molina helped with formatting. Jose Maria Barrero, Andrei Gomberg, Emilio Gutierrez, David Laibson, Aprajit Mahajan, Matt Rabin, Charlie Sprenger, and seminar participants at ITAM, USC, MSU, and UCSD provided valuable feedback.

†Seira: MSU, enrique.seira@gmail.com (corresponding author); McIntosh: University of California San Diego, ctmcintosh@ucsd.edu; Meza: Harvard University, isaacmezalopez@g.harvard.edu; Sadka: ITAM, jsadka@itam.mx; DiTraglia: Oxford, francis.ditraglia@economics.ox.ac.uk

# 1 Introduction

Restrictions of choice are often surreptitiously imposed on consumers. Firms limit their worker's freedom by requiring time-consuming progress reports, schools impose deadlines and pop quizzes, and loan contracts often require monthly repayment. As pointed out by Laibson (2018), although these choice-restricting mechanisms may be beneficial to firms and clients, they are often disliked by the latter, and therefore "shrouded" by firms, i.e. not advertised but nonetheless embedded in products. Laibson (2018) coins the term "private paternalism" for cases when restrictions on the freedom of choice is implemented by a firm in a way that advances a client's interests.

Making a case for private paternalism is not easy. It involves comparing the causal effect of forcing a certain product characteristic against the effect of letting people choose it. Because choice may be correlated with unobservables, identifying these two effects is challenging in observational studies. Moreover, painting a complete picture of the benefits of forcing requires counterfactuals for both choosers and non-choosers alike, since the differential benefit of forcing arises from impacts among those who do not freely choose commitment. While a large behavioral literature has illustrated that allowing borrowers or savers to choose commitment-style products increase savings in an intention-to-treat (ITT) sense (Thaler & Benartzi, 2004; Prina, 2015; Brune *et al.*, 2016; Callen *et al.*, 2019; Dupas & Robinson, 2013; Ashraf *et al.*, 2006), this literature has rarely addressed the benefits of imposing commitment or its differential impacts on those who would not independently choose commitment. Indeed, selection on gains has such wide acceptance that economists often impose it as a part of Roy models on topics as diverse as migration (Borjas, 1989) and job search (Lippman & McCall, 1976). In microfinance, a context close to ours, (Bauer *et al.*, 2012a) claim that there exists positive selection, citing a positive correlation between displaying present biased preferences in surveys and selecting into microfinance as evidence of borrowers' actually demanding the discipline imposed by frequent payments. However, they are unable to show that borrowers indeed benefit or that they indeed positively select based on treatment effects.

This paper studies the benefits of restricting choice in the context of pawnbroker lending. Pawn loans are one of the oldest, most understudied, and most prevalent forms of borrowing (Carter & Skiba, 2012). There are more than 11,000 pawn shops across the US, with 30 million clients and $14 billion yearly revenues (in China it is a $43 billion industry).[1] Our partner pawn lender (henceforth Lender $P$) alone served more than 1 million clients in Mexico in the last 3 years with more than 4 million contracts.[2] Pawnshop lending presents an inverted lending case: since these loans are over-capitalized, the lender in the contract stands to gain the most when borrowers default.

Our partner's standard pawn contract gave 70% of the value of gold collateral in credit, and charged a monthly interest rate of 7% for loans of a three-month duration, with a flexible no-reminders contract typical of the industry that can be paid back anytime before the loan comes due at no penalty. This combination of features, and the fact that the gold pawn is highly liquid,

[1] `https://tinyurl.com/ybm56dpe`, `https://tinyurl.com/y9zdcgws`, `https://tinyurl.com/y59ptdam`.
[2] For comparison there were 2.3 million micro-finance clients in Mexico in 2009 (Pedroza (2010)).

means that the lender makes 90% more profit over three months from a borrower who defaults than one who repays (30% of collateral value recovered under default, 15.8% of collateral value paid in interest if loan fully repaid). While an older literature considers the exploitative potential of over-collateralization and underpriced collateral (Basu, 1984), the implication of such contracts has not been analyzed in the behavioral literature. To repurpose the language of Laibson (2018), we suggest that pawn contracts effectively shroud their *lack* of commitment features to induce default in a manner that is not obvious to borrowers. Given that pawnshops, moneylenders, and paycheck lending institutions are likely to face client pools disproportionately made up of impatient and time-inconsistent individuals, the design of such contracts may have serious consequences for the financial lives of the poor.

Against this backdrop, we implement a multi-arm randomized control trial covering close to 10,000 pawnshop clients in 6 branches of Lender $P$ in Mexico City, using a design built to test the demand for and impact of alternative payment structures in pawn repayment contracts. Our control arm illustrates the costs of the status quo contract: fully *43%* of borrowers default and so lose their pawn as well as any payments made towards principal.[3] We then implement a "commitment choice" arm, in which borrowers can opt into a structured repayment contract at the time of taking a loan. The structured contract requires borrowers to make three monthly payments rather than one balloon payment at the end, with each monthly payment including the accrued interest at that time as well as a nominal fee of 2% of that month's payment if that payment is delinquent. This fee serves as a reminder and a means of reinforcing the importance of these interim payments. In addition, we include a "forced commitment" arm, in which all borrowers are *required* to repay using the structured monthly contract that is offered on an opt-in basis in the commitment choice arm.

The combination of voluntary and forced commitment within the RCT allows us to address several key questions. First, do structured repayment contracts lower financial costs for pawnshop borrowers? Second, do borrowers recognize this benefit, demanding commitment in sufficient numbers? Finally, and most uniquely given the presence of a forced commitment arm, we are able to ask whether the *right* borrowers voluntarily demand commitment. In particular, we use the choice, control, and forced arms to point identify the Treatment on the Treated (TOT) and the Treatment on the Untreated (TUT) effects, along with the average selection on gains (ASG) and average selection bias (ASB).

Our results suggest that commitment is strongly effective in preventing default in pawnshop lending, with the average individual in the forced arm paying financing costs inclusive of fees that are 20% lower than the control, while default decreases by 6.5 percentage points (15% of the mean). In terms of Annual Percentage Rates, the financial cost of borrowing is reduced by 34 percentage points. That is, by imposing the structure of commitment it is possible to save borrowers money by charging them fees! These results are robust qualitatively to including transport costs of going

[3] These high default rates are not uncommon, in the US default is also high at 15% (`https://tinyurl.com/yc2x5bjf`).

to the branch plus losing a day's wage for each visit, using each borrower's subjective value of the pawn rather than the appraised value of the gold, and adjusting for lost liquidity from requiring monthly payments.

The monthly payment contract seems to achieve these cost savings by speeding up payments and by generating an early bifurcation of borrowers into those that will recover the pawn and those that will not. By inducing borrowers either to fully recover or to default without making any payments it saves money that would otherwise have been paid towards loans that ultimately default, money which is lost to the borrower. In the treatment group, the first payment occurs 13 days earlier, and the fraction recovering on the first visit is 7.7 percentage points higher. Commitment contracts also decrease the fraction of borrowers who make a payment but ultimately fail to recover their pawn by 7 percentage points. Treatment effects are concentrated in the intensive margin as treatment does not affect the fraction of clients who pay a positive amount towards pawn recovery. That is, the commitment contract induces people who would otherwise pay something towards recovery but default to instead pay more and faster, making them more likely to recover their pawn at a lower financial cost.

In spite of a large average treatment effect documenting financial cost savings, only 11% of borrowers in the choice arm actually choose commitment. If the effect of commitment were homogeneous, this would be enough to conclude that the 89% who did not choose it would have been financially better off if they had. However, we test and reject the null hypothesis of homogeneous treatment effects using the method of Chernozhukov *et al.* (2018). Can the borrowers who did not choose commitment be those who simply don't need it? We address this question by sequentially imposing more structure on the problem.

We start our analysis of heterogeneity by bounding the distribution of treatment effects $Y_{1i} - Y_{0i}$ using the two marginal distributions in the treatment and control arms as in Fan & Park (2010). We find that at least 24% of individual borrowers benefit from commitment, implying many borrowers did not demand it even though their individual treatment effect is positive. Next, we impose the exclusion restriction that the effect of the contract does not depend on how they got it, that is: choosing a contract has the same treatment effect as being assigned to that contract.[4] We show that in this case our 3-armed experiment point identifies the TUT. We then estimate that the financial cost savings of the TUT is large, at \$356 pesos, equivalent to 24 percentage point savings in APR. That is, on average borrowers that did *not* choose the frequent payment contract would have benefited from having that contract.[5] In a final step, we impose a selection-on-observables assumption and use the Causal Random Forest algorithm of Athey *et al.* (2019) to estimate conditional average treatment effects (both ATEs and TUTs) for the individual borrowers

[4]This is a frequently used assumption even if often implicit. For instance, a significant literature uses schooling compulsory laws to estimate the return of one more year of school, and then often interprets that schooling, in general, has this return.

[5]We estimate a positive TOT that is larger than the TUT, but it is imprecise–this results from the smaller sample size, given that few chose the monthly payment contract–and the hypothesis of TOT=TUT is rejected only with 13 percent confidence for the APR and 12 percent confidence for default.

in our experiment. We estimate that 93% of the borrowers who did not opt for the commitment contract have positive conditional TUT effects, while only 1% of those who chose commitment have negative conditional TOT effects. While targeting commitment products to those that benefit the most is a policy that appears attractive, in this context we find that the usable targeting variables have relatively weak predictive power and hence even the random forest only reduces the overall mis-targeting rate from 8% (all to Forcing) to 5.5% (our best-case feasible targeting mechanism).

What explains the persistence of no-commitment contracts so contrary to borrowers' interests in the real world? First, we show substantial levels of over-optimism among borrowers. Among borrowers who do not choose it, those with the largest benefits from commitment are the individuals who most systematically over-estimate their own probability of repayment at the time of taking the loan. The dramatic improvements in borrower income arising from the commitment contract come directly from the pockets of lenders, reinforcing that the pawn industry as a whole has a strong interest in retaining the status quo.

Does learning eventually overcome the low demand for commitment? We find that borrowers who are assigned the forced commitment contract are 6.7 percentage points more likely to become repeat costumers than those assigned to the status-quo contract. One potential interpretation of this result is that borrowers who experience the frequent payment contract rationally choose to borrow more later on because of the lower effective financial costs that this contract entails.[6] However, a stronger test of liking the monthly payment contract is whether conditional on coming again, they would choose frequent payments. We test this and we cannot reject that demand is higher after experiencing the monthly payment contract, however sample sizes are small and power for this analysis is correspondingly low. Our results suggest that pawn lending markets capture borrower wealth in ways not well summarized by interest rates. By overcollateralizing and then structuring loan contracts in a manner encouraging of default, lenders are able to extract substantial value independent of the interest charged. Further, the 'nudge' approach generally favored by the behavioral literature (voluntary commitment) does not generate adequate demand in this context.

Our paper makes a number of contributions to the literature. First, it speaks to microfinance research on the effects of payment frequency. While experiments in microfinance markets have not shown the same benefits from providing a more regularized repayment environment as we find here (Field & Pande, 2008), these experiments differ from ours in two important ways: they are performed on top of highly structured contracts, and they involve borrower pools who may have selected into that type of lending precisely because it provides structure. These differences may explain why (Field & Pande, 2008) finds almost no default in the control group, in stark contrast to our results. Second, we provide a deeper analysis of both the take-up and efficacy of voluntary

[6] An alternative interpretation is that the monthly payments from the commitment contract sap borrowers liquidity, leading them to take a second loan to pay off the first. This is unlikely because subsequent loans fall outside the 90 day period within which payments for the first loan are due. Moreover, the correlation is also present conditional on having paid down the first loan. Borrowers tend to pledge a different piece for the second loan, so that recovery of the first pawn is also an unlikely explanation.

commitment mechanisms. A number of papers have found low demand for commitment as we do (Ashraf *et al.* (2006), Giné *et al.* (2010), Bai *et al.* (2020), Royer *et al.* (2015), Sadoff *et al.* (2019)) while others have found more robust demand for commitment (Kaur *et al.* (2015), Casaburi & Macchiavello (2019), Schilbach (2019), Tarozzi *et al.* (2009), Dupas & Robinson (2013)). Unlike all of these papers, however, we separately point identify and estimate the effects of commitment for borrowers who would and would not choose it. This allows us to conduct a more rigorous and nuanced analysis of private paternalism in our empirical context.

Two other recent papers also shed some light on private paternalism. In the context of food choice, Sadoff *et al.* (2019) show that individuals with the most time-inconsistent preferences are actually least likely to demand commitment, suggesting a lack of awareness as to self-control problems. In the context of school choice, Walters (2018) finds that students who select into more effective schools have smaller treatment effects from attending than those who do not. Our paper differs from these in both its empirical setting and its methodology. As far as we are aware, ours is the first experiment to combine control, forced treatment, and treatment choice arms and explain how such a design can be used to point identify TUT effects and a range of other interesting causal parameters under minimal assumptions. While Fowlie *et al.* (2021) likewise exploited a three-armed experimental design in their study of the effect of electricity pricing, they identified two TOT effects for different groups of "treated" households, whereas we simultaneously identify the TOT and TUT effects defined with respect to a single "treated" group of borrowers. We call our approach the "Controlled Choice design" for short. Other approaches to going beyond local average treatment effects (LATE) from the literature either assume no unobserved selection-on-gains (Aronow & Carnegie, 2013; Angrist & Fernandez-Val, 2013) (the "LATE-and-reweight" approach) or rely on a combination of instruments with rich support and additional structural assumptions such as additive separability of observed and unobserved determinants of treatment effect heterogeneity (Heckman & Vytlacil, 2007; Cornelissen *et al.*, 2018) (the marginal treatment effects approach). Walters (2018) takes the latter approach, combining a distance instrument with additional structural assumptions, obtaining model-based TUT and ATE estimates that differ substantially from those implied by a LATE-and-reweight approach. In contrast, we leverage our experimental design to point identify the TUT effect using a binary instrument without the need for a structural model, relying instead on relatively simple exclusion restrictions. This restriction has testable implications that we fail to reject. Rather than assuming away unobserved selection-on-gains, as in the LATE-and-reweight approach, we estimate it directly. Unlike Sadoff *et al.* (2019) we directly identify the TUT, obviating the need to first elicit preferences before testing for negative selection. In a returns-to-education setting, Oreopoulos (2006), uses a similar direct approach to identify a LATE *nearly* equals a TUT effect: the 1944 Butler Act increased the share of British children who remained in school until age 15 from 43% to over 90%. Because 100% of borrowers in our forced arm are treated, we identify the TUT rather than an approximation to it.

The remainder of the paper is structured as follows: Section 2 provides context and defines our main outcome variables. Section 3 describes the experiment and data sources, and shows pre-treatment balance across arms. Section 4 provides the standard ITT analysis of the experiment, while Section 5 presents our strategy for estimating ToT and TuT effects, and Section 6 investigates why paternalism functions so well in this context and whether it can be more finely targeted. Section 7 concludes.

# 2 Context

## 2.1 Pawnshop borrowing

Pawn loans involve individuals leaving valuable liquid assets, typically jewelry, as collateral in exchange for an immediate cash loan. Collateral is typically larger than the loan, allowing lenders to skip the checking of credit history and give the loan immediately. This makes pawn loans a popular way to get cash to pay for emergencies. In fact, they are one of the most prevalent forms of borrowing. There are more than 11,000 pawn shops across the US, with 30 million clients and $14 billion yearly revenues.[7] Our partner pawn lender alone served more than 1 million clients in the last 3 years with more than 4 million contracts. For comparison there were 2.3 million micro-finance clients across all lenders in Mexico in 2009 (Pedroza, 2010).

Pawning is also one of the oldest forms of borrowing. Pawn lending existed in antiquity at least since the Roman Empire, and there are records of it in China about 1,500 years ago (Gregg, 2016). In spite of the high prevalence and long history, pawnshop borrowing has not received much attention in the economics literature. The closest widely studied product is payday loans. In developing countries, however, payday lending is likely small compared to pawnshop lending; the latter is faster and requires less documentation, making it more accessible to informal sector workers who receive their salaries in cash.

As with payday lending, pawnshop lending is controversial. Regulators have concerns with the sophistication of borrowers using it, speculating they may suffer from behavioral and cognitive deficiencies which lead to make sub-optimal choices.[8] There is some evidence in support of this view for payday borrowers.[9] Our study reinforces the idea that a lack of sophistication may be an integral part of that way that standard pawn contracts are structured.

[7] `https://tinyurl.com/ybm56dpe`, `https://tinyurl.com/y9zdcgws`, `https://tinyurl.com/y59ptdam`.

[8] The US congress has actually banned the payday lending industry from serving active military personnel, and some States in the US have imposed zoning restrictions, interest caps, and restrictions on serial borrowing as consumer protection measures against payday lending (Stegman, 2007).

[9] Bertrand & Morse (2011) write that "Under the view that the people borrowing from payday lenders are making an informed, utility-maximizing choice given the constraints that they face, one would not expect additional information disclosure about the payday product to alter their borrowing behavior", but to the contrary, they find that simply disclosing how financing costs add up reduced demand by 11%. Melzer (2011) finds that payday loan access leads to increased difficulty paying mortgage, rent and utilities bills.

## 2.2 Pawning Logistics and Contracts

To study this market, we partnered with one of the largest pawn shops in Mexico, an institution with more than one hundred branches spanning multiple states in Mexico. This lender (whom we refer to as 'Lender P') has a simple and typical business model.

**Appraising and lending.** Lender P takes gold jewelry as collateral in exchange for a fraction the value of the piece, in cash. No other collateral and no credit history checks are needed. The transaction takes less than 10 minutes and is conducted at the branch in person between the client and the appraiser (i.e. a teller, see Figure OA-1). The appraiser weighs the gold piece and runs tests on its purity. Based on these she assigns a gold value to the piece, stores it as collateral, and gives 70% of the gold value of the piece in cash to the client. The borrower signs a 2-page contract with the conditions of the loan and leaves with the cash.

**Contract.** Lender P had only one type of contract, henceforth the *status quo* contract. It stipulated that the interest rate was 7% *per month* compounded daily on the outstanding amount of the loan. The loan had a 90 days term with 15 days' grace period. The client could make payments at the branch at anytime with no penalty for pre-payment. No reminders or interim contact exists between the lender and the borrower. If the client returns to pay the principal plus the accumulated interest within 105 days, she received back her pawn, otherwise the pawnbroker kept the piece *and* any payments made. Before the contract expired, the client had the right to renew for another 3 months by going to the pawnshop, paying the accumulated interest, and signing a new contract with exactly the same terms and same piece as the original one a (26% of borrowers renew at least once with a given pawn). This contract was standard in the industry. Pawnshops make money in three ways: by reselling the jewelry left as collateral on defaulted loans, by charging interest on non-defaulted loans, and by keeping the payments made on defaulted loans.

**Borrowers.** The clients that pawned understood these terms well (as we verified in interviews).[10] These clients have little or no access to other types of loans and they value the convenience of pawn borrowing. This population of pawn borrowers is economically vulnerable: 30% of them could not pay either water, electricity & gas or rent in the past 6 months; 87% said they are pawning because of an emergency, and only 13% stated it was to use in a 'non-urgent expense'. When asked why they are pawning this piece, 5% responded 'lost a family member', 'a medical emergency' (11%), or 'an urgent expense' (71%).

**Many borrowers lose their pawn.** Our context is also one with high borrower default: 43% of clients lose their pawn in a time span of 230 days from the date of pawning. One potential explanation for high default is that clients are really just knowingly selling their gold piece through

[10] 87% of clients report in our survey that they have pawned before.

a pawn contract on which they intend to default. This appears unlikely for several reasons: (a) clients can easily sell the gold and obtain a higher amount of instant cash at gold buying stores located close to almost all our pawnshop branches (see Figure OA-2), (b) the reported subjective value of the pawn is larger than the loan size for 86% of clients, (c) among those that lose their pawn, 48% paid a positive amount towards its recovery and on average paid 42% of the value of their loan (see Figure OA-3 in Appendix) — this can only be rationalized if they expected to recover their pawn, and (d) 74% of borrowers report a 100% probability of repaying their loan (and 98% at least a 50% chance of repaying) in our baseline survey at the time they take the loan. Note that high default could be detrimental from the lender's point of view, since it may reduce the likelihood that the client becomes a return customer.

## 2.3 Measuring Borrowers Financial Costs

Borrowers' financial costs are composed of two main categories: the cost of losing their collateral, and the interest incurred during the life of the loan. For each given loan we observe if the client lost her pawn ($\mathbb{1}(\text{Default}_i)$) and when. If the client has renewed her loan several times such that the loan is continuing but has not defaulted, we pursue the strategy of coding default as zero (this approach is conservative in our context in that it does not lead the reduction in repayment time to translate mechanically into a drop in default). In our data 13% of experimental loans are continuing (i.e. censored) when the data period ends. Regarding interest, the administrative data classifies payments made in three types according to their payment allocation rules: payments to principal $P^c$, payments on generated interests $P^i$, and payments on penalty fees $P^f$. We observe each and every payment made under each category, its amount and date.

We define a borrower's financial cost as the total monetary outflow –in cash or pawn value– from the borrower to the lender. This includes all payments the borrower made toward interest and fees, but also the value of the pawn when there is default. When there is no default the borrower gets her pawn back and there is no loss of value for the borrower. Payments towards capital are considered a cost only when the borrower defaults, as she does not get reimbursed for these. Note however that when she does not default payments to capital are not an actual outflow, as they sum up to the value of the loan the lender disbursed in the first place. The formula for financial cost for person $i$ is thus as follows:

$$\text{Financial Cost}_i = \underbrace{\sum_t P^i_{it}}_{\text{Payments to interest}} + \underbrace{\sum_t P^f_{it}}_{\text{Payments of fees}} + \underbrace{\mathbb{1}(\text{Default}_i) \times (\text{Appraised Value}_i}_{\text{Cost of losing pawn}} + \underbrace{\sum_t P^c_{it}}_{\text{Payments to capital}})$$

where $t$ indexes days, and $\mathbb{1}(\text{Default}_i)$ is an indicator function for defaulting. Because the period of the loan is only 90 days we do not use discounting in calculating costs. In robustness checks we show that results are virtually unchanged when applying a wide range of time discounting factors.

We believe the above measure captures financial cost in pesos accurately. However, we will also report results incorporating two costs that are not financial: (i) using the subjective value of the pawn reported by the borrower in place of its appraised gold value, and (ii) adding a measure of travel expenses and the opportunity cost of time, as clients have to go to the branch in order to make payments.

As a second measure of cost we calculate the Annual Percentage Rate (APR) in order to express the cost as a percentage of the loan, per year, inclusive of default costs.[11] The standard definition is given in the following formula:

$$(\text{APR})_i = \left(1 + \frac{\frac{\text{Financial Cost}_i}{\text{Appraised Value}_i}}{\text{loan term}_i}\right)^{\text{loan term}_i} - 1$$

Figure 1(a) plots a histogram of this financial cost in pesos for all loans in the experiment, distinguishing by whether we observed they defaulted (43%), recovered their pawn (44%), or whether they are censored in our data (13%). The first thing to note is that those that default incur in higher costs; part of this is mechanical as losing their collateral is part of the cost. The second thing to note is that those that are not closed at the end of our sample period have already incurred in substantial financial cost, even though they have not defaulted yet. Panel (b) shows the APR; most of the APR costs come from clients that do not recover their pawn.

# 3 Experimental Design

## 3.1 Treatment arms and randomization

**The Commitment Contract.** For the purpose of the experiment we designed a new contract that is identical to the status quo contract except that, informed by the design of micro-lending contracts, it enhances the regularity and salience of payments as a way to encourage repayment (Morduch, 1999; Bauer *et al.*, 2012b). It has the same interest rate (7% *per month*) which accumulates daily on outstanding debt, it has the same loan size/collateral ratio (70%), it has the same loan term (90 days, and a grace period of 15 days), and the gold pawn gets appraised in the same way by the same appraisers. The Commitment contract however requires the client to make regular monthly payments for the duration of the contract, with the principal and interest payments split evenly across the three months of the contract (day 30, 60 and 90 after loan disbursement). The importance of this monthly payment was made salient in the contract and payment receipts (see

[11] Loan term takes into account the entire loan period, including extensions of the loan through refinancing.

Figure 3), and by the levying of a nominal fee (2% of minimum due) on individuals who fell behind in their payments. The idea that the fee itself is not driving the treatment is reinforced later in the paper where we show financial benefits many times larger than the fee, as well as treatment effects even on those who would fail to recover pawns and hence not pay the fee. Transportation cost to go to the branch to pay is on the same order of magnitude as the fee on average. To elicit demand for the monthly payment contract, we include an arm that allows borrowers to opt into this contract if they choose. The existence of both a non-optional "forcing" arm, and a choice arm in our design is key to estimate a battery of treatment effects above an average treatment effect under fairly mild assumptions.

**Treatment Arms.** Treatments were randomized at the branch-day level. Each day a computer randomly assigned which types of contracts were on offer that day in the branch, and the IT system would only offer these. We have 3 different experimental arms[12]

1. *Control* arm: consisted of branch-days offering the status quo contract described in Section 2, and only this contract.
2. *Forced Commitment* arm: consisted of branch-days requiring all borrowers to use the Commitment contract described above.
3. *Commitment Choice* arm: consisted of branch-days offering the client *a choice* between the Commitment contract, and the status quo contract.

We did not allocate an equal number of days across arms, since we were interested in having more power in some of them. The number of branch days allocated to each were 84 to control, 80 to forced commitment, and 93 to choice. See Figure 2 for a CONSORT-style diagram of the study design and recruitment.

**Randomization.** Starting September 6, 2012, we implemented the experiment in 6 branches of Lender $P$. The branches were selected by Lender $P$ to be dispersed across Mexico City and have varying sizes. In four of them the experiment ran for 107 days, and in 2 of them we ran it for a shorter time to economize on data collection costs once we realized we would not be constrained by sample size. Branches are more than 5 km apart from each other, and there is little substitution among them; only 1% of consumers appear in more than one of our branches.

Branch personnel did not know which treatment would happen on which day and were blind to the objective of the intervention. They were told that there were 3 different "types of contract-days", that the system chose randomly which applied on any given date, and that it could happen for instance that two consecutive dates had the same contract. They were also told that this way

[12]The experiment included other arms that involved no fee penalties and did not emphasize the structure of payments. These are being analyzed in a separate paper.

of operating was in place in several of Lender P's branches (they did not know which ones), and that it would be in place for several months. Randomizing at the day level limits the problem of contamination arising from clients realizing that other clients get different contracts than theirs. It also limits potential manipulation by appraisers, who in the presence of individual-level randomization could potentially pick their preferred customer from the line or tell them to wait until their desired contract shows up on the screen. Intra-branch day correlation on the probability of default (ICC) is small, at 0.05, so we lose little power vis-a-vis individual-level randomization.

Some clients pawned more than one time during the duration of the experiment, with 14% pawning 2 times and 8% more than 2 times. To have a clean comparison we are considering only the first pawn in the experiment. It is also the case that 30% of those first pawns involve more than 1 loan, as 2 or more pieces of gold were submitted. We treat each of them as separate loans.

**Timeline.** Figure 2 shows the timeline when the experiment happens and also the length of time for which we observe payments. For contracts in the first week, we observe up to 338 subsequent days, while for contracts in the last week we observe up to 235 days. Figure 2 also illustrates the number branch-days per arm, the number of contracts, and the number of surveys.

**Explanation.** Since we are interested in measuring the effect of different contract terms on client behavior, it is important that clients understand them. To ensure this, we built two 'check-points'. First, two enumerators were present in each branch for the whole day during the duration of our experiment to explain the contract terms to clients. One of the aspects emphasized was that the frequent payment contract involved the commitment to pay a third of the outstanding amount <u>each month</u>, and the nature of the fees associated with the Commitment arm. Figure 4 translates a piece of the materials we used to explain the contracts. The explanation took about 3-5 minutes and was pursued until the client said she understood the contract terms. Enumerators then asked clients to explain back the contract terms and corrected any misunderstandings. The second check-point was that before the clients signed the contract, the appraiser made them read the "Contract Terms Summary" sheet shown in Figure 3. It was a piece of paper given to clients after their piece had been appraised and the size of the loan determined, but before they signed their contract. The appraiser read it and asked the client to sign it as proof of understating. We are confident the overwhelming majority of clients understood the contracts and that those in the choice arm made informed choices.

## 3.2 Data

**Administrative data.** The study exploits two types of data: administrative data from the lender, and a short survey we implemented. The administrative data contains a unique identifier for each client, an identifier for the piece she is pawning, and the transactions relating to that piece. These transactions include the value of the item as assessed by the appraiser, the amount of money loaned (70% of the value), the date of the pawn transaction, and the type of contract for that pawn. Within the period of the loan, we followed each transaction related to that piece in the administrative data: when payments were made and for what amounts, whether there was default (i.e. the client lost her pawn), and whether any late-payment fees were imposed. After the experimental loan, we are able to track subsequent behavior and to see whether that borrower took a subsequent loan. We have this information for all the pawns that occurred in the experiment's 6 branches between August 2, 2012 and August 13, 2013, this includes all the pawns under our experiment but also those that happened 1 month before it started and up to 8 months after it ended. Figure 2 shows the design and timing of the experiment, along with the sample sizes in each arm. The experiment comprises 13,446 pawns, and our administrative data cover a total of 23,974 pawns.

**Survey data.** In addition, we had a team of enumerators in each branch collect surveys and clearly explain the contract terms to walk-in clients. The enumerators were inside the branch and asked clients to complete a 5-minute survey *before* going to the teller window to appraise their piece and before treatment status was known to them. The survey was intentionally short to avoid discouraging the potential clients from pawning, but at the same time it aimed to measure the following: demographics, proxies for income/wealth, education, present-biased preferences, experience pawning, if family or friends commonly asked for money, how time consuming and costly it was to come to the branch, the subjective probability of recovering the piece that they intended to pawn, the subjective value of their piece in money terms (how much money they would sell it for), among others. We surveyed 10,437 clients, and the survey response rate was 78% among clients who took loans.[13]

### 3.3 Experimental Integrity

**Attrition.** There are two main channels through which attrition could complicate the interpretation of our results. The first, and more serious, is the possibility that clients might change their pawning decisions in response to the treatment they encounter in a given branch on the day they enter. If this occurred it would introduce a self-selection dimension which would still reflect the overall impact of a treatment for the lender's portfolio but would no longer deliver *ceteris paribus* effects of treatments on individual borrowers. Narrative reports and the way the treatment was implemented make us believe that selection into treatment is unlikely.[14] If the treatments had induced demand-side selection, we would expect to see that the number of pawns successfully conducted differ in a systematic way across arms. To test for this, we examine the number of loans

[13] Appendix OA-1 transcribes the questionnaire in English.

[14] Potential clients did not know that different days could have different contracts. If they asked, appraisers said that whatever was offered on that day was the only available contract for an undetermined length of time. Anecdotally, appraisers told us that they did not think refusals differed across arms, and our enumerators informed us that potential clients rarely left the branch without pawning. Lender P also never complained to us that our different treatments were hurting sales.

given per branch-day, with specific attention to whether the Forced Commitment arm led to a lower number of borrowers per day. The first row of Table 1 shows there is no difference between the Control and Forced Commitment arm in terms of the number of pawns per branch-day.[15] The second row makes the same point in a more focused way, given that the surveys were conducted prior to the revelation of treatment status. In no arm did more than three percent of individuals who responded to the survey go on to refuse loans, and neither these rates nor the rate of survey response are different across arms. Therefore it appears that the treatments have not induced any endogenous shifts in the composition of borrowers.

A less critical form of attrition would be a differential refusal to answer the survey questions. The survey was done before treatment status was revealed, and we observe loan outcomes regardless of whether the survey was conducted. Our core experimental estimates do not use the survey data as covariates, but the Random Forest analysis in Section 6 is restricted to the subset of borrowers who answered at least some survey questions. The bottom row of Table 1 shows that the survey response rate is broadly similar across arms (about 77 percent).

A more subtle form of sample selection could arise if the treatments induce borrowers to re-pawn in different ways, especially given that their treatment status on subsequent loan/days may not be the same as that initially assigned. To handle this issue our analysis uses only the first loan taken by each individual during the experimental window.

**Balance.** Table 2 presents summary statistics for the sample of actual borrowers across arms, and illustrates that randomization worked to achieve balance across the experimental arms. Panel A uses administrative data for the universe of borrowers in each arm, and shows that loan balances and the days on the week on which individuals pawned are comparable across arms. The average loan size is $2267 MXN ($130 USD). Panel B of Table 2 reports summary statistics across arms from our survey data. 73% of clients are women, with an average age of 43 years; 66% of them have completed high school or more, and 90% have pawned before, so that our sample has mostly experienced borrowers. Finally, the subjective probability of recovery is close to 93% on average, which makes a stark contrast to the actual rate of recovery of 43%; borrowers are highly overconfident on average. The average subjective value they report for the items they pawn is 4084 MXN, much larger than the average appraised gold value of 3238 MXN. While this could arise either from overconfidence in valuation or from undervaluation by the lender, in any case it is *prima facie* evidence that loss of the pawn should be undesirable relative to the quantity of liquidity leveraged by the asset.

## 4 Average Treatment Effects

We begin by estimating average treatment effects of assignment to the Forced commitment and the Choice arms, relative to those assigned to the Control arm. As we explain below, only about 11%

[15] The Choice arm appears to have a somewhat larger number of borrowers than the Control arm (p-value=0.06), but we cannot reject equality with the Forced Commitment arm (p-value=0.41), or across all arms (p-value=0.16).

of those in the choice arm chose the monthly payment option (Figure OA-4 shows coefficient plots for the characteristics that determine choosing commitment in the Choice arm).

**Specification.** Table 3 presents estimates and standard errors from a standard pooled experimental regression

$$y_{ij} = \alpha + \beta^F T_i^F + \beta^C T_i^C + \gamma X_{ij} + \epsilon_{ij} \tag{1}$$

where $i$ indexes client, $j$ indexes branch, $T_i^F$ and $T_i^C$ are indicator variables for receiving the Forced or Choice arms, $X_{ij}$ are branch and day-of-week fixed effects. Standard errors are clustered at the branch-day level, the unit of treatment assignment[16]. Given perfect compliance in the former, the coefficients $\beta^F$ is the ATE of the Forced arm and $\beta^C$ is the ITT of the Choice arm on the outcome variable $y_{ij}$. Our two primary outcome variables are financial cost in pesos and Annual Percentage Rate (APR), as defined in Section 2.3. Results for these outcomes appear in columns (1) and (6) of Table 3. The remaining columns decompose the financial cost and APR outcomes into their components: interest payments (col 2), payment towards any fees incurred (col 3), payments toward the principal (col 4). Column 5 shows the value of lost pawn conditional on losing it. In column 6 the dependent variable is a dummy indicating default. Finally, column 7 rescales financial cost as a function of loan size to estimate causal effects on incurred APR.[17]

**Results.** The results are stark. The Forced Commitment arm yields large and significant decreases in the cost of loans to clients, as measured either by financial cost or APR. Despite causing an increase in fees, the Forced arm leads to a decrease of 379 pesos in the costs of borrowing (out of a group mean of 1851 in the status quo), equivalent to 20% reduction as a fraction of mean cost. These cost savings arise from a 6.5 percentage point (pp) decrease in the probability of default (out of a baseline mean of 44pp, implying cost savings of 254 pesos), and also from lower interest payments, since as we will document the treatment is effective at speeding up payments, with the benefit that the interest rate applies to a smaller principal. This translates into a large reduction in APR. A credit product that has an effective average APR of 184% in the status quo arm (inclusive of default) is reduced to a cost of 150% through the imposition of a more regularized payment structure.

In contrast, to the impressive effectiveness of the Forced commitment arm, the Choice arm fails to deliver significant changes in any measure, with the exception of an increase in fees, for which we are highly powered since this outcome is zero for every control observation. Giving borrowers the

[16] A minority of clients pawned on more than one day during the experiment: 15% pawned on two distinct days, and 7.5% on three or more days. To avoid contamination from earlier treatments to which these individuals were exposed, we restrict our sample to each client's *first visit*. Note that a client may pawn multiple items her first visit. We include these as separate observations. Because our standard errors are clustered at the branch-day level, they automatically account for any dependence in error terms arising from multiple pawns by the same client on her first visit.

[17] As we explained above, loans can be extended for an additional 3 months by paying the interest owed and restarting the loan under the same treatment conditions. This means that some loans extend for more than 3 months. We consider the entire flow of cost from the duration of our sample.

choice of contract did not decrease financial cost, whereas forcing them into a structured payment contract dramatically reduced it. As we explore later in the paper, the null effect of the Choice arm arises because few people had demand for it, with 89% choosing the less effective status-quo contract.

**Intermediate outcomes.** To shed light on the mechanisms behind the ATEs discussed above, Table 4 shows how treatment affects a number of intermediate outcomes. One can group the types of intermediate outcomes into two categories: measures of the speed of pawn recovery, and measures of the decision of when to default. While the first payment for borrowers in the status-quo contract occurs on average only on day 82 (on a 90 days contract), borrowers in the forced commitment arm start paying 13.8 days earlier on average (col 1). Not only do they start paying earlier, the first payment is also 7.9% larger (col 2), and a larger fraction of 7.9% actually pay in full and recover their pawn in the first payment , compared to 30% in the status quo contract (col 3). The resolution of the loan (either by payment or default) is shortened by 27.9 days (col 4), and conditional on recovery (an endogenous control) by 17.9 days (col 5).

A very undesirable outcome from the borrower's perspective is to pay towards the loan without paying in full, i.e. defaulting on the loan. In this case, they lose both the collateral and any payments made toward recovery. One could be concerned that by encouraging them to pay monthly, more borrowers could end up in this dire scenario in the Forced commitment contract. Column 6 shows this is not the case. On the contrary, 7 percentage points fewer borrowers end up in this situation, compared to 12 percent in the status quo contract. Conditional on defaulting those assigned to the Forced commitment contract have paid 3.9% less of their loan (col 7), and 14 percentage higher proportion of borrowers pay zero conditional on defaulting, an outcome analogous to selling their pawn (col 8). One interpretation of these results is that the Forced commitment contract forces borrowers to think earlier about whether they will indeed be able to eventually recover their pawn, and separates borrowers into those "selling" their pawn and those recovering it, reducing the share of undecided borrowers that end up paying interest and end up losing the pawn anyway. This mechanism may also help to explain why the Forced commitment contract does not increase the number of visits to the branch to pay (col 10): those recovering their pawn visit more, but those defaulting have 0.19 fewer visits (col 11). Figure OA-5 plots the histogram of the timing of payments in each arm, and shows that the commitment does appear to bind in that it generates a highly regular pattern of payments at 30, 60, and 90 days after the loan is taken. Finally, observe that Column 9 shows that treatment effects are concentrated in the intensive margin as treatment does not affect the fraction of clients who pay a positive amount towards pawn recovery.

**Other costs.** We have shown that forcing borrowers to take the monthly payment contract significantly reduces financial cost. Although the paper focuses on financial cost, we consider three additional costs here. First, we include the cost of going to the branch. This includes the self-reported transport cost (most use public transport), as well as the opportunity cost of time. To err on the conservative side, we subtract a whole day's minimum wage the day of the visit, instead of just the wage corresponding to a couple of hours. Second, we consider a rough proxy of the value of liquidity they lose by paying earlier. To do this we add the interest costs on to the payments in the Forcing and recompute treatment effects with these payments compounding daily (as if they had to borrow in order to make the more rapid payments). Thirdly, so far we have valued the collateral at the gold value appraised by the lender, but the piece may be worth more to the borrower than its gold value. For many of them the pawned jewelry has sentimental value. This is reflected in the subjective valuation they reported in the survey which is 87% higher on average. Our third extra cost considers this higher value.

Table 5 shows results. Panel A reports financial cost in pesos, while Panel B shows APR. Columns (1) and (6) replicate our previous results for comparability. Columns (2) and (7) of Table 5 use the subjective value of the pawn reported by the borrower rather than its appraised value. Columns (3) and (8) adjust for self-reported transport costs per visit plus an entire day's wage, both multiplied by the number of visits that each individual made.[18] Columns (4) and (9) adjust to consider the liquidity cost. Finally columns (5) and (10) include all three changes together. The main takeaway from the table is that results are quite robust to including a much expanded measure of costs.

**Repeat Pawns.** Table 6 explores the effects of treatment on future pawning behavior. Column (1) shows that participants assigned to the the forced arm are 6.7% more likely to be repeat clients later. While this appears to be *prima facie* evidence of greater satisfaction among borrowers in the forced arm, the interpretation is complicated by the fact that monthly payments may themselves trigger more borrowing to pay them. This is unlikely to be the case given that the effect on re-pawning comes after 90 days (during the period of contract demanded payments) and not before (see columns 2 and 3). Column 4 only considers new loans which use different collateral from that of the initial one. We do this to foreclose the explanation that those in the Forced arm repeat more because they are more likely to have collateral, as they recover the collateral of the first loan more often. Column 5 focuses on the (endogenous) subsample of those recovering their pawn in both arms of the experiment. This means that both have the pawn to re-pawn and also that the liquidity demands from the monthly contract are not longer binding as the contract has been closed. We find that the difference between the Forcing contract and the status quo is even larger in this subsample, with the former having 11pp higher likelihood of being a repeat client during our sample period.

**Censoring of Loan Completion** The window of time during which we were able to observe borrower behavior was limited in each branch, meaning that there were loans that we do not see

[18]For clients who did not complete the individual survey, we adjust using the mean self-reported transport cost among respondents of the respective branch.

completed (particularly those pawns that were rolled over for one or two further 90-day spells). Overall, 13% of all experimental loans are censored, meaning that they neither default nor repay within the observation window. In the prior analysis we handle this issue by using outcomes such as ‘did not default’ which are defined even when we do not observe the completion of the loan, and only considering costs that we observe directly. We have also showed, however, that one effect of the forcing arm is to accelerate repayment, meaning that it is less likely for loans in this arm to be censored. This issue is illustrated in Figure OA-7, which shows the CDF of loan completion (either default or recovery in Panel (a)) and loan recovery (Panel (b)) by the number of days since first pawn. Two features of these graphs are salient for our analysis. The first is the extent to which loan outcomes are observed more quickly in the forcing arm. This is primarily due to the the substantially higher rate of repayment of Forced Commitment loans at 120 days (15 pp higher than the other arms). The second is the very low rate at which loans are recovered in any arm after 120 days. In the 180 day window after the second rollover period we see a further 10 percent of loans repaid in all arms, but these loans are largely dormant, suggesting that many of the censored loans will in fact be defaulted upon.

The confluence of censoring and a treatment effect on censoring is problematic from an experimental point of view. The approach taken so far is a conservative one in that it inherently assumes that all of the loans outstanding at the end of the observation window will be repaid, making it so that the acceleration of payment observed in the Forced arm does not translate mechanically into the treatments decreasing default. Given the centrality of this issue to the observed costs, we now delve into it more detail. One way of considering the effect that this issue could have on our results is to make extreme assumptions about the outcome of these loans in the treatment and control so as to bound the possible influence of censoring. In Table OA-5 we compare the Forced and Control arms, making the bracketing assumptions about repayment on censored loans. Panel A assumes all censored loans are repaid, and Panel D that all default. Panel B provides the lower bound for the treatment effect by assuming censored control loans are always repaid and treatment loans never are, and Panel C the upper bound by making the reverse assumption. Comfortingly, even with these extreme assumptions the significance on the main treatment effects never flips and treatment effects on financial cost and interests payments remain negative and significant in all scenarios. So there appears to be no scope for the censoring issue to overturn our main results.

Finally, Panel E of this table conducts a logit prediction model that uses all of the available information on loans that were completed to predict the outcome of loans that were not. This is a best guess of the outcome on censored loans. Using this prediction, we replicate the main experimental results and find that the treatment effect on financial cost increases from -378 (main results) to -525 (censored loans predicted), and the APR from -34% to -62%. Hence, while the censoring issue does have a substantial influence on the magnitude of our estimated treatment effects, these checks confirm that a) the core results are robust to censoring, and b) the headline approach that we take to the issue is conservative and is likely understating the true magnitude of impacts.

# 5 Choice and Heterogeneous Treatment Effects

The results from Section 4 show that commitment *works*: clients who were assigned to the forced commitment arm experienced substantially lower financial costs on average. In spite of this, given the opportunity, only 11% of borrowers chose commitment. If the effect of commitment were homogeneous, this would be enough to conclude that the 89% who did not choose it would have been financially better-off if they had. In a world of heterogeneous treatment effects, however, low demand for commitment could still be consistent with borrowers adhering to a standard model of rational choice. The borrowers who did not choose commitment could simply be those who don't need it. Indeed, we find strong evidence that the effect of forced commitment varies substantially across individuals in our experiment: Appendix D. tests and rejects the null hypothesis of homogeneous treatment effects, following the methodology of Chernozhukov *et al.* (2018). So the question remains: do the 89% who do not choose commitment know something about their personal situations that we as researchers do not, or are most people in the choice arm making a costly mistake? In this section we present of econometric exercises that shed light on this question, leveraging unique features of our experimental design. Among other results, we show that commitment lowers average financial costs even for the subset of borrowers who would *not* choose to commit voluntarily. To simplify the exposition in this and all sections that follow, we re-define all outcome variables so that *beneficial* treatment effects are *positive*. Using this convention, a positive treatment effect of commitment on financial cost, for example, reflects the average cost *savings* caused by commitment.

## 5.1 Bounding the Distribution of Individual Treatment Effects

If we knew the distribution of individual treatment effects, it would be easy assess how many individuals would win and lose under a policy of universal forced commitment, along with the magnitude of any individual harm. Because we can never simultaneously observe the same borrower in both the control and commitment arms, however, the distribution of individual treatment effects cannot be point identified. It can, however, be bounded. We begin our exploration of heterogeneous treatment effects by computing assumption-free bounds on the share of individuals who benefit from forced commitment.

Let $Y_{i0}$ be borrower $i$'s potential outcome under the control condition, $Y_{i1}$ be her potential outcome under forced commitment, and $\Delta_i \equiv Y_{i0} - Y_{i1}$. Because it randomly assigns borrowers to the control and forced arms, our experimental design point identifies the marginal distributions of $Y_{i0}$ and $Y_{i1}$, call them $F_0$ and $F_1$. Now, define the functions $\underline{F}$ and $\overline{F}$ as follows:

$$\underline{F}(\delta) \equiv \max\left\{0, \sup_y F_1(y) - F_0(y-\delta)\right\}, \quad \overline{F}(\delta) \equiv 1 + \min\left\{0, \inf_y F_1(y) - F_0(y-\delta)\right\}.$$

Since $F_0$ and $F_1$ are point identified, so are $\underline{F}$ and $\overline{F}$. Fan & Park (2010) show that the sharp (best possible) pointwise bounds for $F_\Delta$ are given by $\underline{F}(\delta) \leq F_\Delta(\delta) \leq \overline{F}(\delta)$. These bounds are simple to compute, and can be surprisingly informative. Here we use them to bound the fraction of borrowers who benefit from forced commitment. Given the way that we have defined our outcome variables, this is the share of borrowers whose treatment effect is positive, i.e. $\mathbb{P}(\Delta_i > 0) = 1 - F_\Delta(0)$. To construct the sharp bounds for this quantity, we simply substitute $\delta = 0$ into the preceding equations and estimate $F_0$ and $F_1$ using their empirical analogues constructed from the forced commitment and forced no-commitment arms of the experiment.

Figure 5 plots the sharp bounds for $F_\delta$ based on the APR outcome defined above in Section 2.3. Our point estimates of $\underline{F}(0)$ and $\overline{F}_0$ are .03 and .76 respectively, with associated 95% confidence intervals of [0.027, 0.040] and [0.73, 0.78], which says that at least 24% of individual borrowers benefit from commitment (and at most 97%).[19]

The bounds we have just described are constructed by considering all possible joint distributions of $Y_0$ and $Y_1$ that are compatible with the observed marginal distributions $F_1$ and $F_0$. With stronger assumptions, it is possible to say more. One such assumption is *rank invariance*, which posits that a person's rank in the distribution of $Y_0$ equals her rank in the distribution of $Y_1$. While strong, this assumption or variants of it has appeared in a number of settings in the literature, e.g. Chernozhukov & Hansen (2005). Under rank invariance, the distribution of treatment effects is point identified and given by

$$F_\Delta(\delta) = \int_0^1 \mathbb{1}\{F_1^{-1}(u) - F_0^{-1}(u) \leq \delta\}\, \mathrm{d}u$$

where $F_1^{-1}$ and $F_0^{-1}$ are the quantile functions of $Y_1$ and $Y_0$. See figure OA-11 in Appendix C.3. As we show in Appendix C.3, $Y_1$ first-order stochastically dominates $Y_0$ in our experiment: every quantile of the distribution of financial cost savings under forced commitment is higher than the corresponding quantile under the control condition, and the same is true for the APR outcome. Since this implies that $F_1^{-1}(u) - F_0^{-1}(u)$ is always positive, rank invariance yields an extremely stark conclusion: all borrowers benefit from forced commitment.

## 5.2 Potential Outcomes and Exclusion

Our assumption-free bounds from the preceding section show that more than 24% of borrowers would benefit from a policy of forced commitment. Intuitively, this seems to suggest that some of the 89% of borrowers in the choice arm who did *not* choose commitment would have faced lower borrowing costs if they had. Making this intuition precise, however, requires a careful consideration of the relationship between choice and heterogeneous treatment effects. To this end, we now provide a full definition of the potential outcomes in our empirical setting, and introduce a pair of assumptions that will allow us to go beyond the assumption-free bounds from above.

[19] Confidence intervals are constructed using the asymptotic distribution for the bounds. See Fan & Park (2010) for details.

Let $Z_i \in \{0, 1, 2\}$ denote the treatment arm to which to participant $i$ was assigned: $Z_i = 0$ denotes the forced no-commitment arm, $Z_i = 1$ denotes the forced commitment arm, and $Z_i = 2$ denotes the choice arm. Now let $D_i$ be the treatment that participant $i$ actually *received*, where $D_i = 0$ denotes no-commitment and $D_i = 1$ denotes commitment. We assume perfect compliance in the $Z_i = 0$ and $Z_i = 1$ arms.[20] It is only in the $Z_i = 2$ arm that participants are free to choose between alternative contracts. Let $C_i \in \{0, 1\}$ denote a participant's "choice type." If $C_i = 1$ then participant $i$ *would choose commitment*, given the option; if $C_i = 0$ she would not. As shorthand, we call borrowers with $C_i = 1$ "choosers" and those with $C_i = 0$ "non-choosers." Whereas a participant's choice type $C_i$ is only observed if she is allocated to the choice arm ($Z_i = 2$), her treatment $D_i$ and experimental arm $Z_i$ are always observed. Given the design of our experiment, these quantities are related by

$$D_i = \mathbb{1}(Z_i \neq 2)Z_i + \mathbb{1}(Z_i = 2)C_i. \tag{2}$$

We maintain the stable unit treatment value assumption (SUTVA) throughout. This means that borrower $i$'s outcomes depend only on her *own* values of $Z_i$ and $D_i$, not those of any other person in the experiment. Under this assumption, a fully general model for the potential outcomes in our experiment would take the form $Y_i(d, z)$ for $d \in \{0, 1\}$ and $z \in \{0, 1, 2\}$, allowing participant $i$'s potential outcome to depend *both* on the treatment she actually receives, $D_i$, and the experimental arm to which she is assigned, $Z_i$. This model is too general, however, to point identify meaningful causal effects. For this reason, we consider two exclusion restrictions.

Before stating these restrictions, we first define some additional notation. Because our experimental design implies that any borrower with with $Z_i = 0$ has $D_i = 0$, we abbreviate the potential outcome $Y_i(d = 0, z = 0)$ as $Y_{i0}$. Similarly, since any borrower with $Z_i = 1$ has $D_i = 1$, we abbreviate $Y_i(d = 1, z = 1)$ as $Y_{i1}$. This is in keeping with our notation from section 5.1 above. Using this notation, our first exclusion restriction is given by

$$Y_i(d = 0, z = 2) = Y_i(d = 0, z = 0) \equiv Y_{i0}. \tag{3}$$

Equation 3 only restricts the potential outcomes of non-choosers, individuals with $C_i = 0$, because they are the only borrowers for whom $D_i = 0$ when $Z_i = 2$. In words, this condition assumes that every non-chooser experiences the same potential outcome regardless of whether she is assigned to the choice arm or the control arm. Similarly, our second exclusion restriction is given by

$$Y_i(d = 1, z = 2) = Y_i(d = 1, z = 1) \equiv Y_{i1}. \tag{4}$$

Equation 4 only restricts the potential outcomes of choosers, individuals with $C_i = 1$, because they are the only borrowers for whom $D_i = 1$ when $Z_i = 2$. In words, this condition assumes that

[20] For more discussion on this point, see Section 3.3 above.

every chooser experiences the same potential outcome regardless of whether she is assigned to the treatment arm or the choice arm.

Mathematically (3) and (4) have the same structure as the standard LATE exclusion restriction that $Y_i(d, z)$ depends only on $d$, not on $z$. Substantively, however, they are slightly different, given that there is no explicit reference to the "chosen" versus "forced" treatment distinction in the usual LATE setup.[21] In essence, (3) and (4) assume that being assigned a particular treatment has the same result as choosing it for yourself, provided that you are assigned the same treatment that you *would have chosen*. If the mere fact of having been given a choice has a direct effect on outcomes, one or both of our exclusion restrictions will be violated. One can imagine situations in which this might be the case. For example, even someone who would have voluntarily chosen to undergo drug rehabilitation, given the choice, might respond differently when coerced. In our empirical setting, however, both (3) and (4) are plausible. Moreover, each has testable implications that we fail to reject. For details, see Appendix F.

Under the exclusion restrictions from (3) and (4), each participant's observed outcome $Y_i$ is related to her potential outcomes $(Y_{i0}, Y_{i1})$ according to

$$Y_i = \mathbb{1}(Z_i = 0)Y_{i0} + \mathbb{1}(Z_i = 1)Y_{i1} + \mathbb{1}(Z_i = 2)\left[(1 - C_i)Y_{i0} + C_iY_{i1}\right]. \tag{5}$$

Equation 5 is the key to understanding the results that follows. As noted above, by randomly assigning $Z_i = 0$ and $Z_i = 1$ our experiment identifies the marginal distributions of $Y_{i0}$ and $Y_{i1}$ in the population as a whole. By randomly assigning $Z_i = 2$, it likewise point identifies the share of choosers ($C_i = 1$), the distribution of $Y_{i1}$ for choosers, and the distribution of $Y_{i0}$ for non-choosers ($C_i = 0$). Because $Z_i$ is assigned independently of pre-treatment covariates $X_i$, our design also identifies the corresponding *conditional* distributions of $Y_{i0}$ and $Y_{i1}$ given $X_i$.

### 5.3 The "Controlled Choice" Design

We now show how our experimental design, henceforth the "controlled choice design," combined with the exclusion restrictions from (3) and (4) can be used to point identify a number of interesting and economically-relevant causal quantities without the need for additional structural restrictions. First we identify the treatment on the treated (TOT) and treatment on the untreated (TUT) effects, defined as follows:

$$\text{TOT} \equiv \mathbb{E}(Y_{i1} - Y_{i0}|C_i = 1), \quad \text{TUT} \equiv \mathbb{E}(Y_{i1} - Y_{i0}|C_i = 0).$$

If the TUT is positive, this means that those who did not choose commitment would have experienced better outcomes, *on average*, if they had. In a canonical Roy model, the TOT should exceed both the TUT and average treatment effect (ATE). If the TOT is statistically distinguishable

[21] For related discussion, see Chamberlain (2011) who uses an assumption similar to our exclusion restrictions to develop a theory of optimal treatment choice for an individual who has access to data from an RCT.

from and substantially larger than the TUT, this provides empirical support for the relevance of selection-on-gains in real-world decision-making. Because our design identifies all three quantities (as shown in Appendix Appendix E.), it allows us to test this implication directly and to calculate the average selection on gains (ASG):

$$\text{ASG} \equiv \mathbb{E}[Y_{i1} - Y_{i0}|C_i = 1] - \mathbb{E}[Y_{i1} - Y_{i0}|C_i = 0] = \text{TOT} - \text{TUT}.$$

The controlled choice design also identifies both the average selection bias (ASB) and the average selection on levels (ASL).

$$\text{ASB} \equiv \mathbb{E}[Y_{i0}|C_i = 1] - \mathbb{E}[Y_{i0}|C_i = 0], \quad \text{ASL} \equiv \mathbb{E}[Y_{i1}|C_i = 1] - \mathbb{E}[Y_{i1}|C_i = 0].$$

A companion STATA package accompanying this paper provides estimators of the TOT, TUT, ASG, ASB, and ASL, along with cluster-robust standard errors for each, and a test for the validity of the exclusion restriction following Huber & Mellace (2015) [22].

A number of recent papers compare estimates of the TOT and TUT to better understand who selects into treatment and why, e.g. Cornelissen *et al.* (2018) and Walters (2018). This line of work relies, at least to some extent, upon structural modeling assumptions to extrapolate from the reduced-form quantities that are identified by the data alone to more interesting, and economically relevant, causal parameters.[23] An alternative approach aims to avoid structural assumptions by calculating conditional local average treatment effects (LATE) given observed covariates $X$ and re-weighting them according to the distribution of covariates in some population of interest to yield, for example, an average treatment effect (Aronow & Carnegie, 2013; Angrist & Fernandez-Val, 2013). But there is no free lunch: this "LATE-and-reweight" approach relies upon assumptions of its own, most crucially the assumption that there is *no selection-on-gains* conditional on $X$, i.e. that the conditional LATE equals the conditional ATE. In contrast to both approaches, the controlled choice design uses exogenous experimental variation to point identify the ATE, TOT, and TUT without ruling out unobserved selection-on-gains or relying on additional structural modeling assumptions. Figure 6 provides graphical intuition for our identification approach; derivations appear in Appendix E.

The key insight can be read directly from (2) and (5). Viewing $Z_i$ as an instrumental variable, the controlled choice design can be interpreted as a *pair* of RCTs, each subject to one-sided non-compliance. The first of these compares $Z_i = 0$ to $Z_i = 2$. For each individual with $Z_i = 0$ we have $D_i = 0$ and observe $Y_{i0}$. For those with $Z_i = 2$ we have $D_i = C_i$ and observe $(1 - C_i)Y_{i0} + C_iY_{i1}$. This is identical to a "randomized encouragement" design in which treatment is only available to

[22] See Appendix Appendix E. for more details.

[23] While the marginal treatment effects (MTE) approach (Heckman & Vytlacil, 2007) can in principle be used to identify the TOT and TUT without parametric restrictions, doing so requires an instrumental variable $Z$ with sufficiently rich support that the probability of treatment take-up given $Z$ varies continuously between zero and one. In practice, instrumental variables are usually discrete and, even when continuous, typically have a more modest effect on take-up.

those who are encouraged: $Z_i = 2$. Under this interpretation, those with $C_i = 1$ are "the compliers" and it follows that

$$\frac{\mathbb{E}(Y_i|Z_i = 2) - \mathbb{E}(Y_i|Z_i = 0)}{\mathbb{E}(D_i|Z_i = 2) - \mathbb{E}(D_i|Z_i = 0)} = \frac{\mathbb{E}(Y_i|Z_i = 2) - \mathbb{E}(Y_i|Z_i = 0)}{\mathbb{E}(D_i|Z_i = 2)} = \mathbb{E}(Y_{i1} - Y_{i0}|C_i = 1) \tag{6}$$

since $\mathbb{E}(D_i|Z_i = 0) = 0$ by (2). A closely related argument can be used to construct a Wald estimand that identifies the TUT. Here we consider $Z_i = 1$ to be the "encouragement" and compare the outcomes for these individuals to those with $Z_i = 2$. If $Z_i = 1$ then $D_i = 1$ and we observe $Y_{i1}$. If instead $Z_i = 2$ then $D_i = C_i$ and we observe $(1 - C_i)Y_{i0} + C_i Y_{i1}$. Again, we can view this as an experiment with one-sided non-compliance, but now the situation is reversed. Everyone with $Z_i = 1$ is treated, but some people with $Z_i = 2$ are "always-takers" who obtain the treatment ($D_i = 1$) despite having been allocated to the "control" arm $Z_i = 2$. Under this interpretation, the "compliers" are those with $C_i = 0$: when $Z_i = 1$ they take the treatment, and when $Z_i = 2$, they do not. Thus,

$$\frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 2)}{\mathbb{E}(D_i|Z_i = 1) - \mathbb{E}(D_i|Z_i = 2)} = \frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 2)}{1 - \mathbb{E}(D_i|Z_i = 2)} = \mathbb{E}(Y_{i1} - Y_{i0}|C_i = 0) \tag{7}$$

since $\mathbb{E}(D_i|Z_i = 1) = 1$ by (2). Equations (6) and (7) are useful for understanding why the controlled choice design identifies the TOT and TUT, but they are less convenient for estimation and inference. In Appendix E., we show that

$$Y_i = \mathbb{E}(Y_{i0}) + (\text{ATE})\mathbb{1}(Z_i = 1) + (\text{TOT})\left[\mathbb{1}(Z_i = 2) \times D_i\right] + U_i \tag{8}$$

$$Y_i = \mathbb{E}(Y_{i1}) + (\text{ATE})\left[-\mathbb{1}(Z_i = 0)\right] + (\text{TUT})\left[-\mathbb{1}(Z_i = 2) \times (1 - D_i)\right] + V_i \tag{9}$$

where $\mathbb{E}(U_i|Z_i) = \mathbb{E}(V_i|Z_i) = 0$. It follows that a pair of just-identified, linear instrumental variables regressions can be used to estimate and carry out inference for the ATE, TOT, and TUT. To identify the ATE and TOT, run IV on (8) with instruments $\mathbb{1}(Z_i = 1)$ and $\mathbb{1}(Z_i = 2)$. An F-test based on this regression can then be used to test the restriction ATE = TOT, and the usual IV output can be used to carry out inference for the TOT. Similarly, to identify the ATE and TUT run IV on 9 with instruments $\mathbb{1}(Z_i = 0)$ and $\mathbb{1}(Z_i = 2)$. An F-test based on this regression can then be used to test the restriction ATE = TUT, and the usual IV output can be used to carry out inference for the TUT.

Because they identify both the TOT and TUT, (8) and (9) also identify the average selection on gains: ASG = TOT − TUT. Carrying out inference for this quantity is a bit more involved, because ASG is a difference of coefficients from two separate IV regressions. In Appendix E. we show how to use the residuals from (9) and (8) to carry out cluster-robust inference for ASG, a procedure that is automated in our companion STATA package. As mentioned above, the controlled choice design also identifies the average selection bias (ASB) and average selection on levels (ASL). In particular,

$$\text{ASB} \equiv \mathbb{E}(Y_{i0}|C_i = 1) - \mathbb{E}(Y_{i0}|C_i = 0) = \frac{\mathbb{E}(Y|Z=0) - \mathbb{E}(Y|Z=2, D=0)}{\mathbb{E}(D|Z=2)} \tag{10}$$

$$\text{ASL} \equiv \mathbb{E}(Y_{i1}|C_i = 1) - \mathbb{E}(Y_{i1}|C_i = 0) = \frac{\mathbb{E}(Y|Z=2, D=1) - \mathbb{E}(Y|Z=1)}{1 - \mathbb{E}(D|Z=2)} \tag{11}$$

as shown in Appendix E. For purposes of inference, both the ASB and ASL can be re-written as the difference of coefficients from a pair of just-identified linear IV regressions. For details, along with a description of how our companion STATA package carries out cluster-robust inference for these quantities, see Appendix E.

Table 7 calculates the causal quantities described above–TOT, TUT, ASG, ASB, and ASL–for our experimental data, along with robust standard errors for each. For purposes of comparison, the table also presents the ATE results from Section 4 above (row 1), along with the corresponding average potential outcomes $\mathbb{E}[Y_0]$ and $\mathbb{E}[Y_1]$ (rows 4–5). The columns of the table correspond to different outcome variables, defined as in Section 2.3 above. For all four outcome definitions, the TUT effect is positive, statistically and economically significant, and comparable in magnitude to the ATE. In other words: commitment is *beneficial*, on average, to the people who *would not choose it*, and these benefits are large. The estimated TOT effects are positive and larger still. Although we cannot quite reject the null hypothesis that the TUT and TOT effects are equal–our power is limited because of the low take-up rate of commitment in the choice arm–the pattern of coefficients is consistent with some degree of selection-on-gains. Commitment is more beneficial to people who are willing to choose it than it is to people who are not. For the APR benefit and (1 - Default) outcomes, we have sufficient precision to conclude that the average selection bias (ASB) is large and *negative*. This means that borrowers who choose commitment would have faced a *higher* APR and probability of default under the status quo contract than borrowers who do not choose commitment. Overall, Table 7 suggests that commitment works and the "right people" choose to commit: those who are most likely to benefit from it and those whose outcomes are most adverse under the status quo. At the same time, *not enough* people choose to commit: the commitment contract is beneficial on average even to those who would *not* choose it voluntarily. Below we explore possible explanations for this result.

# 6 The Case for Paternalism

## 6.1 Why does paternalism work in this context?

The behavioral literature has highlighted voluntary commitment as an attractive way of allowing the "right" people to self-select. A discussion of compulsory commitment is necessarily paternalistic, and by its nature focuses on impacts among those who would not have selected commitment voluntarily. Viewed through the lens of rational choice theory, it is unsurprising that we estimate

TOT > TUT. The argument for compulsory treatment, however, centers on the more surprising result that TUT > 0. We now investigate several potential explanations for this result. For simplicity, we focus on causal effects for the APR outcome throughout this section.

We conduct an exploratory analysis that examines four potential explanations for the positive TOT: the need to learn, time discounting, present bias, and overconfidence. To explore the last two dimensions we will use survey data for which our response rate was 78%. Appendix A.2 demonstrates that the subsample on whom we have survey data appears to be representative in terms of loan outcomes.

**Learning.** A first explanation involves learning. Our experiment introduced a new contract into an environment that had not previously featured commitment; perhaps clients required experience to understand its benefits. Given the strongly positive impact of forcing, this explanation would suggest that clients who were forced once would subsequently choose commitment if given the chance to do so. A subset of 16% clients from our sample returned to pawn again on another day before the end of the experiment. Whereas all of the analyses from above restrict attention to the first day on which a given individual pawned, Table OA-6 presents information about participants' *immediate subsequent* pawning behavior.[24] Column (1) considers the 228 clients who returned (only a second time) to pawn again at a day/branch that was randomly assigned to the choice arm. Each of the two rows in this column presents a difference of mean commitment take-up rates, and associated standard error. The first row compares those who were *initially* assigned to forced commitment against those where were assigned to control; the second row compares those who were initially assigned to the choice commitment arm to those who were assigned to the other two arms. In each case, there is no statistically discernible difference in the rates of commitment take-up. Granted, this is a selected sample because the decision to pawn again is potentially endogenous to the initial treatment allocation. For this reason, Column (2) considers the full sample of 4436 borrowers by re-defining the outcome variable to be an indicator for returning to pawn again at a branch/day when commitment was offered *and* choosing commitment. This composite outcome variable is not subject to the sample selection problem (although it is directly driven by the decision to repeat borrow). The comparison in the two rows remains the same: forced commitment versus control in row one and choice commitment versus forced arms in row two. Again, there is no statistically discernible difference in commitment take-up rates in either row. While these exercises cannot completely exclude the possibility that learning plays a role, they provide no indication that the lack of voluntary compliance is simply a matter of inexperience with commitment.

**Time discounting.** Discounting is a second potential explanation for our results. Our estimated decrease in the financial cost of credit from above ignores discounting, but commitment involves incurring up-front costs (early payment) in return for delayed benefits (a higher probability of

[24]In other words, for borrowers who returned to pawn again more than once, this analysis considers only their first repeat pawn.

recovering one's pawn). Highly impatient individuals might therefore rationally choose the *status quo* contract, despite the benefits that commitment yields in terms of raw (undiscounted) returns. To investigate this explanation we calculate the net present value (NPV) of the financial cost TUT effect under different hypothetical discount rates, given the timing of repayment and pawn recovery. Figure OA-9 presents the results of this exercise. The solid line gives the TUT effect adjusted for a specified annual discount rate, while the shaded regions gives the associated 95% & 90% confidence interval. We see that non-choosers continue to experience significant decreases in NPV financial costs up to annual discount rates of 4,000%, and the NPV remains positive, although not significant, at 10,000% discount rates. As such, discounting is unlikely to explain why those who benefit, on average, from commitment fail to choose it when offered.

**Present bias.** If the benefits of commitment among non-choosers cannot be explained by standard models of rational choice, the canonical behavioral story would center on time inconsistency. While commitment is useful to anyone with hyperbolic time preferences, only those who are sophisticated–i.e. aware that they are hyperbolic discounters–will demand it. A large share of "naïve" hyperbolics in the population–borrowers who are unaware that they are hyperbolic discounters–could therefore drive a large and positive TUT. Our baseline survey included standard questions about discount rates between today and a month in the future versus discount rates between three and four months in the future. This allows us to classify borrowers who display more impatience over immediate delays as present biased. This measure of financial hyperbolicity is widely used in survey research, although it is not without problems.[25]

If we could perfectly measure present bias, we could divide the sample into three groups: sophisticated hyperbolics (who chose commitment), time-consistent non-choosers (for whom forcing will not be effective), and naïve hyperbolic non-choosers (who will benefit from forced commitment). If present bias fully explains the low take-up rate of voluntary commitment, we should find that the TUT for present biased borrowers is positive while the TUT for all other borrowers is not. This is because TUT effects already restrict attention to borrowers who would not choose commitment. Within this group, a comparison of present-biased borrowers against everyone else is a comparison of naïve hyperbolics against time-consistent non-choosers. The right panel of Figure 7 carries out a feasible version of this exercise using our survey measure of present bias. The overall TUT estimate for all borrowers who answered our present-bias survey questions is given in blue, along with a 95% confidence interval. The corresponding TUT estimate and confidence interval for present-biased borrowers is given in green; results for all other borrowers are shown in red. By the law of iterated expectations, the blue overall estimate must lie between the green and red sub-group estimates. In

[25] Our measure is dichotomous, and it is not incentivized. Recent empirical work has shown the superiority of more elaborate measures such as "convex time budgets" (Andreoni *et al.*, 2015) while questioning the interpretation of measures of hyperbolicity that are not based on consumption (Andreoni & Sprenger, 2012; Cohen *et al.*, 2020), suggesting that real effort tasks provide a better measure (Augenblick *et al.*, 2015). Given that we had only a few minutes to interview real pawnshop clients prior to a commercial transaction, our simple measure was a necessary compromise.

fact, all three estimates are positive and similar in magnitude, although imprecisely estimated given the extent of survey non-response for these questions. Taking our survey measure of hyperbolicity at face value, we find no indication that present-bias explains our positive estimated TUT.

**Sure confidence.** While 72% of survey respondents believe they have a 100% chance of recovering their pawn, in reality only 48% will go on to do so. This suggests a borrower pool characterized by over-optimism. Incorrect expectations about recovery probabilities could explain low take-up if individuals who *believe* that they are certain to repay choose, rationally given their incorrect beliefs, to forgo the costs associated with commitment that are designed to induce repayment. We now explore whether over-optimistic expectations of recovery probability can explain our positive overall TUT estimate. To do this, we carry out an analysis that is analogous to our present bias exercise from the preceding paragraph, comparing the overall TUT estimate to estimates computed for two sub-groups. Here, however, the groups are defined by a binary variable that we call "sure-confidence." This measure equals one for any individuals who say at the time of borrowing that they have a 100% probability of recovering their pawn, zero otherwise. The right panel of Figure 7 presents the results for this exercise. The overall TUT estimate for all borrowers who answered our sure-confidence survey questions is given in blue, along with a 95% confidence interval. The corresponding TUT estimate and confidence interval for sure-confident borrowers is given in green; results for all other borrowers are shown in red. As above, the blue overall estimate must lie between the green and red sub-group estimates by the law of iterated expectations. In contrast to our results for present bias, shown in the left panel of Figure 7, we estimate a higher TUT effect for sure-confident borrowers compared to all other borrowers and a negative, although insignificant, effect for borrowers who are not sure-confident.

As discussed above, our measure of hyperbolicity is based on un-incentivized responses in a short survey and so is likely to be noisy; nonetheless we see no evidence here that it drives the forcing effect. Rather, the analysis from Figure 7 suggests that the effectiveness of paternalism in our experiment may be driven by a group of *overconfident* borrowers who, heedless of the risk of default, fail to choose commitment despite benefiting substantially when they are forced to commit.

If the attribute of "sure confidence" proves so important in explaining the positive TUT, what are its determinants? In Figure OA-12 we plot the coefficient estimates from a regression that predicts sure confidence with a battery of individual-level characteristics. Older males are more likely to be sure-confident, as are those with more education. Taken at face value, the sure-confident also appear to face less financial stress, to have less trouble paying bills, and to be more frequently relied upon financially by family members. Viewed through a behavioral lens, however, it is also possible that the type of person who is over-confident in their ability to repay a loan also maintains fictions in other domains of their financial life, providing answers to the baseline survey questions that exaggerate their degree of economic security. In any case it appears that sure confidence may be difficult to predict with easily-observed and objective demographic criteria, a point to which we return below.

### 6.2 Analyzing Choice versus Paternalism Using Causal Forests

**Causal forests.** On average, the commitment contract benefits both those who would choose it and those who would not. In Section 5.1, we briefly went *beyond* average effects by presenting bounds on the distribution of individual treatment effects. Because they made no assumptions beyond random assignment, these bounds were relatively wide. Adding the assumption of rank invariance yielded a distribution of treatment effects that was bounded below by zero, implying that *everyone* would benefit from treatment and hence that there would be *no losers* from paternalism. Rank invariance, however, is an extremely strong assumption. In this section, we explore a middle way between the two extremes, by considering conditional average treatment effects and conditional TOT and TUT effects, given our survey measures (note that this approach also imposes the exclusion restriction within each leaf of the tree). This exercise provides more fine-grained information about treatment effect heterogeneity. Among other things, it will potentially allow us to identify groups of borrowers who are on average *harmed* by commitment. Under the stronger assumption that our observed survey measures capture the main sources of treatment effect heterogeneity, this exercise will allow us to approximate individual-level counterfactuals, to consider whether particular borrowers made "mistakes" in their choice of contract.

To estimate conditional average treatment effects given administrative and survey data, we use the function `causal_forest()` of the `grf` R package; to estimate conditional TOT and TUT effects we use the function `instrumental_forest()` function from the same package. In each case, we use the default parameter values from the `grf` package with one exception: we increase the number of trees from the default value of 2000 to 5000. The functions `causal_forest()` and `instrumental_forest()` implement special cases of the "generalized random forest" methods of Athey *et al.* (2019). In broad strokes, these functions combine a large number of regression trees that recursively partition the covariate space to estimate conditional average effects. Figure OA-13 illustrates the partition that emerges from one of the trees in our causal forest implementation. The trees are "honest" in that observations used to determine the optimal partition are not used to estimate effects, and vice-versa. While closely related to more familiar "regression-tree" random forests, the generalized random forest approach explicitly targets the parameter of interest–a conditional ATE or IV estimand–when choosing the optimal covariate partition.[26] When constructing our random forest estimates of heterogeneous treatment effects, we use observations for all borrowers who answered at least *part* of the intake survey. We impute the median response for the missing values, while also including an indicator whether the variable originally had a missing value.[27]

Figure 8 plots densities of the estimated conditional ATE, TOT, and TUT effects from the

[26] For more details, see Athey *et al.* (2019) and the `grf` documentation: `https://grf-labs.github.io/grf/`.

[27] Results are similar if we manually include interactions between the original/imputed variable and an indicator for missingness. This is as expected, given that tree-based methods by their nature "automatically" consider interactions of arbitrary orders.

generalized random forest models described above. In each case, the outcome variable is APR benefit, i.e. the reduction in APR from a commitment contract. As we see from the figure, the conditional average effects are overwhelmingly positive. The TUT density is particularly interesting for the question of paternalism since, as emphasized above, it presents conditional average effects for borrowers who would not voluntarily choose commitment. Only 13% of our estimated conditional TUTs are negative, with a 95% confidence interval of [11%, 17%].[28] To be clear, this is a probability statement about conditional average effects over the distribution of *covariates*. In particular, we estimate that

$$\int \mathbb{1}\{\mathbb{E}[Y_1 - Y_0 | X = \mathbf{x}, C = 0] > 0\} \, f(\mathbf{x}|C = 0) \, \mathrm{d}\mathbf{x}$$

is approximately 0.13. The share of non-choosers with a positive conditional *average* treatment effects need not equal the share with a positive *individual* effects, i.e. $\mathbb{P}(Y_1 > Y_0 | C = 0)$. But the more treatment effect heterogeneity that $X_i$ explains, the closer these two values become. Figure 8 strengthens our argument, introduced in section 5, that not enough borrowers choose commitment.

Under the assumption that our instrumental forest estimates capture the main sources of treatment effect heterogeneity, we can use them to assess whether particular borrowers in the choice arm made "mistakes" in their decision to accept or refuse the commitment contract, in terms of predicted financial costs of the loan. To do this we use the same information that is depicted in Figure 8, but present it in a different way. For each non-chooser in the choice arm, we use our instrumental forest from above to estimate the conditional TUT effect, given her observed covariates. Of course conditional average effects need not equal individual treatment effects, and our APR outcome may not capture all of the costs and benefits that are relevant for individual borrowers' decisions. To account for this, we define a "mistake" for a non-chooser to be a conditional TUT estimate that significantly exceeds some large and positive APR threshold, e.g. 25%. At any such threshold, we can calculate the percentage of non-choosers in the choice arm who made a "mistake" by not choosing commitment. (If treatment it beneficial, it should be chosen.) The results of this exercise can be read off from the green curve in Figure 9. Defining $F_{\text{TUT}}(\delta)$ to be the CDF corresponding to the density of conditional TUT estimates from Figure 8, the green curve in Figure 9 is merely $[1 - F_{\text{TUT}}(\delta)] \times 100\%$. In other words, the green curve gives the percentage of non-choosers who made a "mistake" when mistakes are defined at a particular APR threshold. The green shaded region gives associated 95% pointwise confidence bounds.

For choosers we follow an analogous approach, defining a "mistake" as a *negative* conditional TOT effect that exceeds a particular APR threshold. (If treatment is harmful, it should not be chosen.) The results for choosers can be read from the the red curve in Figure 9. If $F_{\text{TOT}}(\delta)$

[28] The generalized random forest approach of Athey *et al.* (2019) produces conditional average effect estimators that are asymptotically normal, and includes methods for computing correct standard errors. These methods are relatively straightforward because the trees are "honest" in that observations used to determine splits in the recursive partitioning algorithm are not used for causal effect estimation and vice-versa. Our inferences in this section are carried out by "bootstrapping the limit experiment," i.e. simulating from the normal limit distributions using the estimated standard errors.

denotes the CDF corresponding to the density of conditional TOT estimates from Figure 8, then the red curve is merely $F_{\text{TOT}}(-\delta) \times 100\%$. In other words, the red curve gives the percentage of choosers who made a "mistake" when mistakes are defined at a particular APR threshold. The red shaded region gives associated 95% pointwise confidence bounds. Note that we use a *positive* APR threshold to denote a mistake for both choosers and non-choosers. This ensures that bigger mistakes are always to the *right* of smaller mistakes for both the green and red curves. The blue curve in Figure 9 gives the *overall* percentage of borrowers in the choice arm who made a "mistake" at a particular APR threshold. This total is computed by taking a weighted average of the green (non-choosers) and red (choosers) curves, with weights equal to their shares in the choice arm.[29]

The results in Figure 9 suggest that a large fraction of non-choosers made mistakes by not choosing commitment. Even at an APR threshold as large as 25%, we estimate that more than half of them should have chosen commitment in order to lower financial costs. In contrast, relatively few choosers appear to have made mistakes by choosing commitment. This now allows us to make a stronger statement in favor of paternalism; not only does forced commitment generate large benefits on average, but it also benefits the vast majority of borrowers *who would be coerced* under a policy of forced commitment.

### 6.3 Can we target paternalism?

Despite having shown such a large majority of individuals benefit from paternalism, it is still natural to ask whether we are able to target commitment in such a manner as to only force it upon those who benefit. For a financial firm to engage in this type of targeting under real-world circumstances, we must impose several constraints on the targeting rule. First, it will in general not be possible to use the numerous subjective and unverifiable questions we asked in the baseline survey as inputs to the rule because the answers to these questions can be manipulated by the clients, and would likely change when they become incentivized. This leaves us with only a few objective covariates that could be used to target: age, gender, HS education or above, desired loan size, and whether that individual has ever pawned before. We call these the "narrow" covariate set below, to contrast with the full set of survey variables, which we call the "wide" covariate set. Secondly, it will not be attractive for a commercial firm to ask individuals whether they want to voluntarily accept commitment and then to force upon those who say no against their will. So the choice variable itself cannot be used to target other than in a voluntary program. For this reason, the relevant causal effect for this exercise is the conditional average treatment effect (ATE) rather than the conditional TUT or TOT as considered above. As shown in Figure 10, an estimated 92% of borrowers, with a 95% confidence interval of [89%, 94%] have a positive conditional ATE as estimated from the causal forest from section 6.2 above. In the remainder of this section we ask how accurately it is possible to identify these borrowers.

[29]The blue curve is very similar to the green curve because 89% of borrowers in the choice arm are non-choosers.

**Narrow inputs targeting.** To consider targeting effectiveness we must begin from an individual-specific ground truth, which we take from our estimated conditional average treatment effects (CATEs) $\widehat{\text{ATE}}(X_i)$ from the causal forest described in Section 6.2, which used the "wide" set of covariates (wide RF) and data for all borrowers who replied to at least part of the intake survey. We then analyze how well we can predict the losers from paternalism using the "narrow" set of non-manipulable covariates listed above (narrow RF) for the same subset of borrowers. Figure 11 shows the relationship between the CATEs estimated from the RF using the "wide" and "narrow" covariate sets. It is visually obvious from this figure that we generate substantially less heterogeneity when using the narrow covariate set, although we still reject the null hypothesis of no treatment effect heterogeneity, following the approach of Chernozhukov *et al.* (2018).[30]

**Targeting assessment.** To more directly investigate our ability to target commitment, we use two different methods to predict who should be assigned to the commitment arm. In each, we estimate a classification model using the "narrow" set of covariates to predict whether the CATE from the wide random forest model is positive. The first method uses a simple logistic regression, while the second uses a random forest classification model.[31] Each of these methods yields an estimated probability of a positive CATE, allowing us to rank borrowers from highest to lowest probability. Since 92% of borrowers have a positive CATE as estimated from the wide random forest model from above, we consider a decision rule that assigns a borrower to forced commitment if her estimated probability is in the top 92% of the sample. Figure 12 compares the in-sample performance of these targeting rules against a policy of universal forced commitment. The plot presents the estimated proportion of borrowers who benefit from a particular treatment assignment rule, along with associated 95% confidence bands, where "benefit" is defined as having a conditional average treatment effect above a particular threshold. As above, we take the conditional average effects from the "wide" causal forest as ground truth for the purposes of this exercise. Because this is an in-sample exercise, the logistic and random forest classification rules have an unfair advantage: there is no adjustment for overfitting. In spite of this, their performance is quite unimpressive relative to a policy of universal forced commitment. The fraction of the whole sample incorrectly targeted when moving from universal forced commitment to targeting based on the narrow RF falls only from 8.07% to 5.49%. The logit targeting rule actually makes things marginally worse. Table 8 shows error rates for six possible assignment rules: assigning all borrowers to control, all to treatment (Forcing), the optimal (infeasible) assignment, narrow RF targeting, logit targeting, and we evaluate ex-post the voluntary choice. While the narrow RF correctly assigns roughly half of those who do not benefit from commitment to control, it also incorrectly allocates 3.7% of the sample that would have gained from treatment to control. As such, it only improves the correct targeting rate by about a half of a percentage point on net relative to universal Forcing. The logit

[30] See Appendix D. for details of this method.

[31] We use the STATA package *rforest* and use the default parameters for the estimation. For further detail see Schonlau & Zou (2020).

assignment rule is less successful at predicting benefits and harms, with a higher share of borrowers incorrectly assigned to both treatment and control. The underlying cause of the weak performance of targeting is that the attribute that predicts gains from forcing (being sure confident and not choosing the commitment) is largely behavioral and not easily predicted with basic covariates, as seen in Figure 7. Self-targeting through choice proves to be little better than assigning everyone to the control condition, given the low take-up rate and the presence of both Type I and Type II errors in the choice arm. Fully 79% of individuals in the Choice arm made a "mistake", weighting the error rates among choosers and non-choosers by their relative frequency in the Choice arm. The takeaway is that given low take-up, the large fraction of the sample benefiting from commitment, and the weak predictive power of the narrow covariates, in this case universal paternalism appears to be an attractive contract.

# 7 Conclusion

Laibson (2018) discusses the idea of "veiled paternalism", whereby principals embed forms of commitment into their products but mask this fact from consumers who may need but do not desire commitment. In the context of pawnbroking, over-collateralization means that lenders stand to make more money from unreliable borrowers than those who repay. The fact that a simple change to contract terms results in a substantial financial saving for borrowers implies that the pawn contract involves "veiled non-paternalism": features that lead to high borrower costs are embedded in non-obvious ways. Potentially due to the nature of the borrower pool, voluntary commitment choice does not result in substantial improvements in borrower returns; a mandated reform to the contract induced significant cost savings for the overall group of clients.

We are able to arrive at a nuanced set of conclusions about the relationship between take-up and heterogeneity in returns due to a novel experimental setup, the "Controlled Choice" design. This three-armed experiment, featuring a control group, an arm with borrower choice, and an arm with compulsory treatment, allows us to point-identify the impact of the treatment on those who would naturally choose it as well as those who only experience commitment when forced. While we find evidence of selection on gains (as would be expected from standard economic models) there remain substantial benefits of treatment even for non-choosers. Given that the rate of voluntary commitment in our sample is only 11%, in order to achieve widespread benefits in this context compulsory commitment may be necessary. In the Forced arm the APR (inclusive of the cost of default) falls from 184% to 150%, and the fraction of borrowers defaulting on loans drops by 6.5 pp, or 15%. Clients appear to prefer the forced commitment after they have experienced it, in that assignment to this arm increases the fraction of individuals who return to pawn again.

Why do borrowers leave such substantial returns on the table? Our investigation of treatment effect heterogeneity suggests that over-optimism is the characteristic most strongly associated with benefiting from the intervention and yet not choosing it. Overall the borrower pool overestimates their own probability of repayment by more than 50%, and we find that the positive TUT is almost entirely confined to borrowers who report that they are certain to repay. The takeaway is that inefficiently low demand for tools to remediate default is strongest among those who incorrectly believe that they have no chance of defaulting. More standard explanations such as discounting, learning, or time inconsistency find little support in our data. Importantly, we find that not only are the benefits of commitment close to universal, but the determinants that might allow us to target it more finely are primarily behavioral and hence not easily predicted in an fast incentive-compatible manner by lenders. Even a sophisticated machine learning-based exercise is only able to decrease the fraction of borrowers mis-targeted by about 2.5 percentage points relative to simply assigning everyone to commitment (from 8% to 5.5%), and a logit-based targeting rule actually does slightly worse than universal commitment.

This result has immediate policy implications. Pawnshops, along with other over-collateralized credit products such as payday lending, exist in an environment where the bank may desire customers to lose their collateral on the loan. Given the high prevalence of the use of this form of credit by low-income households, this has an obviously detrimental effect on the dynamics of asset accumulation among the poor. Given that the specific items being pawned (such crucifixes and family jewelry) may have sentimental, non-pecuniary meaning to households, the efficiency implications of these contract terms may be even larger (as evidenced by the fact that borrowers' subjective value of their own pawns was 26% higher than the appraised collateral value). With a now well-established toolkit of regular small payments and incentives delivering vanishingly small default rates in microfinance, regulators in the banking sector may fruitfully investigate the possibility of requiring pawnbrokers to embed features of commitment and regularity into their repayment structures in more consistent ways. If employed at scale in a competitive lending sector this could redistribute welfare from those who would have repaid (whose interest rates must now rise to cover lower returns from collateral seizure) towards those who would only repay in the presence of commitment. In a setting of lender market power however, some redistribution from lenders to borrowers could occur.

Where banks have no incentive to engage in veiled paternalism and customers display inefficiently low demand for it, financial policy regulation may prove an attractive option. An important question for future research will be the extent to which borrowers are able to learn about the benefits of commitment over time, making it so that temporary, lighter-touch policies could achieve lasting benefits for borrowers. Pawning with commitment may provide an important mechanism to preserve flexible credit access while allowing more poor borrowers to retain their assets.

## References

Abdulkadiroğlu, Atila, Angrist, Joshua D, Dynarski, Susan M, Kane, Thomas J, & Pathak, Parag A. 2011. Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *The Quarterly Journal of Economics*, **126**(2), 699–748.

Andreoni, James, & Sprenger, Charles. 2012. Estimating time preferences from convex budgets. *American Economic Review*, **102**(7), 3333–3356.

Andreoni, James, Kuhn, Michael A, & Sprenger, Charles. 2015. Measuring time preferences: A comparison of experimental methods. *Journal of Economic Behavior & Organization*, **116**, 451–464.

Angrist, Joshua D, & Fernandez-Val, Ivan. 2013. ExtrapoLATE-ing: External Validity and. *Page 401 of: Advances in Economics and Econometrics: Tenth World Congress*, vol. 3. Cambridge University Press.

Aronow, Peter M, & Carnegie, Allison. 2013. Beyond LATE: Estimation of the average treatment effect with an instrumental variable. *Political Analysis*, **21**(4), 492–506.

Ashraf, Nava, Karlan, Dean, & Yin, Wesley. 2006. Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines*. *The Quarterly Journal of Economics*, **121**(2), 635–672.

Athey, Susan, Tibshirani, Julie, & Wager, Stefan. 2019. Generalized random forests. *Ann. Statist.*, **47**(2), 1148–1178.

Augenblick, Ned, Niederle, Muriel, & Sprenger, Charles. 2015. Working over time: Dynamic inconsistency in real effort tasks. *The Quarterly Journal of Economics*, **130**(3), 1067–1115.

Bai, Liang, Handel, Benjamin, Miguel, Edward, & Rao, Gautam. 2020. Self-Control and Demand for Preventive Health: Evidence from Hypertension in India. *Review of Economics and Statistics, Forthcoming.*

Basu, Kaushik. 1984. Implicit interest rates, usury and isolation in backward agriculture. *Cambridge Journal of Economics*, **8**(2), 145–159.

Bauer, Michal, Chytilová, Julie, & Morduch, Jonathan. 2012a. Behavioral Foundations of Microcredit: Experimental and Survey Evidence from Rural India. *The American Economic Review*, **102**(2), 1118–1139.

Bauer, Michal, Chytilová, Julie, & Morduch, Jonathan. 2012b. Behavioral foundations of microcredit: Experimental and survey evidence from rural India. *American Economic Review*, **102**(2), 1118–39.

Bertrand, Marianne, & Morse, Adair. 2011. Information Disclosure, Cognitive Biases, and Payday Borrowing. *The Journal of Finance*, **66**(6), 1865–1893.

Borjas, George J. 1989. Economic theory and international migration. *International migration review*, **23**(3), 457–485.

Brune, Lasse, Giné, Xavier, Goldberg, Jessica, & Yang, Dean. 2016. Facilitating savings for agriculture: Field experimental evidence from Malawi. *Economic Development and Cultural Change*, **64**(2), 187–220.

Callen, Michael, De Mel, Suresh, McIntosh, Craig, & Woodruff, Christopher. 2019. What are the headwaters of formal savings? Experimental evidence from Sri Lanka. *The Review of Economic Studies*, **86**(6), 2491–2529.

Carter, Susan Payne, & Skiba, Paige Marta. 2012. Pawnshops, behavioral economics, and self-regulation. *Rev. Banking & Fin. L.*, **32**, 193.

Casaburi, Lorenzo, & Macchiavello, Rocco. 2019. Demand and Supply of Infrequent Payments as a Commitment Device: Evidence from Kenya. *American Economic Review*, **109**(2), 523–55.

Chamberlain, Gary. 2011. Bayesian aspects of treatment choice. *In: The Oxford Handbook of Bayesian Econometrics*.

Chernozhukov, Victor, & Hansen, Christian. 2005. An IV model of quantile treatment effects. *Econometrica*, **73**(1), 245–261.

Chernozhukov, Victor, Demirer, Mert, Duflo, Esther, & Fernandez-Val, Ivan. 2018. *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India*. Tech. rept. National Bureau of Economic Research.

Cohen, Jonathan, Ericson, Keith Marzilli, Laibson, David, & White, John Myles. 2020. Measuring time preferences. *Journal of Economic Literature*, **58**(2), 299–347.

Cornelissen, Thomas, Dustmann, Christian, Raute, Anna, & Schönberg, Uta. 2018. Who benefits from universal child care? Estimating marginal returns to early child care attendance. *Journal of Political Economy*, **126**(6), 2356–2409.

Cullen, Julie Berry, Jacob, Brian A, & Levitt, Steven. 2006. The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, **74**(5), 1191–1230.

DiTraglia, Francis J. & Garcia-Jimeno, Camilo. 2019. Identifying the Effect of a Mis-classified, Binary, Endogenous Regressor. *Journal of Econometrics*, **209**(2), 376–390.

Dupas, Pascaline, & Robinson, Jonathan. 2013. Why Don't the Poor Save More? Evidence from Health Savings Experiments. *American Economic Review*, **103**(4), 1138–71.

Fan, Yanqin, & Park, Sang Soo. 2010. Sharp bounds on the distribution of treatment effects and their statistical inference. *Econometric Theory*, **26**(3), 931–951.

Field, Erica, & Pande, Rohini. 2008. Repayment Frequency and Default in Microfinance: Evidence from India. *Journal of the European Economic Association*, **6**(2/3), 501–509.

Fowlie, Meredith, Wolfram, Catherine, Baylis, Patrick, Spurlock, C Anna, Todd-Blick, Annika, & Cappers, Peter. 2021. Default effects and follow-on behaviour: Evidence from an electricity pricing program. *The Review of Economic Studies*, **88**(6), 2886–2934.

Giné, Xavier, Karlan, Dean, & Zinman, Jonathan. 2010. Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation. *American Economic Journal: Applied Economics*, **2**(4), 213–35.

Gregg, Samuel. 2016. How Medieval Monks Changed the Face of Banking. *American Banker*, **1**(88).

Heckman, James J, & Vytlacil, Edward J. 2007. Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments. *Handbook of econometrics*, **6**, 4875–5143.

Huber, Martin, & Mellace, Giovanni. 2015. Testing instrument validity for late identification based on inequality moment constraints. *The Review of Economics and Statistics*, **97**(2), 398–411.

Kaur, Supreet, Kremer, Michael, & Mullainathan, Sendhil. 2015. Self-Control at Work. *Journal of Political Economy*, **123**(6), 1227–1277.

Laibson, David. 2018. Private Paternalism, the Commitment Puzzle, and Model-Free Equilibrium. *AEA Papers and Proceedings*, **108**(May), 1–21.

Lippman, Steven A, & McCall, John J. 1976. The economics of job search: A survey. *Economic inquiry*, **14**(2), 155–189.

Melzer, Brian T. 2011. The real costs of credit access: evidence from the payday lending market. *The Quarterly Journal of Economics*, **126**(1), 517–555.

Morduch, Jonathan. 1999. The microfinance promise. *Journal of economic literature*, **37**(4), 1569–1614.

Oreopoulos, Philip. 2006. Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, **96**(1), 152–175.

Pedroza, Paola. 2010. *Microfinanzas en América Latina y el Caribe: El sector en Cifras*. Tech. rept. Interamerican Development Bank Report.

Prina, Silvia. 2015. Banking the poor via savings accounts: Evidence from a field experiment. *Journal of development economics*, **115**, 16–31.

Royer, Heather, Stehr, Mark, & Sydnor, Justin. 2015. Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company. *American Economic Journal: Applied Economics*, **7**(3), 51–84.

Sadoff, Sally, Samek, Anya, & Sprenger, Charles. 2019. Dynamic Inconsistency in Food Choice: Experimental Evidence from Two Food Deserts. *The Review of Economic Studies*, **87**(4), 1954–1988.

Schilbach, Frank. 2019. Alcohol and Self-Control: A Field Experiment in India. *American Economic Review*, **109**(4), 1290–1322.

Schonlau, Matthias, & Zou, Rosie Yuyan. 2020. The random forest algorithm for statistical learning. *The Stata Journal*, **20**(1), 3–29.

Stegman, Michael A. 2007. Payday Lending. *Journal of Economic Perspectives*, **21**(1), 169–190.

Tarozzi, Alessandro, Mahajan, Aprajit, Yoong, Joanne, & Blackburn, Brian. 2009. Commitment Mechanisms and Compliance with Health-Protecting Behavior: Preliminary Evidence from Orissa, India. *American Economic Review*, **99**(2), 231–35.

Thaler, Richard H, & Benartzi, Shlomo. 2004. Save more tomorrow™: Using behavioral economics to increase employee saving. *Journal of political Economy*, **112**(S1), S164–S187.

Walters, Christopher R. 2018. The Demand for Effective Charter Schools. *Journal of Political Economy*, **126**(6), 2179–2223.

# 8 Tables

Table 1: Limited and balanced attrition

| | | Commitment arms | | |
|---|---|---|---|---|
| | Control | Forced | Choice | p-value |
| Number of branch-day pawns | 30 | 34 | 35 | 0.16 |
| | (2.8) | (2.9) | (1.8) | |
| Ended up pawning | 0.98 | 0.97 | 0.97 | 0.62 |
| | (0.01) | (0.01) | (0.01) | |
| Survey response rate | 0.79 | 0.76 | 0.77 | 0.62 |
| | (0.02) | (0.02) | (0.01) | |
| Obs | 1770 | 1954 | 2580 | |

Each row in this table corresponds to a regression, where the level of observation is the individual loan originated. The dependent variables of these regressions are displayed in the first column. Each dependent variable is regressed in a multivariate OLS regression against the experimental arms indicators (control, forced commitment, choice). The table reports the coefficients on each of these indicators, as well as the p-value an F-test of the null hypothesis of equality of the three coefficients. The dependent variables are: the number of pawns-loans originated per day per branch, a variable indicating whether a person who answered the baseline survey (before knowing contract terms) ended up pawning, and an indicator of whether the person that obtained the loan answered the baseline survey. By design we had the choice arm be larger.

Table 2: Summary statistics and Balance

| | Control | Commitment arms: Forced | Commitment arms: Choice | p-value |
|---|---|---|---|---|
| | | Panel A : Administrative Data | | |
| Loan amount | 2267 | 2162 | 2223 | 0.65 |
| | (76) | (83) | (66) | |
| Weekday | 0.88 | 0.89 | 0.83 | 0.56 |
| | (0.044) | (0.035) | (0.048) | |
| Obs | 1770 | 1954 | 2580 | |
| | | Panel B : Survey Data | | |
| Subjective value | 4084 | 3877 | 4173 | 0.51 |
| | (186) | (193) | (172) | |
| Trouble paying bills | 0.19 | 0.21 | 0.18 | 0.67 |
| | (0.024) | (0.023) | (0.02) | |
| Present bias | 0.14 | 0.13 | 0.13 | 0.89 |
| | (0.02) | (0.01) | (0.01) | |
| Makes budget | 0.62 | 0.59 | 0.65 | 0.29 |
| | (0.028) | (0.036) | (0.021) | |
| Subj. pr. of recovery | 91.89 | 91.65 | 93.61 | 0.09 |
| | (0.721) | (1.031) | (0.582) | |
| Pawn before | 0.87 | 0.89 | 0.9 | 0.25 |
| | 0.02 | (0.013) | (0.011) | |
| Age | 43.32 | 42.85 | 43.82 | 0.73 |
| | (0.688) | (0.949) | (0.792) | |
| Female | 0.73 | 0.72 | 0.71 | 0.88 |
| | (0.023) | (0.019) | (0.02) | |
| + High-school | 0.66 | 0.67 | 0.65 | 0.84 |
| | (0.027) | (0.022) | (0.018) | |
| Obs | 1386 | 1469 | 1982 | |

This table has two panels. Panel A uses administrative data at the loan level, while Panel B uses survey data. Each row in this table corresponds to a regression, where the level of observation is the individual loan originated. The dependent variables of these regressions are displayed in the first column. Each dependent variable is regressed in a multivariate OLS regression against the experimental arms indicators (control, forced commitment, choice). The table reports the coefficients on each of these indicators, as well as the p-value an F-test of the null hypothesis of equality of the three coefficients. The admin data was a very limited set of pre-determined variables. The dependent variables in Panel A are the loan amount in pesos, and an indicator for whether the day of the loan origination was a weekday (as opposed to weekend). The dependent variables in Panel B from the survey (see the questions in Table OA-1. Subjective value of the pawn (how much would the client be willing to sell it for (Q3), an indicator for having trouble paying bills in the last 6-months (Q28), present bias (constructed from questions Q10 and Q29 in the standard way as in Ashraf *et al.* (2006)), an indicator for whether they make expenses budget for the month ahead of time. The subjective probability of recovery was elicited a la Manski (from 0 to 100 what is the probability that you will recoup your pawn), pawned before is a dummy=1 if the client declares to have pawned before (although not necessarily with Lender $P$) age is in year, +High-school is a dummy that indicates if the client has completed high school.

Table 3: Effects on Financial Cost

| | FC | Components of FC | | | | | APR |
|---|---|---|---|---|---|---|---|
| | | Interest pymnt | Fee pymnt | Principal pymnt | Lost pawn value | Default | |
| | | $\sum_t P_{it}^i$ | $\sum_t P_{it}^f$ | $\mathbb{1}(\text{Def}_i) \times \sum_t P_{it}^c$ | $\mathbb{1}(\text{Def}_i) \times \text{Appr. Val.}_i$ | $\mathbb{1}(\text{Def}_i)$ | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Forced commitment | -379.9*** | -157.3*** | 32.1*** | -0.61 | -254.7** | -0.064*** | -0.34*** |
| | (111.4) | (34.9) | (1.43) | (3.04) | (104.7) | (0.023) | (0.081) |
| Choice commitment | -78.8 | -24.9 | 1.34** | -2.76 | -55.3 | -0.023 | -0.10 |
| | (114.5) | (38.4) | (0.54) | (2.56) | (109.1) | (0.021) | (0.074) |
| Observations | 6304 | 6304 | 6304 | 6304 | 6304 | 6304 | 6304 |
| R-squared | 0.007 | 0.022 | 0.151 | 0.003 | 0.007 | 0.013 | 0.011 |
| Control Mean | 1850.6 | 545.9 | 0 | 5.75 | 1304.7 | 0.43 | 1.83 |

This table shows the treatment effects for our core pecuniary outcomes. Each column is a different regression for different outcomes on an indicator for the forced and choice arms, following specification in equation 1. Columns (1) & (7) analyze our core financial cost measures, while the rest of the columns decompose these into finer components as follows:

$$\underbrace{\text{Financial Cost}_i}_{(1)} = \underbrace{\sum_t P_{it}^i}_{\substack{\text{Payments to interests} \\ (2)}} + \underbrace{\sum_t P_{it}^f}_{\substack{\text{Payments of fees} \\ (3)}} + \underbrace{\overbrace{\mathbb{1}(\text{Default}_i)}^{(6)} \times (\text{Appraised Value}_i}_{\substack{\text{Cost of losing pawn} \\ (5)}} + \underbrace{\sum_t P_{it}^c}_{\substack{\text{Payments to capital} \\ (4)}} )$$

A few borrowers take more than one loan on the first day they appear in an experimental branch. These are treated as different observations. Table OA-4 shows our results are similar when dealing with the multiplicity of loans per borrower. Each regression includes branch and day-of-week FE. Standard errors are clustered at the branch-day level.

## Table 4: Effects on intermediate outcomes

| | Panel A : Speed of payment | | | | |
|---|---|---|---|---|---|
| | Days to 1st payment | % of payment in 1st visit | Pr(Recovery in 1st visit) | Loan duration (days) | Loan duration \| recovery |
| | (1) | (2) | (3) | (4) | (5) |
| Forced commitment | -13.8*** | 7.94*** | 0.079*** | -27.9*** | -17.9*** |
| | (1.61) | (2.77) | (0.026) | (4.35) | (3.88) |
| Choice commitment | -3.51** | -0.78 | -0.010 | -0.18 | -1.35 |
| | (1.57) | (2.20) | (0.022) | (4.33) | (4.19) |
| Observations | 4412 | 6304 | 6304 | 6304 | 3031 |
| R-squared | 0.055 | 0.014 | 0.016 | 0.054 | 0.041 |
| Control Mean | 82.8 | 44.7 | 0.30 | 136.6 | 103.9 |

| | Panel B : Variables related to default | | | | Panel C : Visit variables | |
|---|---|---|---|---|---|---|
| | Pr(+ payment & default) | % of pay \| def | Pr(Selling pawn \| def) | Pr(Selling pawn) | # of visits | # of visits \| def |
| | (6) | (7) | (8) | (9) | (10) | (11) |
| Forced commitment | -0.069*** | -3.93*** | 0.14*** | 0.0050 | -0.031 | -0.19*** |
| | (0.015) | (1.27) | (0.034) | (0.021) | (0.049) | (0.049) |
| Choice commitment | -0.027* | -1.84* | 0.050* | 0.0035 | 0.085 | -0.082* |
| | (0.014) | (1.04) | (0.029) | (0.019) | (0.053) | (0.042) |
| Observations | 6304 | 2487 | 2487 | 6304 | 6304 | 2487 |
| R-squared | 0.011 | 0.023 | 0.033 | 0.016 | 0.022 | 0.026 |
| Control Mean | 0.12 | 9.51 | 0.72 | 0.31 | 1.14 | 0.39 |

This table explores treatment effects in "intermediate variables". Each column represents regression output for different dependent variables following equation (1). Panel A focuses on variables related to the speed of payment. While Panel B focuses on variables related to default, and Panel C related to visits. The outcome variables are as follows: number of days elapsed between origination and first payment (col 1); percentage of the loan paid in the first payment (col 2); probability of recovery in the first visit (col 3); loan duration is the number of days the borrower took to payoff her loan for those that recover, the number of days until default for those that default, and the maximum number of days we observe them in the sample for those that have not recovered or defaulted (col 4); loan duration conditional on recovery; an indicator for paying a positive amount towards recovery but nonetheless losing the pawn (col 6); the percentage or the loan paid, conditional on defaulting –'wasted payments' (col 7). Column 8 uses the phrase 'selling the pawn' for a dummy variable indicating the borrower did not pay any amount towards recovery and lost the pawn. Moreover, (col 9) shows that treatment effects are concentrated in the intensive margin as treatment does not affect the fraction of clients who pay a positive amount towards pawn recovery. The dependent variable in column 10 is the number of day-visits to the branch (measured by the existence of transactions that day associated with our particular pawn), while column 11 conditions on borrowers that lost the pawn. Each regression includes branch and day-of-week FE. Standard errors are clustered at the branch-day level.

Table 5: Effects on more comprehensive cost measures

| | FC | FC (subj.value) | FC + tc | FC - interest | FC (subj.value) + tc - int |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Forced commitment | -379.9*** | -473.9*** | -383.5*** | -272.4** | -320.3** |
| | (111.4) | (150.9) | (110.9) | (108.8) | (144.2) |
| Choice comitment | -78.8 | -96.5 | -72.6 | -71.6 | -65.4 |
| | (114.5) | (153.7) | (114.2) | (114.1) | (148.9) |
| Observations | 6304 | 6304 | 6304 | 6304 | 6304 |
| R-squared | 0.007 | 0.007 | 0.007 | 0.006 | 0.006 |
| Control Mean | 1850.6 | 2296.7 | 1934.3 | 1387.4 | 1834.4 |

| | APR | APR (subj.value) | APR + tc | APR - interest | APR (subj.value) + tc - int |
|---|---|---|---|---|---|
| | (6) | (7) | (8) | (9) | (10) |
| Forced commitment | -0.34*** | -0.62*** | -0.39*** | -0.29*** | -0.40** |
| | (0.081) | (0.17) | (0.086) | (0.086) | (0.16) |
| Choice comitment | -0.10 | -0.22 | -0.11 | -0.12 | -0.17 |
| | (0.074) | (0.16) | (0.080) | (0.081) | (0.15) |
| Observations | 6304 | 6304 | 6304 | 6304 | 6304 |
| R-squared | 0.011 | 0.009 | 0.012 | 0.011 | 0.009 |
| Control Mean | 1.83 | 3.29 | 2.03 | 1.58 | 2.86 |

This table augments the measure of financial cost presented in Table 3 with measures of transaction costs, subjective costs, and adjustments for liquidity costs. Column 1 replicates the result of column 1 of Table 3 to ease comparability. Column 2 uses the subjective value of the pawn instead of the assessed gold value. Column 3 uses the cost from column 1 but adds to it the self-reported transport cost (most use public transport, imputing missing values with the average) as well as a whole day's minimum wage as a proxy for lost time. Column 4 ignores the 'liquidity' cost in the financial cost formula by subtracting the interest of 7% per month for each peso paid. Column 5 applies all these extra costs at the same time. Columns 6 to 10 apply the analogous changes to the APR formula. Each regression includes branch and day-of-week FE. Standard errors are clustered at the branch-day level.

Table 6: Effects on Repeat Pawning

| | Ever pawns again (ITT) | | | | |
|---|---|---|---|---|---|
| | | After 90 days | Within 90 days | Different collateral | Cond. on rec |
| | (1) | (2) | (3) | (4) | (5) |
| Forced commitment | 0.067* | 0.037*** | 0.032 | 0.048* | 0.11** |
| | (0.035) | (0.013) | (0.027) | (0.029) | (0.046) |
| Choice commitment | 0.040 | 0.0098 | 0.030 | 0.036 | 0.058 |
| | (0.031) | (0.0087) | (0.026) | (0.026) | (0.041) |
| Observations | 4441 | 4441 | 4441 | 4441 | 2173 |
| R-squared | 0.003 | 0.006 | 0.001 | 0.002 | 0.008 |
| Control Mean | 0.32 | 0.020 | 0.30 | 0.29 | 0.35 |

This table estimates the specification of equation 1 but at the level of the borrower (not the loan). Each column represents a regression with a different outcome variable. In column 1, the dependent variable indicates, for each borrower in the experiment, whether he/she pawned again after the first loan in the experiment (up to the end period of our data set 338 days after the experiment began). Column 2 is analogous, but only pawning after 90 days of the first loan is considered. Column 3 instead considers pawning before 90 days. Column 4 is analogous to column 1 but focuses on the pawning of a gold piece that is different from the one in the first experimental loan. Column 5 is analogous to column 1, but conditioning on the sample that recovered the first loan. Each regression includes branch and day-of-week FE. Standard errors are clustered at the branch-day level.

Table 7: Treatment on the Treated (TOT), Treatment on the Untreated (TUT), Selection-on-gains (TOT - TUT), Average Selection Bias (ASB), and Average Selection Bias, calculated using the results from Section 5.3.

| | APR % benefit | FC benefit | % (1-Default) | % (1-Refinance) |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| ATE | 35.2*** | 389.0*** | 7.62*** | 6.34** |
| | (8.31) | (117.6) | (2.50) | (2.90) |
| ToT | 122.0* | 644.7 | 37.4* | -25.9 |
| | (68.9) | (1085.5) | (21.7) | (29.1) |
| TuT | 24.9*** | 358.5*** | 4.07* | 10.2*** |
| | (8.31) | (107.7) | (2.41) | (2.90) |
| $\mathbb{E}[Y_1]$ | -183.1*** | -1850.6*** | 56.6*** | 60.9*** |
| | (5.92) | (69.4) | (1.69) | (1.70) |
| $\mathbb{E}[Y_0]$ | -147.9*** | -1461.6*** | 64.2*** | 67.2*** |
| | (5.83) | (95.0) | (1.84) | (2.35) |
| ToT-TuT | 97.1 | 286.2 | 33.4 | -36.1 |
| | (72.3) | (1132.5) | (22.6) | (30.6) |
| ASB | -110.7 | -51.9 | -39.3* | 22.7 |
| | (71.6) | (1126.5) | (22.3) | (30.1) |
| ASL | -13.6 | 234.3 | -5.90 | -13.4*** |
| | (15.9) | (154.4) | (4.29) | (4.20) |
| Observations | 6304 | 6304 | 6304 | 6304 |
| $H_0$ : ATE-TuT=0 | 0.18 | 0.80 | 0.14 | 0.23 |
| $H_0$ : ATE-ToT=0 | 0.18 | 0.80 | 0.14 | 0.24 |
| $H_0$ : ToT-TuT=0 | 0.18 | 0.80 | 0.14 | 0.24 |
| $H_0$ : ToT-TuT$\geq$ 0 | 0.090 | 0.40 | 0.071 | 0.88 |

This table presents results computed using the derivations from Section 5.3 and Appendix E.. Each column corresponds to an outcome variable: APR in column (1), financial cost in column (2), an indicator for *not* defaulting in column (3), and an indicator for *not* refinancing in column (4). The APR and financial cost outcomes have been multiplied by $-1$ so that a positive causal effect *benefits* the borrower in each of the four columns. The top panel presents treatment on the treated (TOT) and treatment on the untreated (TUT) effect estimates, with standard errors clustered at the branch-day level. The average treatment effect (ATE) results from Section 4 are presented for purposes of comparison, along with average treated and untreated potential outcomes. The middle panel presents estimates of the average selection on gains (TOT $-$ TUT), along with the average selection bias (ASB) and the average selection on levels (ASL), again with standard errors clustered at the branch-day level. See Appendix E.3 for more details on how we compute standard errors for the middle panel. The bottom panel present sample sizes, and p-values for a number of null hypothesis tests of treatment effect heterogeneity.

Table 8: Type I & II errors using targeting narrow rules

| Rule | % incorrectly assigned to control | % incorrectly assigned to treatment | Overall Error Rate |
|---|---|---|---|
| All to control | 91.93 | 0 | 91.93 |
| All to forcing | 0 | 8.07 | 8.07 |
| Optimal | 0 | 0 | 0 |
| Narrow rule (RF) | 2.31 | 3.18 | 5.49 |
| Narrow rule (Logit) | 3.69 | 4.56 | 8.25 |
| Allow choice | 86.9 | 22.31 | 79.6 |

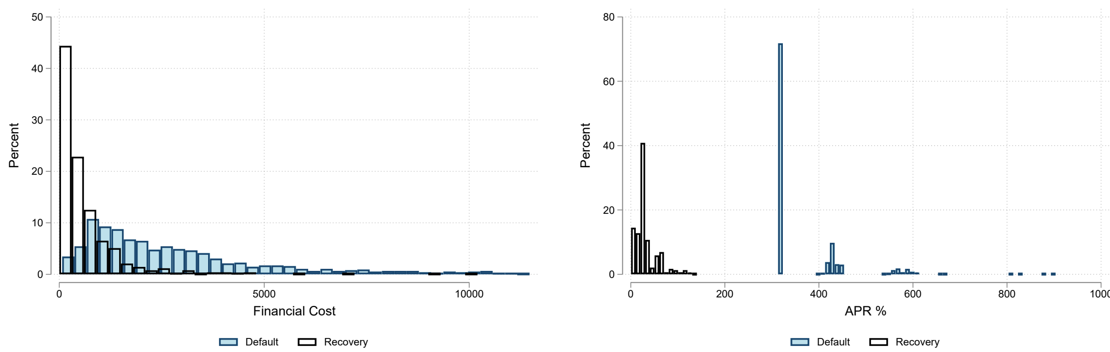
This table reports error rates for six different rules for allocating individuals to commitment. The first row allocates all borrowers to control. Taking the conditional ATE estimates from the "wide" causal forest, described in Section 6.2 as ground truth, this results in 87% of individuals losing by not receiving commitment. The second row assigns all borrowers to forced commitment, yielding error rates that are the exact mirror image of those from the first row. The third row considers the infeasible optimum in which each borrower is allocated the "correct" treatment, treating the estimate from the "wide" causal forest as ground truth. The fourth and fifth rows assign borrowers to commitment based on a classification model that uses the narrow covariate set to predict an indicator for whether the "wide" random forest CATE estimate is positive. Row four uses a random forest classification model while row five uses a simple logistic regression model. In each of these rows, the assignment rule ranks borrowers from highest to lowest by their estimated probability of having a positive CATE. The 91% of borrowers with the highest estimated probabilities are assigned to treatment, matching the overall rate of positive CATE estimates from the "wide" RF model. The sixth row gives the percentage of miss-classification that would follow from the observed choice.

# 9 Figures

Figure 1: Financial cost

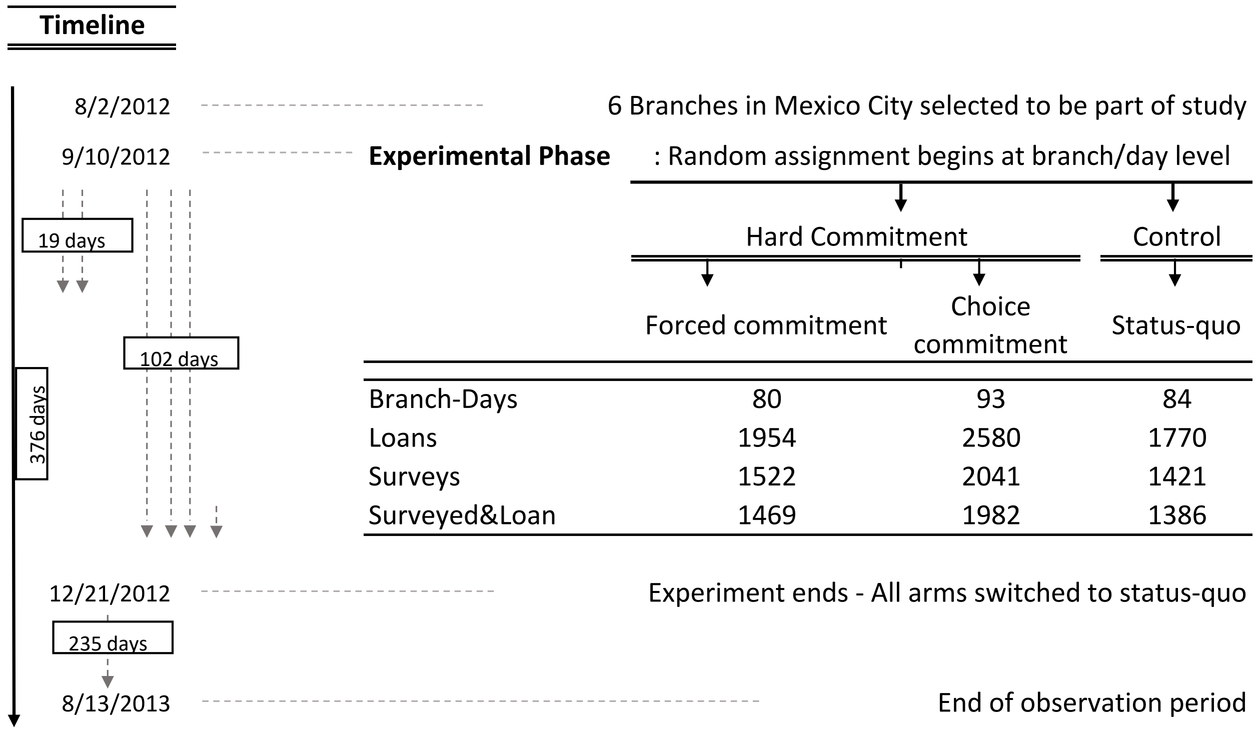
(a) In pesos MXN

(b) Effective APR



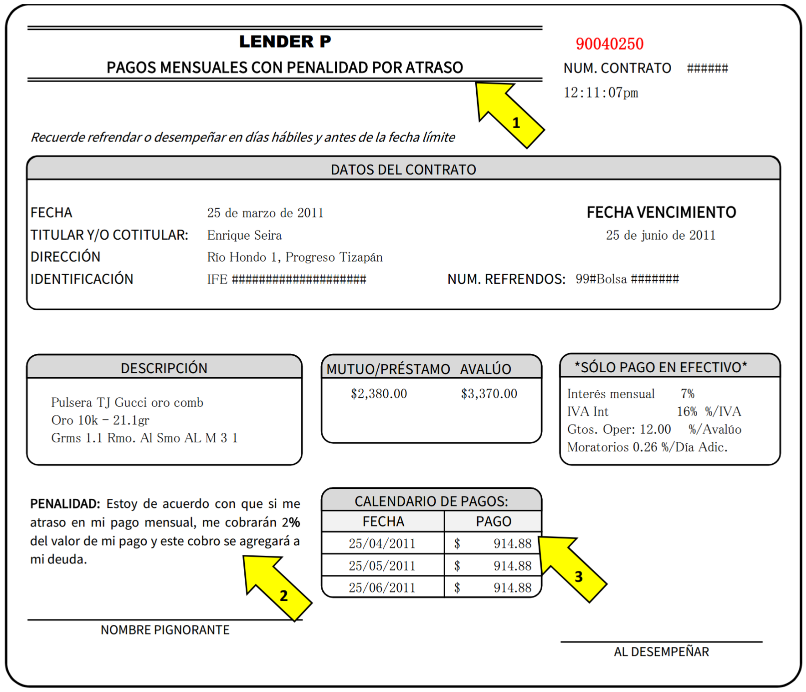

Panel (a) presents a histogram of financial cost (defined as in Section 2.3) for the control group. The figure plots separately the financial cost separately for borrowers who lose the pawn (blue) and those that recover it (transparent). Panel (b) is analogous but displays the APR.

Figure 2: Experiment description




| | Forced commitment | Choice commitment | Status-quo |
|---|---|---|---|
| Branch-Days | 80 | 93 | 84 |
| Loans | 1954 | 2580 | 1770 |
| Surveys | 1522 | 2041 | 1421 |
| Surveyed&Loan | 1469 | 1982 | 1386 |

Figure 3: Contract Terms Summary



**LENDER P**

PAGOS MENSUALES CON PENALIDAD POR ATRASO

90040250

NUM. CONTRATO ######

12:11:07pm

1

*Recuerde refrendar o desempeñar en días hábiles y antes de la fecha límite*

DATOS DEL CONTRATO

| | | |
|---|---|---|
| FECHA | 25 de marzo de 2011 | **FECHA VENCIMIENTO** |
| TITULAR Y/O COTITULAR: | Enrique Seira | 25 de junio de 2011 |
| DIRECCIÓN | Río Hondo 1, Progreso Tizapán | |
| IDENTIFICACIÓN | IFE #################### | NUM. REFRENDOS: 99#Bolsa ####### |

DESCRIPCIÓN

Pulsera TJ Gucci oro comb
Oro 10k - 21.1gr
Grms 1.1 Rmo. Al Smo AL M 3 1

| MUTUO/PRÉSTAMO | AVALÚO |
|---|---|
| $2,380.00 | $3,370.00 |

*SÓLO PAGO EN EFECTIVO*

Interés mensual 7%
IVA Int 16% %/IVA
Gtos. Oper: 12.00 %/Avalúo
Moratorios 0.26 %/Día Adic.

**PENALIDAD:** Estoy de acuerdo con que si me atraso en mi pago mensual, me cobrarán **2%** del valor de mi pago y este cobro se agregará a mi deuda.

2

CALENDARIO DE PAGOS:

| FECHA | PAGO |
|---|---|
| 25/04/2011 | $ 914.88 |
| 25/05/2011 | $ 914.88 |
| 25/06/2011 | $ 914.88 |

3

NOMBRE PIGNORANTE

AL DESEMPEÑAR

We show a sample receipt that was given to clients that got assigned to the fee-forcing contract (the font and format were changed to protect Lender's P identity). We want to highlight the salience of some items. First the title clearly indicated which contract the client has (arrow 1). Second, in the case of the fee contract it clearly indicates that there is a fee for paying late equivalent to 2% of the value of the monthly payment (arrow 2). Third, there is a calendar for payments clearly specifying the dates and amounts to pay each month.

Figure 4: Explanatory Material

**Status-quo Contract**

**Pay when you want (before 3 months)**

✓**Term**: loan must be paid **before 3 months**.

✓**Amount owed**: Loan + Accumulated interest before loan term ends**.** Interest accumulates daily on outstanding amount.

✓**Flexibility**: you can pay any quantity at any time before 90 days **with no prepayment penalty.**

**Forced-commitment Contract**

**3 mandatory monthly payments**

✓**Term**: loan must be paid **before 3 months**.

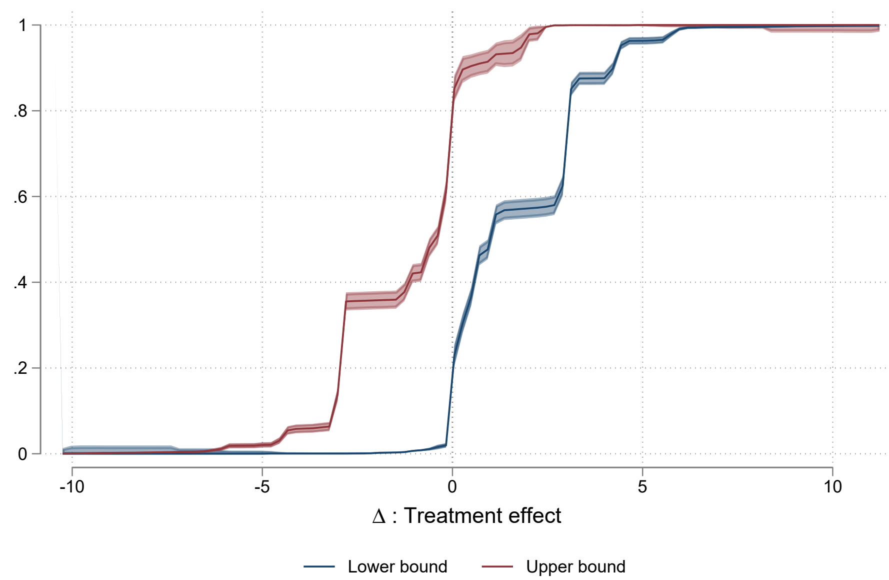
✓**Amount owed**: Loan + Accumulated interest before loan term ends**.** Interest accumulates daily on outstanding amount.

✓**Commitment**: to give you structure, each month you must pay at least 1/3 of the loan; that is: 3 equal sized payments. By missing it you incur in a **penalty fee of 2%** of the monthly payment due.

This is a (translated) sample information slide shown to clients. The real ones were twice the size of this figure and were laminated. Different ones were shown for each treatment arm.

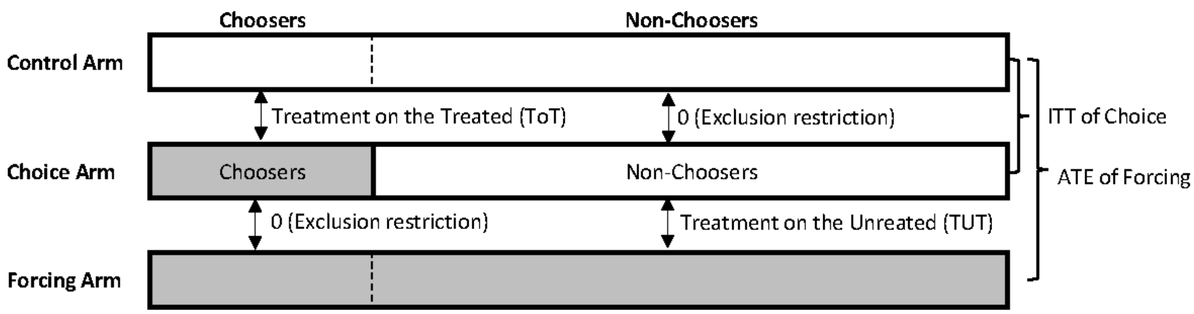
Figure 5: Fan & Park bounds for benefit in APR%

(a) APR



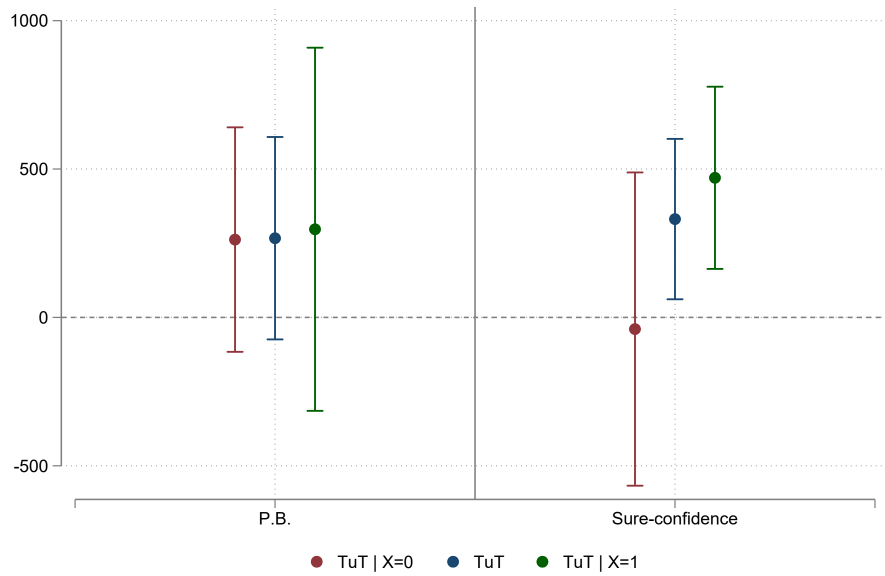

This figure depicts the Fan & Park (2010) bounds on the distribution $F_\Delta$ of individual treatment effects $\Delta \equiv (Y_1 - Y_0)$, described in Section 5.1, for the APR outcome. The dark red curve and light red shaded region give the estimated upper bound function $\overline{F}$ for $F_\Delta$ and associated (pointwise) 95% confidence interval. The dark blue curve and light blue shaded region give the estimated lower bound function $\underline{F}$ for $F_\Delta$ and associated (pointwise) 95% confidence interval. Confidence intervals are computed using the asymptotic distribution for the bounds. See Fan & Park (2010) for details. The bounds are pointwise sharp: at any specified value of $\delta$ the bounds $\underline{F}(\delta) \leq F_\Delta(\delta) \leq \overline{F}(\delta)$ cannot be improved without imposing additional assumptions. Evaluating the bounds at $\delta = 0$, we see that between 24% and 97% of borrowers have a positive individual treatment effect. This is greater than the share of borrowers who chose commitment: 11%.

Figure 6: The Controlled Choice Design




This figure provides graphical intuition for the way in which the controlled choice design from Section point identifies both the treatment on the treated (TOT) and treatment on the untreated (TUT) effects, as discussed in Section 5.3 and Appendix E.. The gray shaded regions denote borrowers with a commitment contract; the white shaded regions denote borrowers with a status quo contract. A comparison of means across control and forcing arms identifies the ATE of forcing commitment. The TOT and TUT effects of commitment are identified as follows. In the choice arm, anyone with a commitment contract is a chooser; in the control and forcing arms, we do not know who is a chooser and who is not. Because borrowers are randomly assigned to experimental treatment arms, however, the share of choosers will be the same, on average, across experimental arms. This is depicted using dashed vertical lines in the control and forcing arms. This overall share of choosers and non-choosers is point identified from the choice arm. Now, the difference of mean outcomes across the choice and control arms gives the intent-to-treat effect of offering choice. Under the exclusion restriction that moving non-choosers between the choice and forcing arms does not change their outcomes, this comparison "nets out" the non-choosers. Hence, the ITT of offering choice equals the TOT multiplied by the share of choosers. Similarly, under the exclusion restriction that moving choosers between the choice and forcing arm does not affect their outcomes, the difference of means across the forcing and choice arms "nets out" the choosers and hence equals the TUT multiplied by the share of non-choosers.

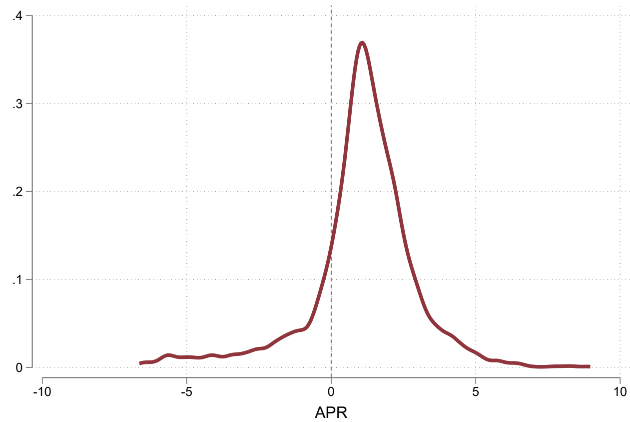
Figure 7: Partition of TUT by behavioral variables.



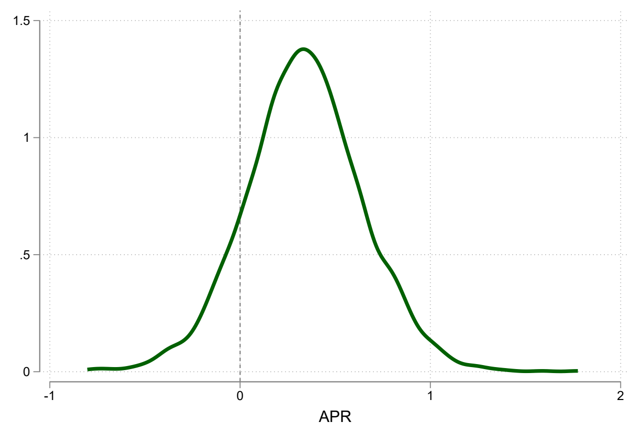

Each panel in this figure shows how the estimated treatment on the untreated (TUT) effect varies with a binary survey variable $X_i$. In the left panel (P.B.), $X_i = 1$ if borrower $i$ is "present-biased" based on her responses to the time preference questions from our survey. In the right panel (Sure-confidence) $X_i = 1$ if borrower $i$ reported that she was certain to recover her pawn, zero otherwise. Each panel depicts three estimates and associated 95% confidence bands: the overall TUT (blue), the TUT for borrowers with $X = 1$ (green), and the TUT for borrowers with $X = 0$ (red). By the law of iterated expectations, the overall TUT in each panel must lie between the group-specific TUTs. The overall TUT is slightly different in each panel, because some borrowers only completed part of the survey. In each panel, we present results for the full set of borrowers who answered the survey question or questions needed to determine $X_i$. From the left panel, we see that TUT does not appear to vary with present bias. In contrast, the right panel suggests that the positive and significant overall TUT is driven by individuals who are "sure-confident," i.e. borrowers who report that they are certain to recover their pawn.

Figure 8: Heterogeneous Treatment Effects

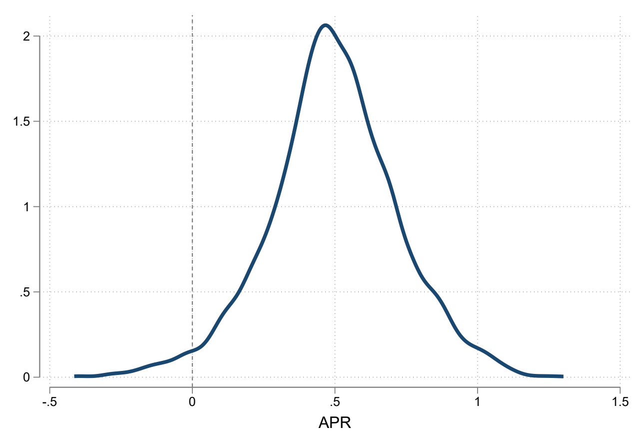
(a) ToT




(b) TuT

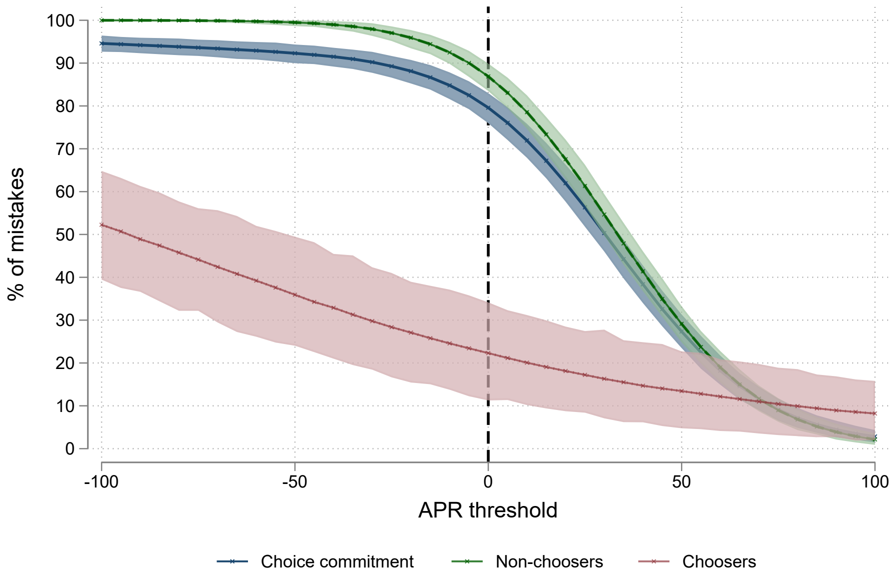



(c) ATE



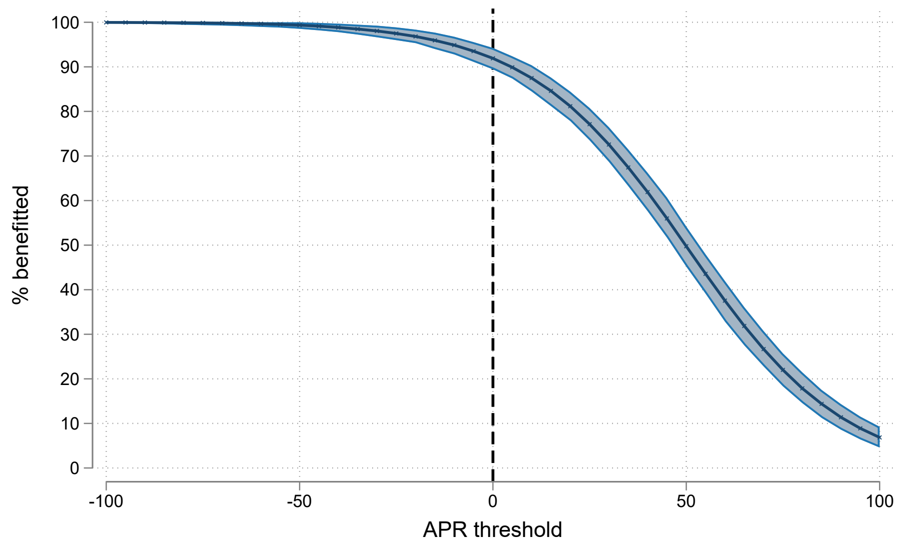

This Figure plots the densities of treatment on the untreated effects (TUT), treatment on the treated (TOT) effects, and conditional average treatment effects (ATE) based on the instrumental and causal forest models from Section 6.2.

Figure 9: "Mistakes" in the choice arm



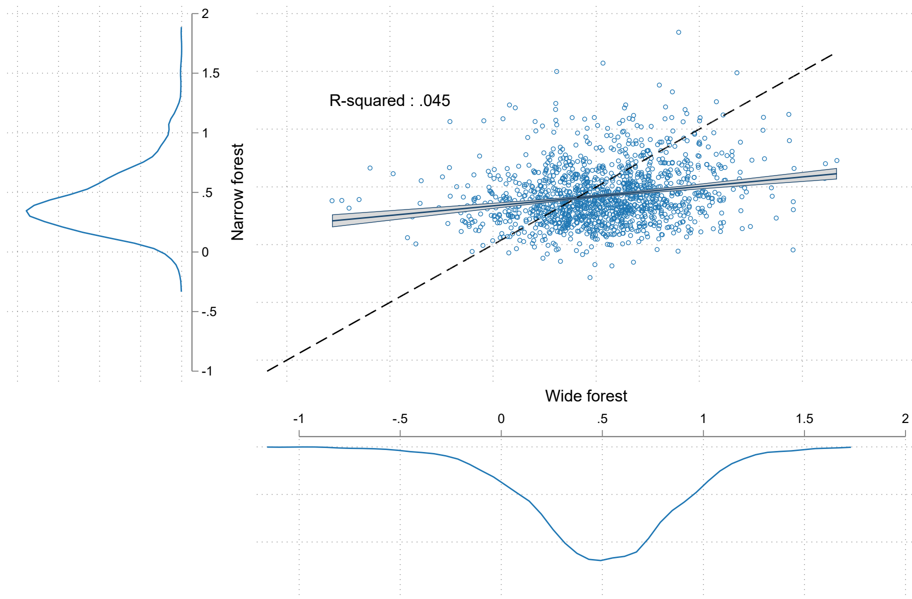

This figure presents conditional average TUT and TOT effects for the APR outcome from Figure 8 in an alternative manner, to consider the fraction of borrowers in the choice arm who made "mistakes" in their decision to accept or refuse the commitment contract. A "mistake" for a non-chooser is defined as a positive conditional TUT effect that significantly exceeds a specified threshold APR value. The green curve equals $[1 - F_{\text{TUT}}(\delta)] \times 100\%$, where $F_{\text{TUT}}$ is the CDF corresponding to density of conditional TUT effects from Figure 8, computed using the instrumental forest approach from Athey *et al.* (2019). Evaluated at any positive value on the horizontal axis, it gives the fraction of non-choosers in the choice arm who made a "mistake" by not choosing commitment. The green shaded region gives associated 95% pointwise confidence bands. Analogously, a "mistake" for a chooser is defined as a negative conditional TOT effect that exceeds a specified threshold APR value. The red curve equals $F_{\text{TOT}}(-\delta)$ where $F_{\text{TOT}}$ is the CDF corresponding to the density of conditional TOT effects from Figure 8. The red shaded region gives associated 95% pointwise confidence bands. For both the red and green curves, we define the APR threshold so that larger mistakes are to the *right* of smaller mistakes. This allows us to construct the overall fraction of mistakes in the choice arm, the blue curve, as a weighted average of the green and red curves. The weights equal the share of choosers and non-choosers in the choice arm.

Figure 10: Cumulative Distribution Function of Conditional ATE Estimates



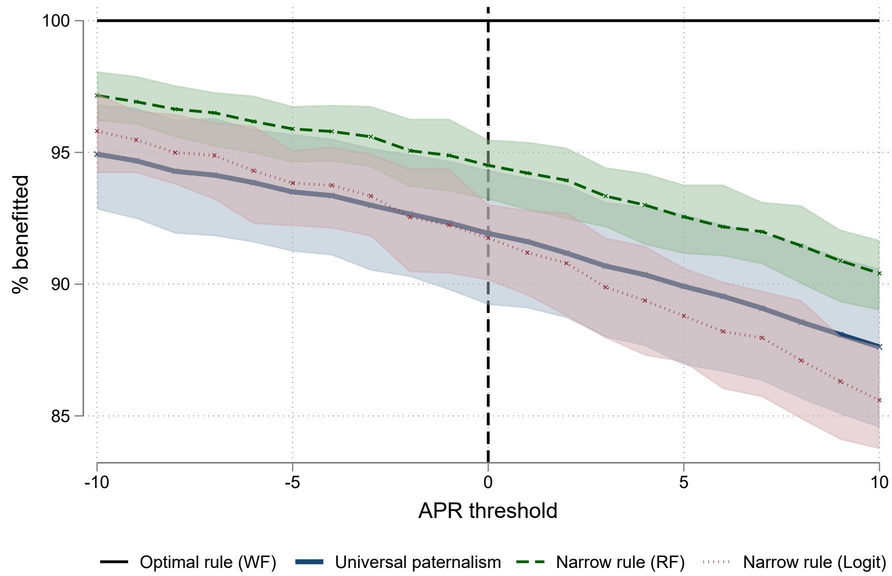

This figure shows the share of borrowers who benefit from forced commitment, where "benefit" is defined as having an estimated conditional average treatment effect (ATE) above a specified threshold value. Results are based on the causal forest model from Section 6.2. The solid line equals $1 - F_{\text{ATE}}(\delta)$ where $F_{\text{ATE}}(\delta)$ denotes the CDF corresponding to the density of conditional ATE estimates from Figure 8. The blue shaded regions are associated 95% confidence intervals. We estimate that at least 91% of borrowers have a positive conditional ATE.

Figure 11: Conditional ATEs from "wide" and "narrow" covariate sets




This figure plots the relationship between the causal forest conditional ATE estimates from Section 6.2 that use the "wide" set of covariates (all intake survey responses) and those based on a restricted "narrow" set of covariates (age, gender, HS education, and previous borrowing). The scatterplot graphs one estimate versus the other, with the "wide" covariate set on the horizontal axis and the "narrow" set on the vertical axis. The density plots on each axis show the estimated marginal distribution of conditional ATEs under each covariate set. The density for the "wide" covariate set is considerably more dispersed, as the causal forest based on this set of covariates captures considerably more treatment effect heterogeneity.

Figure 12: Targeting rules




This figure plots the percentage of borrowers who would benefit from a program in which commitment is targeted using the "narrow" covariate set (age, gender, HS education, and previous borrowing). In this exercise "benefit" is defined to mean having a conditional ATE above a particular threshold, taking the estimated conditional effects from the "wide" causal forest as ground truth. The plot compares the estimated percentage of borrowers who benefit, along with associated 95% confidence bands, for three targeting rules: universal forced commitment in blue, targeting based on a random forest classification model in green, and targeting based on the logit model in red. For each of the classification models, the outcome variable is an indicator for whether the CATE estimate from the "wide" random forest model is positive. Each of the classification models produces an estimated probability of a positive CATE. For each, the assignment rule ranks borrowers by this estimated probability, and assigns the highest 92% to forced commitment, to match the overall percentage of borrowers with positive estimated CATEs from the "wide" random forest model. Because this is an in-sample exercise, it overstates the actual performance of logit and RF-based targeting.

# The limits of self-commitment and private paternalism

## Appendix – For Online Publication

Craig McIntosh Isaac Meza Joyce Sadka Enrique Seira

## Contents

# Appendix A.

## A.1 Pictures

Figure OA-1: Some Pawnshops

(a) Appraiser/tellers inside a pawnshop


(b) Pawnshop

(c) Pawnshop


(d) Lost pawns which are for sale


This figure shows pictures of pawnshops in Mexico city. They do not necessarily coincide with Lender P for confidentiality.

Figure OA-2: Gold buyers next to pawnshops

(a) Gold buyer next to pawnshop 1


(b) Gold buyer next to pawnshop 1


This figure shows pictures of gold buyers next to pawnshops in Mexico city. They do not necessarily coincide with Lender P for confidentiality.

## A.2 Survey

We now present analysis of the extent to which the survey-responding sample is representative and balanced. Table OA-2 presents information about survey non-response. Panel A is a balance table that compares loan amount and day of the week across treatment arms for the subset of borrowers who responded to at least one survey question. The p-values in the fourth column are for the F-test of no difference of means across treatment arms. Panel B presents response-rates by question for each arm of the experiment, along with p-values for the F-test of no difference in question-specific response rates across treatment arms. This panel uses data from participants who answered at least one survey question and went on to pawn on the same day as their survey response. From the table, we see that loan amount and weekday are balanced across treatment arms among survey respondents and that response rates are likewise stable across treatment arms. Table OA-3 shows how the estimated TUT effect changes if we restrict our estimation sample based on survey response. The first row of the table presents TUT results for the financial cost outcome, while the second presents corresponding results for the APR outcome. In each row, column (1) presents the full-sample TUT estimate and standard error while the remaining columns restrict the sample to borrowers who answered a particular survey question or set of questions. For example, column (4) presents results for borrowers who answered the two questions needed to compute our measure of "present bias" discussed in the next paragraph, while column (9) present results restricted to participants who disclosed their sex. As seen from the table, our TUT estimates are quite stable across sub-samples defined by survey response. In each case they are positive and of the same magnitude as the corresponding full-sample estimate, although statistical significance varies depending on the size of the corresponding sub-sample. For these reasons, we are comfortable relying on data for survey respondents in the empirical exercises that follow. In exercises that rely on a single survey question, we use data for every borrower who answered that question. In the random forest exercises described below, we use data for every borrower who answered at least one survey question.

Table OA-1: Baseline survey questions (translated to English)

| | Baseline Survey |
|---|---|
| 1 | **Your pawn was:** |
| | (a) Inherite, (b) a gift, (c) bought by me, (d) lend to me, (e) other ____________ |
| 2 | **Mark with an "X" in the line below how likely is that you recover your pawn.** |
| | **Where 0 is impossible and 100 is completely certain** |
| 3 | **How much do you think the item you plan to pawn is worth? ______________ pesos** |
| 4 | **Gender** |
| 5 | **Age** |
| 6 | **Civil Status** |
| | (a) married, (b) single, (c) divorced, (d) widowed |
| 7 | **Work status** |
| | (a) employed, (b) own business, (c) houseshores, (d) don't work, (e) retired, (f) study |
| 8 | **Education** |
| | (a) no formal education, (b) primary, (c) middle school, (d) highschool, (e) more than highschool |
| 9 | **In the last month, did a friend or family member asked you for money?** |
| | (a) yes (b) no |
| 10 | **What would you like to have: 100 pesos tomorrow or 150 pesos in one month?** |
| 11 | **How often do you feel stressed by your economic situation?** |
| | (a) always, (b) very often, (c) sometimes, (d) never |
| 12 | **What is the main reason you want to pawn?** |
| | (a) Need the money because somebody in my family lost his/her job |
| | (b) Need the money to pay for a sickness in the family |
| | (c) Need the money for an urgent expense |
| | (d) Need the money for some non urgent expense. |
| 13 | **How stressed do you feel from the situation that led to to pawn?** |
| | (a) very stressed, (b) somwhat stressed, (c) a little stressed, (d) not stressed |
| 14 | **In 3 months, I expect to have a ________________ situation** |
| | (a) better, (b) similar, (c) worse |
| 15 | **Have you panwned before?** |
| | (a) yes (b) no |
| 16 | **How many times have you pawned on a Lender P branch?** |
| | (a) NO___ (b) 1-2 times ___ (c) 3-5 times____ (d) More than 5____ |
| 17 | **If you are saving money and a family member wants to use it for something** |
| | (a) I would only give him the money for an urgent expenze |
| | (b) I would give him the money even if it was not an urgent expense |
| | (c) I would not give him/her the money regardless |
| | (d) No one would ask me for my money |
| 18 | **Do you make an expenses budget for the month ahead of time?** |
| | (a) always, (b) very often, (c) sometimes, (d) never |
| 19 | **Do you have other items you could pawn?** |
| | (a) yes (b) no |
| 20 | **Do you have savings?** |
| | (a) yes (b) no |
| 21 | **Do you participate in a ROSCA?** |
| | (a) yes (b) no |
| 22 | **Is it common that family or friends ask for money?** |
| | (a) yes (b) no |
| 23 | **How much did you spend to come to the branch today? $______________ pesos** |
| 24 | **How much time does it usually take to come to this branch? ___________** |
| 25 | **How much does your family spend in a normal week? $______________ pesos** |
| 26 | **How much do you manage to save in a normal week? $______________ pesos** |
| 27 | **Does it happen to you that you spend more than you wanted because you fall into temptation?** |
| | (a) never, (b) almost never, (c) sometimes, (d) very often |
| 28 | **In the last 6 months, has it happened that at some point you lacked money to pay** |
| | (a) rent? (b) food (c)food (d) medicine (e) electricity (f) heating (g) telephone (i) water |
| 29 | **What would you like to have: 100 pesos in 3 months or 150 pesos in four months?** |
| 30 | **Would you like to receive (free) reminders for upcomming payments?** |
| | (a) yes (b) no |

Translation of the baseline questionnaire.

Table OA-2: Balance conditional on survey response and question-by-question response rates by treatment arm

| | | Commitment arms | | |
|---|---|---|---|---|
| | Control | Forced | Choice | p-value |
| | Panel A : Administrative Data (conditional on survey) | | | |
| Loan amount | 2199 | 2196 | 2216 | 0.98 |
| | (86) | (106) | (81) | |
| Weekday | 0.88 | 0.89 | 0.85 | 0.8 |
| | (0.045) | (0.038) | (0.045) | |
| Obs | 1386 | 1469 | 1982 | |
| | Panel B : Survey Data - response rate | | | |
| Subjective value | 0.73 | 0.69 | 0.71 | 0.39 |
| | (0.022) | (0.024) | (0.024) | |
| Trouble paying bills | 0.49 | 0.48 | 0.44 | 0.29 |
| | (0.023) | (0.022) | (0.022) | |
| Present bias | 0.43 | 0.44 | 0.39 | 0.17 |
| | (0.02) | (0.02) | (0.02) | |
| Makes budget | 0.55 | 0.56 | 0.53 | 0.56 |
| | (0.023) | (0.023) | (0.019) | |
| Subj. pr. of recovery | 0.74 | 0.72 | 0.74 | 0.71 |
| | (0.024) | (0.025) | (0.025) | |
| Pawn before | 0.55 | 0.57 | 0.53 | 0.57 |
| | (0.023) | (0.024) | (0.02) | |
| Age | 0.54 | 0.55 | 0.52 | 0.59 |
| | (0.023) | (0.023) | (0.018) | |
| Woman | 0.58 | 0.59 | 0.58 | 0.93 |
| | (0.025) | (0.024) | (0.02) | |
| + High-school | 0.54 | 0.55 | 0.51 | 0.39 |
| | (0.025) | (0.025) | (0.019) | |
| Obs | 1386 | 1469 | 1982 | |

Panel A of this table presents the same information as Panel A of Table 2, but restricts attention to the subsample of borrowers who answered at least one question in our baseline survey. Conditional on survey response, loan amount and weekday are balanced across treatment arms. Panel B presents response rates to each individual survey question across treatment arms, along with p-values for the F-test of no difference in response rates across survey arms. This panel uses data for borrowers who pawned on the same day as their survey response.

Table OA-3: Survey Non-response: TUT estimates for respondents to each survey question.

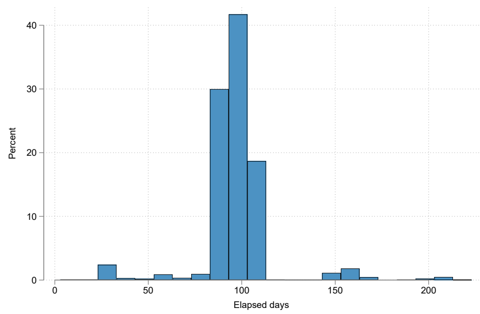
| | Full-sample | Subjective value | Trouble paying bills | Present bias | Makes budget | Subj. pr. of recovery | Pawn before | Age | Woman | + High-school |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| FC benefit TuT | 24.9*** | 20.9** | 15.2 | 21.3* | 18.4* | 26.8*** | 18.4* | 17.9* | 14.9 | 16.3 |
| | (8.31) | (9.10) | (11.6) | (11.4) | (10.5) | (8.59) | (10.3) | (10.6) | (9.88) | (10.1) |
| | (11) | (12) | (13) | (14) | (15) | (16) | (17) | (18) | (19) | (20) |
| Apr % benefit | 358.5*** | 331.8** | 163.3 | 266.9 | 242.5* | 331.3** | 225.9 | 229.9 | 120.9 | 111.6 |
| | (107.7) | (130.3) | (150.0) | (173.2) | (143.8) | (137.2) | (143.7) | (148.5) | (139.0) | (155.4) |
| Observations | 6304 | 4465 | 2948 | 2613 | 3433 | 4625 | 3468 | 3393 | 3677 | 3352 |

Some borrowers did not respond to our survey; others only completed part of the survey. This table computes the TUT effect for a number of sub-groups by survey response, and compares them against the overall TUT effect. The top panel uses financial cost as the outcome variable, while the bottom panel uses APR. Each group is defined by the borrowers who responded to a particular *subset* of the survey questions. For example, column (4) computes TUT effects for the subgroup of borrowers who responded to the two questions needed to compute our measure of present bias, while column (8) computes the TUT for the subgroup of borrowers who provided their age. The bottom row of the table counts the number of borrowers in each category. TUT estimates are quite stable across sub-groups and generally similar to the full-sample estimates. Because standard errors increase as sample size falls, some of the sub-group estimates are not statistically significant.

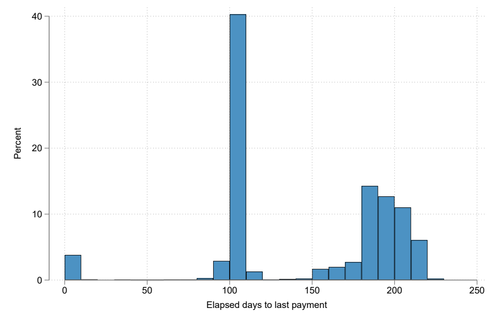
## A.3 Some more evidence of overconfidence

Figure OA-3: Behavior of borrowers who lost their pawn

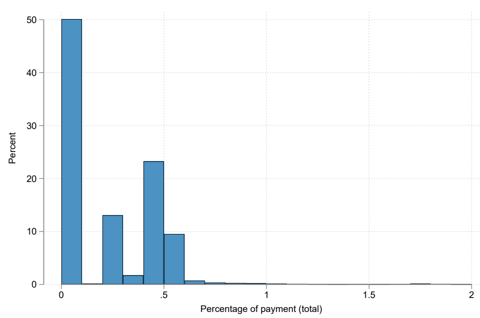
(a) Elapsed days to first payment



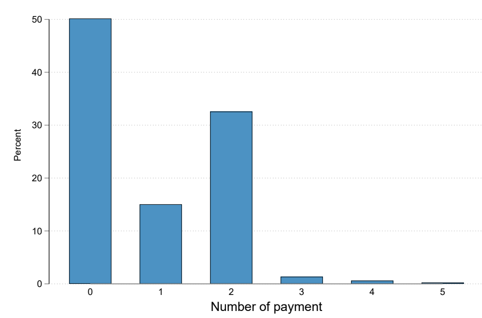

(b) Elapsed days to last payment




(c) Payments as % of loan




(d) Number of payments




This figure provides more details on the behavior of clients who were assigned to the control group and did not recover their pawn. Panel (a) shows a histogram of days elapsed from the pawn to the first payment, while panel (b) displays a histogram of days elapsed until the last payment. Some borrowers make payments after day 105, the end of the grace period: if they pay all interest owed, they can "restart" the loan. This amounts to starting a new loan with the same conditions and same pawn. Panel (c) shows a histogram of the fraction of the loan paid, while panel (d) presents a barplot of the number of times that borrowers went to the branch to make payments.

# Appendix B. Main treatment effects: Additional material

## B.1 Multiple loans

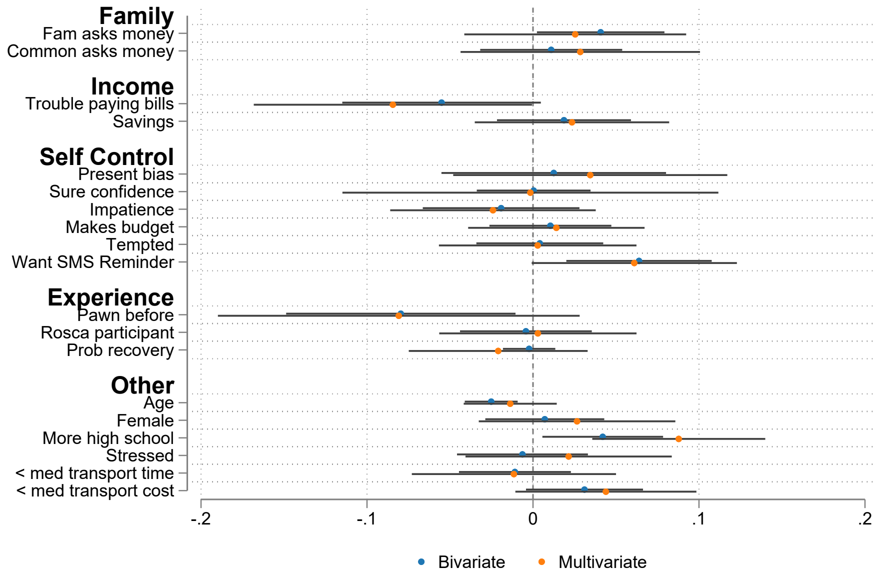
Table OA-4: Multiple-loans robustness check

| | First visit (Baseline approach) | | | | Multiple visits - multiple treatments | | | | First treatment (ITT) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | FC | APR | Recovery | Default | FC | APR | Recovery | Default | FC | APR | Recovery | Default |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
| Forced Commitment | -379.9*** | -0.34*** | 0.14*** | -0.064*** | -328.1*** | -0.20*** | 0.099*** | -0.032 | -285.3*** | -0.27*** | 0.11*** | -0.051*** |
| | (111.4) | (0.081) | (0.025) | (0.023) | (94.5) | (0.071) | (0.021) | (0.020) | (91.6) | (0.066) | (0.021) | (0.019) |
| Choice Commitment | -78.8 | -0.10 | 0.0094 | -0.023 | -71.0 | -0.024 | 0.00054 | -0.0046 | -105.0 | -0.13** | 0.029 | -0.032* |
| | (114.5) | (0.074) | (0.022) | (0.021) | (94.2) | (0.065) | (0.018) | (0.018) | (91.9) | (0.059) | (0.018) | (0.018) |
| Observations | 6304 | 6304 | 6304 | 6304 | 8519 | 8519 | 8519 | 8519 | 8813 | 8813 | 8813 | 8813 |
| R-sq | 0.007 | 0.011 | 0.019 | 0.013 | 0.010 | 0.008 | 0.018 | 0.010 | 0.011 | 0.012 | 0.020 | 0.015 |
| Control Mean | 1850.6 | 1.83 | 0.43 | 0.43 | 1794.4 | 1.74 | 0.46 | 0.42 | 1805.7 | 1.82 | 0.44 | 0.44 |

An additional empirical issue generated by repeat pawning is the question of how to handle the treatment status of those who take multiple loans within the experiment. 43% of borrowers take more than one loan within the study period, and 19% are assigned multiple treatment statuses on their different loans. The issue of sequential endogenous treatments has been extensively studied in the context of school lotteries, where the convention is to use the treatment status from the first exposure (ITT) (Cullen *et al.*, 2006; Abdulkadiroğlu *et al.*, 2011).

This table provides robustness for our main results for different ways to handling the multiplicity of loans. Columns 1-4 repeat the main analysis of financial cost and APR using the baseline approach from our main results,i.e. it only considers the first visit, while allowing multiple loans for the same-day. Columns 5-8 considers all loans including dummies for the order of pawns for an individual, essentially making the identification within-order. Columns 9-12 pursue the ITT strategy of always assigning the first treatment status to all subsequent pawns. The core treatment effects are robust to any of these ways of handling repeat pawning.

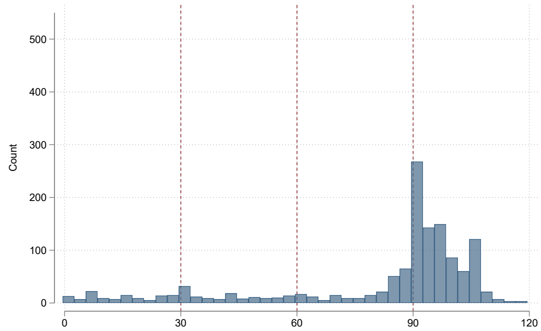
Figure OA-4: Determinants of choice



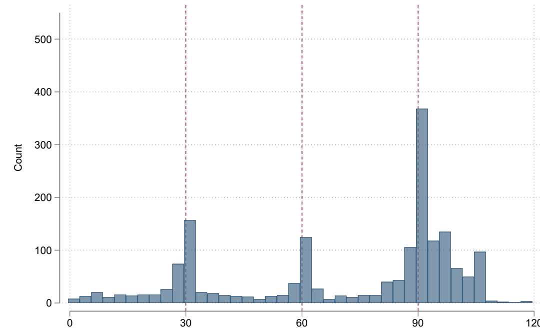

The above figure shows the determinants in a bivariate and multivariate OLS regression of choosing commitment. Choice commitment is a binary variable equal to one, whenever subjects choose the forced commitment contract in the choice arm.

Figure OA-5: Histogram of payments

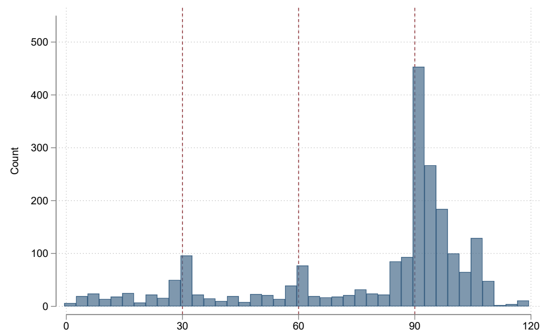
(a) Status-quo




(b) Forced commitment




(c) Choice commitment




This figure shows the schedule of payments for each treatment arm. In other words it records the time after the loan origination when the borrower makes a payment toward recovery of the pawn. We can see that for both the Forced commitment and Choice commitment arm, payments are bunching at the 30 and 60 days, while most payments are done around the due date of the loan (90 days).

## B.2 Censoring

Some loans in our sample are "censored" in that they continue beyond our observation period. For these loans, we do not know whether the borrower ultimately defaulted or recovered her pawn. In this appendix, we show that our main results are robust to different ways of addressing censoring.

Our main results, presented above, make no assumption on the default/recovery status of this loans. However, we compute the Financial Cost & APR considering all the payments done so far. We now consider some alternative approaches.

One way of considering the effect that this issue could have on our results is to make extreme assumptions about the outcome of these loans in the treatment and control so as to bound the possible influence of censoring. In Table OA-5 we compare the Forced and Control arms, making the bracketing assumptions about repayment on censored loans. Panel A assumes all censored loans are repaid, and Panel D that all default. Because forced commitment causes borrowers to make payments earlier than they otherwise would have, In Panel A we make a conservative assumption. To see why, suppose for the sake of argument that commitment affects payment timing, but does not not default. In other words, suppose that treatment makes censoring *less likely* but has no effect on uncensored outcomes. If we were to treat all outstanding loans as defaults–the opposite of our convention–this would artificially inflate the rate of default among control loans relative to treatment loans.

Panel B provides the lower bound for the treatment effect by assuming censored control loans are always repaid and treatment loans never are, and Panel C the upper bound by making the reverse assumption. Comfortingly, even with these extreme assumptions the significance on the main treatment effects never flips and treatment effects on financial cost and interests payments remain negative and significant in all scenarios. So there appears to be no scope for the censoring issue to overturn our main results.

Finally, Panel E of this table conducts a lasso logit prediction model that uses all of the available information on loans that were completed to predict the outcome of loans that were not. This is a best guess of the outcome on censored loans. Using this prediction, we replicate the main experimental results and find that the treatment effect on financial cost increases from -378 (main results) to -530 (censored loans predicted), and the APR from -34% to -62%. Hence, while the censoring issue does have a substantial influence on the magnitude of our estimated treatment effects, these checks confirm that a) the core results are robust to censoring, and b) the headline approach that we take to the issue is conservative and is likely understating the true magnitude of impacts.

Table OA-5: Bounding censoring

| | FC | Interest pymnt | Principal pymnt | Lost pawn value | Default | APR |
|---|---|---|---|---|---|---|
| | | Panel A : | Control = 0 | Forced Commitment = 0 | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Forced commitment | -408.3*** | -191.8*** | -0.42 | -248.1** | -0.063*** | -0.37*** |
| | (107.2) | (37.6) | (3.01) | (101.4) | (0.023) | (0.078) |
| Observations | 3724 | 3724 | 3724 | 3724 | 3724 | 3724 |
| R-sq | 0.012 | 0.025 | 0.004 | 0.012 | 0.019 | 0.022 |
| Control Mean | 1898.1 | 593.5 | 5.75 | 1304.7 | 0.43 | 1.88 |
| | | Panel B : | Control = 0 | Forced Commitment = 1 | | |
| | (7) | (8) | (9) | (10) | (11) | (12) |
| Forced commitment | -226.1** | -207.7*** | 1.38 | -50.0 | 0.0094 | 0.095 |
| | (110.8) | (37.4) | (3.45) | (103.3) | (0.024) | (0.096) |
| Observations | 3724 | 3724 | 3724 | 3724 | 3724 | 3724 |
| R-sq | 0.009 | 0.026 | 0.004 | 0.009 | 0.014 | 0.014 |
| Control Mean | 1898.1 | 593.5 | 5.75 | 1304.7 | 0.43 | 1.88 |
| | | Panel C : | Control = 1 | Forced Commitment = 0 | | |
| | (13) | (14) | (15) | (16) | (17) | (18) |
| Forced commitment | -804.2*** | -140.4*** | -2.33 | -695.5*** | -0.21*** | -1.17*** |
| | (113.3) | (34.1) | (3.16) | (100.8) | (0.023) | (0.10) |
| Observations | 3724 | 3724 | 3724 | 3724 | 3724 | 3724 |
| R-sq | 0.022 | 0.020 | 0.004 | 0.022 | 0.053 | 0.082 |
| Control Mean | 2272.4 | 545.9 | 7.69 | 1726.5 | 0.57 | 2.62 |
| | | Panel D : | Control = 1 | Forced Commitment = 1 | | |
| | (19) | (20) | (21) | (22) | (23) | (24) |
| Forced commitment | -622.0*** | -156.3*** | -0.53 | -497.3*** | -0.13*** | -0.71*** |
| | (117.3) | (33.8) | (3.58) | (103.1) | (0.024) | (0.12) |
| Observations | 3724 | 3724 | 3724 | 3724 | 3724 | 3724 |
| R-sq | 0.015 | 0.021 | 0.003 | 0.013 | 0.028 | 0.030 |
| Control Mean | 2272.4 | 545.9 | 7.69 | 1726.5 | 0.57 | 2.62 |
| | | Panel E : | Prediction with lasso-logit model | | | |
| | (25) | (26) | (27) | (28) | (29) | (30) |
| Forced commitment | -529.7*** | -172.4*** | -1.26 | -389.4*** | -0.12*** | -0.62*** |
| | (120.5) | (37.4) | (3.22) | (110.2) | (0.025) | (0.11) |
| Choice commitment | -61.5 | -30.6 | -4.52 | -32.3 | -0.016 | -0.039 |
| | (124.6) | (42.0) | (2.78) | (114.5) | (0.023) | (0.11) |
| Observations | 6304 | 6304 | 6304 | 6304 | 6304 | 6304 |
| R-sq | 0.012 | 0.022 | 0.003 | 0.009 | 0.016 | 0.024 |
| Control Mean | 2077.4 | 567.5 | 7.51 | 1509.9 | 0.51 | 2.28 |

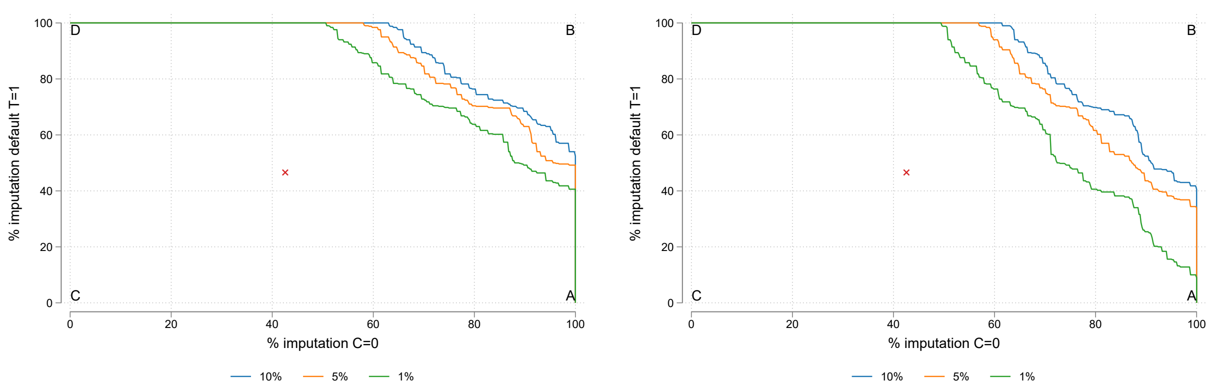
Given the censored loans, i.e. loans that have not finished by the end of the observation period, we estimate 'a la Manski' bounds for these loans, meaning that we impute all loans to either *default*= 1 or *recovery*= 0 depending on the treatment arm. Different panels perform different imputations for the censored loans for all possible combinations for the imputation, and computes the ATE for the same outcomes of Table 3. Panel A, for instance, assumes that all outstanding loans are fully payed. Panel B is the most conservative imputation since it assumes all outstanding loans in the control arm are payed, while all the outstanding loans in the forced commitment arm default. Panel C, on the other hand, is the most optimistic scenario opposite to that of Panel B. Panel D assumes all remaining loans default. The last panel makes the imputation to the censored loans according to the best prediction using a piecewise lasso logit model for default. In concrete, we build two logit models with lasso regularization, depending whether the loan duration is less than 220 days (two cycles) or more than 220 days. For prediction we use the former whenever the last recorded payment was done within 220 days, and the latter otherwise. Both models includes loan characteristics (loan size, branch), and payment behavior (loan duration so far, days to first payment, % of first payment, % of payments at 30, 60, 90, and 105 days, and % of interest payed at 105 days), but the latter model also includes % of payments at 150, 180, and 210 days. This predictive model achieves an accuracy rate of 92% both in-sample and out-of-sample.

Note that in all panels we maintain significant results for Financial Cost as dependent variable, while only in the most conservative scenario (Panel B) we lose significance for the APR outcome.

Figure OA-6: Interpolation on bounding censoring

(a) Significance area for APR

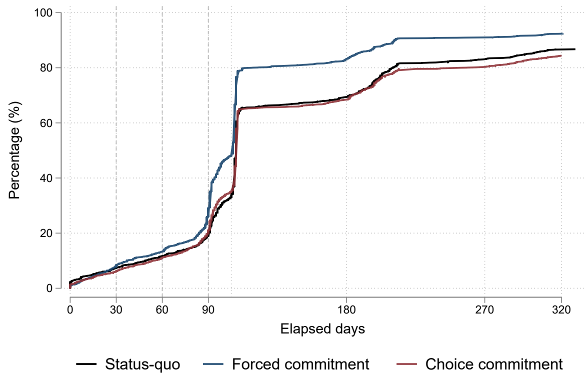
(b) Significance area for Default



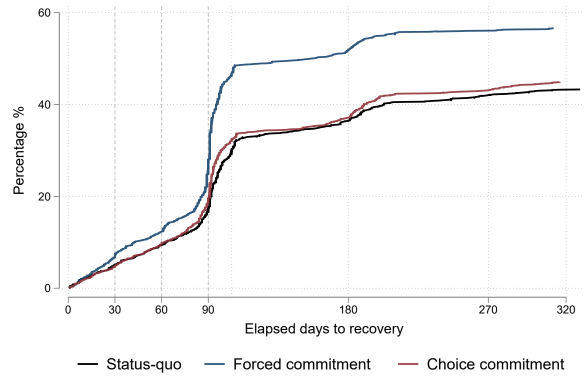

The next figure aims to answer the following question: For how many loans in the control arm can we impute recovery, and for how many in the treatment arm can we impute default and still have significance? This figure shows exactly the boundary separating significance when we vary the percentage of imputed censored loans with recovery and default respectively for control and treatment. Each corner in the square will correspond to one of the panels from the Table OA-5. For instance, the origin is the best-case scenario (Panel C) and the point (100,100) (Panel B) is the worst-case scenario. Thus we can think of this graph as an 'interpolation' from the four extreme cases. The 'x' indicates the proportions imputed by the lasso logit model, and the different lines correspond to setting different significance levels. We do not include the plot for financial cost, since for any imputation we still have significant results.

Figure OA-7: Survival graph

(a) Ended contract



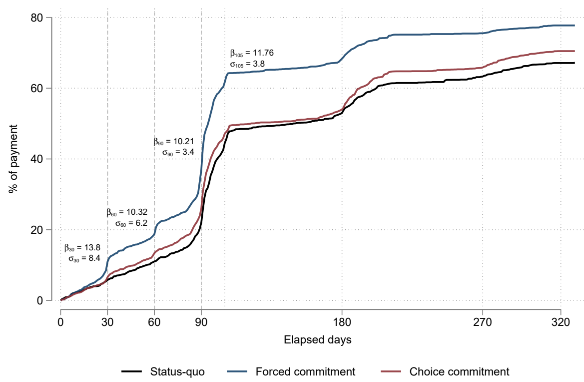

(b) Recovery



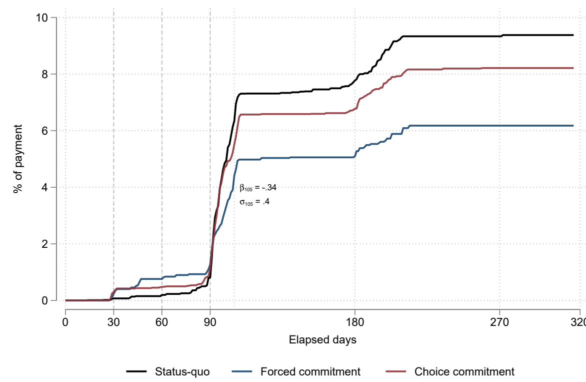

This Figure shows the accumulated percentage of recovery in time by treatment arm.

Figure OA-8: % of payment over time

(a) Unconditional




(b) Conditional on default




This Figure shows the accumulated percentage of recovery in time by treatment arm.

# Appendix C. Alternative explanations

## C.1 Learning

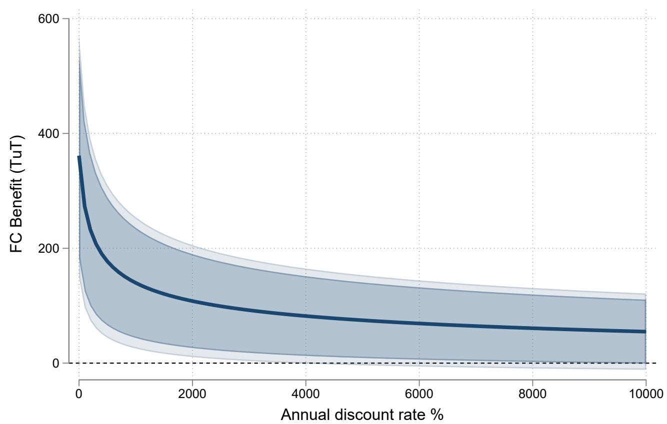
Table OA-6: Effect of Prior Assignment on Subsequent Choice

| | Choose commitment in $t+1$ | Ever choose commitment in $t+1$ |
|---|---|---|
| $t$ | (1) | (2) |
| Forced commitment (ATE) | -0.0047 | 0.00014 |
| | (0.048) | (0.0027) |
| Choice commitment (ITT) | 0.034 | 0.0015 |
| | (0.057) | (0.0030) |
| Observations | 228 | 4436 |
| R-sq | 0.004 | 0.000 |
| DepVarMean | 0.092 | 0.0047 |

This table presents information about borrowers' *future* pawning behavior. Column (1) reports results for the 228 borrowers who returned to pawn again at a day/branch that was randomly assigned to the choice arm, enabling us to observe whether they chose commitment or the status quo contract. Each row presents a difference in mean commitment take-up rates and associated standard errors. The first row (ATE) compares borrowers who were initially assigned to forced commitment against those were assigned to the control condition. The second row (ITT) compares borrowers who were initially assigned to the choice commitment condition to those who were not. Whereas column (1) conditions on the (endogenously) selected sample of borrowers who return to pawn again, column (2) considers the full sample by re-defining the "outcome" to be an indicator for whether a borrower pawned again on a day when choice was offered *and* chose commitment.

## C.2 Discount rates

Figure OA-9: Financial benefit TUT effect for different discount rates

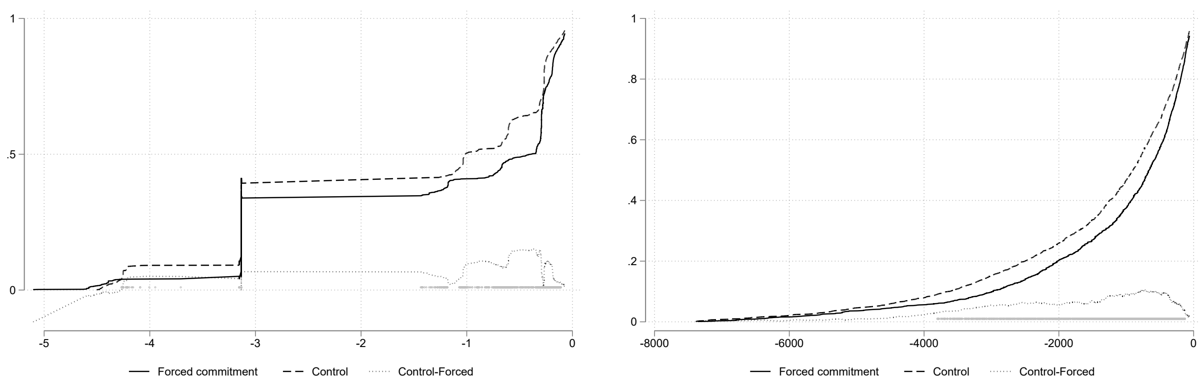



This Figure re-estimates the treatment on the untreated (TUT) effect from Table 7, introducing a daily discount factor in the definition of financial benefit. The discount factor reflects time preference, i.e. the fact that payments made in the present are more costly for the borrower. At a given annual discount rate in percentage points (x-axis) the solid line gives the adjusted TUT and the shaded regions 90% & 95% confidence bands. A discount factor of one corresponds to the estimate from Table 7. For higher discount rates, the adjusted financial cost will be higher, so the financial cost *savings* will be lower. As seen from the figure, borrowers would need to face unrealistically large discount rates to reverse our headline result of a large, positive, and statistically significant TUT effect.

## C.3 First order stochastic dominance

Figure OA-10: Empirical CDF of Financial Cost: Forced commitment vs Control

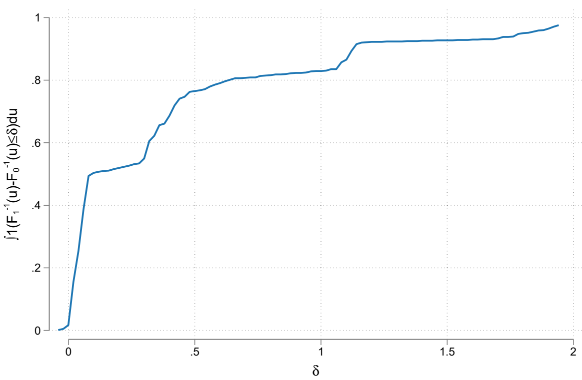
(a) APR benefit (b) Financial benefit



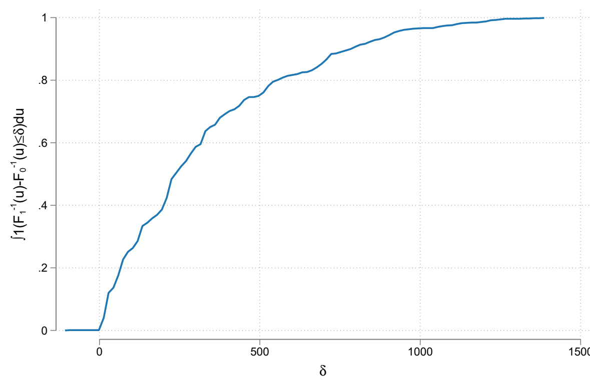

This figure plots plots the empirical CDF of experimental outcomes separately for the control / status quo (dashed) and forced commitment contracts (solid). In panel (a), the outcome is APR benefit; in panel (b) the outcome is Financial benefit. In each case, the empirical CDF under forced commitment first-order stochastically dominates the empirical CDF under the status quo. This can be seen by examining the dotted line at the bottom of panels (a) and (b), which shows the difference (Control − Commitment).

Figure OA-11: Distribution of treatment effects under rank invariance.

(a) APR benefit




(b) Financial benefit



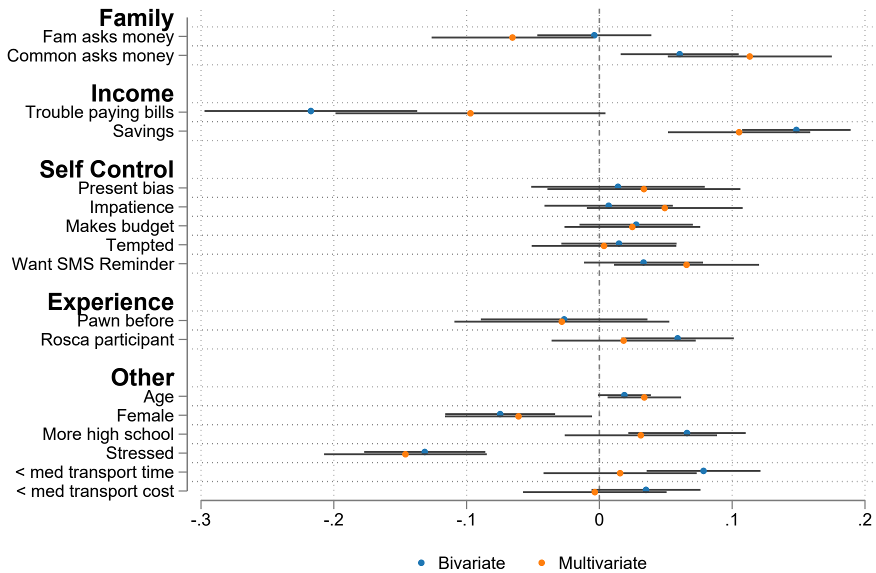

This figure show the distribution of treatment effects under rank invariance

$$F_\Delta(\delta) = \int_0^1 \mathbb{1}\{F_1^{-1}(u) - F_0^{-1}(u) \leq \delta\}\, \mathrm{d}u$$

where $F_1^{-1}$ and $F_0^{-1}$ are the quantile functions of $Y_1$ and $Y_0$.

## C.4 Sure Confidence

Figure OA-12: Determinants sure confidence




The above figure shows the determinants in a bivariate and multivariate OLS regression of sure confidence among the non-choosers. Sure confidence is a binary variable defined to be one when people report a 100% probability of recovery.

## Appendix D. Testing for heterogeneity

If the treatment effects $Y_{i1} - Y_{i0}$ are constant across $i$, then we must have

$$\text{ATE}(X_i) \equiv \mathbb{E}[Y_{i1} - Y_{i0}|X_i] = \mathbb{E}[Y_{i1} - Y_{i0}] \equiv \text{ATE}$$

for any covariates $X_i$ that vary across $i$. If, on the other hand, $\text{ATE}(X_i)$ can be predicted using some scalar function $\tau(\cdot)$ of $X_i$, then the average treatment effect function is not constant so there must be treatment effect heterogeneity.

We operationalize this idea using a two-step approach proposed by Chernozhukov *et al.* (2018). We begin by randomly dividing the participants in the forced arms of the experiment ($Z_i \neq 2$) into two groups: a training set and a test set. These sets are constructed to ensure that all observations from a given branch-day cluster are allocated to the same set. This avoids inferential problems that could arise from correlated unobservables within clusters. In the first step, we apply the generalized random forest approach of Athey *et al.* (2019) to the training set to estimate two proxy predictors: $\psi(\cdot|\text{Training})$ approximates the untreated potential outcome function, $\mathbb{E}[Y_{i0}|X_i] = \mathbb{E}[Y_i|Z_i = 0, X_i]$, while $\tau(\cdot|\text{Training})$, approximates the ATE function

$$\text{ATE}(X_i) = \mathbb{E}[Y_i|Z_i = 1, X_i] - \mathbb{E}[Y_i|Z_i = 0, X_i].$$

The proxy predictors need not be unbiased or even consistent estimators of the functions they aim to approximate: the goal of this exercise is merely to find a scalar function of $X_i$ that *accurately predicts* $\text{ATE}(X_i)$. In the second step we fit a linear regression model to data from the training set using regressors constructed from the proxy functions $\psi(\cdot|\text{Training})$ and $\tau(\cdot|\text{Training})$ constructed in the first step. In particular, we estimate

$$Y_i = \alpha_0 + \alpha_1\psi_i + \beta_1(Z_i - \mathbb{E}[Z_i]) + \beta_2(Z_i - \mathbb{E}[Z_i])(\tau_i - \mathbb{E}[\tau_i]) + \epsilon_i \tag{12}$$

where $\psi_i \equiv \psi(X_i|\text{Training})$ and $\tau_i \equiv \tau(X_i|\text{Training})$.[32] As shown by Chernozhukov *et al.* (2018), the coefficients $\beta_1$ and $\beta_2$ from (12) identify the *best linear predictor* of the conditional ATE based on $\tau(\cdot|\text{Training})$, namely

$$\beta_1 = \mathbb{E}[\text{ATE}(X_i)] = \text{ATE}, \quad \beta_2 = \frac{\text{Cov}[\text{ATE}(X_i), \tau_i]}{\text{Var}(\tau_i)}.$$

If treatment effects are homogeneous we must have $\beta_2 = 0$. Rejecting this hypothesis establishes that $\tau_i$ predicts $\text{ATE}(X_i)$ and hence that $\Delta_i$ varies. Since $\tau_i$ and $\psi_i$ do not depend on the test set, inference for the regression in (12) is straightforward conditional on the Training/Test split. Our

[32] This is a slightly simpler regression than the one proposed in equation (3.1) of Chernozhukov *et al.* (2018), which involves propensity score weights. Because the random assignment of $Z$ in our experiment does *not* condition on $X$, the propensity score weights in our case are constant over $X$ and hence drop out.

estimate for $\beta_2$ is 1.68 with a one-sided heteroskedasticity-robust (HC3) of 0.50. Thus we easily reject the null hypothesis of homogeneous treatment effects.

# Appendix E. Derivations for Section 5.3

This appendix provides proofs of the results described in Section 5.3, using the notation and assumptions described in Section 5.2. To simplify the presentation, we omit $i$ subscripts throughout this section. We also use the shorthand $Z_0 \equiv \mathbb{1}(Z=0)$, $Z_1 \equiv \mathbb{1}(Z=1)$, and $Z_2 \equiv \mathbb{1}(Z=2)$. For convenience, the following assumption collects our exclusion restriction and the key features of the constrained choice design.

**Assumption 1.**

*(i)* $Z$ *is independent of* $(Y_0, Y_1, C)$

*(ii)* $D = \mathbb{1}(Z \neq 2)Z + \mathbb{1}(Z=2)C$

*(iii)* $Y = \mathbb{1}(Z=0)Y_0 + \mathbb{1}(Z=1)Y_1 + \mathbb{1}(Z=2)[(1-C)Y_0 + CY_1]$

## E.1 Point Identification

We first show that the TOT, TUT, ASB, and ASL effects are point identified under the constrained choice design. It follows that the ASG effect, (TOT − TUT), is likewise point identified.

**Lemma 1.** *Under Assumption 1,*

*(i)* $\mathbb{E}(D|Z=2) = \mathbb{P}(C=1)$

*(ii)* $\mathbb{E}(Y|Z=0) = \mathbb{E}(Y_0)$

*(iii)* $\mathbb{E}(Y|Z=1) = \mathbb{E}(Y_1)$

*(iv)* $\mathbb{E}(Y|D=0, Z=2) = \mathbb{E}(Y_0|C=0)$

*(v)* $\mathbb{E}(Y|D=1, Z=2) = \mathbb{E}(Y_1|C=1)$.

*Proof.* Part (i) follows because $Z=2$ implies $D=C$ and $Z$ is independent of $C$. Parts (ii) and (iii) follow similarly: given $Z=0$ we have $Y=Y_0$, given $Z=1$ we have $Y=Y_1$, and $Z$ is independent of $(Y_0, Y_1)$. For parts (iv) and (v), first note that Assumption 1 (iii) implies that $Z$ is conditionally independent of $(Y_0, Y_1)$ given $C$. Now, $Z=2$ implies that $D=0$ if and only if $C=0$. Hence,

$$\mathbb{E}(Y|D=0, Z=2) = \mathbb{E}(Y_0|D=0, Z=2) = \mathbb{E}(Y_0|C=0, Z=2) = \mathbb{E}(Y_0|C=0)$$

establishing part (iv). For part (v) $Z=2$ implies that $D=1$ if and only if $C=1$ and hence

$$\mathbb{E}(Y|D=1, Z=2) = \mathbb{E}(Y_1|D=1, Z=2) = \mathbb{E}(Y_1|C=1, Z=2) = \mathbb{E}(Y_1|C=1).$$

□

**Proposition 1.** *Under Assumption 1,*

(i) $TOT \equiv \mathbb{E}(Y_1 - Y_0|C=1) = \dfrac{\mathbb{E}(Y|Z=2) - \mathbb{E}(Y|Z=0)}{\mathbb{E}(D|Z=2)}$

(ii) $TUT \equiv \mathbb{E}(Y_1 - Y_0|C=0) = \dfrac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=2)}{1-\mathbb{E}(D|Z=2)}$

(iii) $ASB \equiv \mathbb{E}(Y_0|C=1) - \mathbb{E}(Y_0|C=0) = \dfrac{\mathbb{E}(Y|Z=0) - \mathbb{E}(Y|Z=2, D=0)}{\mathbb{E}(D|Z=2)}$

(iv) $ASL \equiv \mathbb{E}(Y_1|C=1) - \mathbb{E}(Y_1|C=0) = \dfrac{\mathbb{E}(Y|Z=2, D=1) - \mathbb{E}(Y|Z=1)}{1-\mathbb{E}(D|Z=2)}$.

*Proof.* For parts (i) and (iii) we require an expression for $\mathbb{E}(Y_0|C=1)$ in terms of the observables $(Y, D, Z)$. By Lemma 1(ii) and iterated expectations

$$\mathbb{E}(Y|Z=0) = \mathbb{E}(Y_0) = \mathbb{E}(Y_0|C=0)\mathbb{P}(C=0) + \mathbb{E}(Y_0|C=1)\mathbb{P}(C=1).$$

Re-arranging and substituting Lemma 1(i) and (iv),

$$\begin{aligned}\mathbb{E}(Y_0|C=1) &= \frac{\mathbb{E}(Y|Z=0) - \mathbb{E}(Y_0|C=0)\mathbb{P}(C=0)}{\mathbb{P}(C=1)}\\ &= \frac{\mathbb{E}(Y|Z=0) - \mathbb{E}(Y|Z=2, D=0)\mathbb{E}(1-D|Z=2)}{\mathbb{E}(D|Z=2)}.\end{aligned} \tag{13}$$

Part (i) follows by combining (13) with Lemma 1(v) and simplifying; part (iii) follows by combining (13) with Lemma 1(iv) and simplifying. Similarly, for parts (ii) and (iv) we require an expression for $\mathbb{E}(Y_1|C=0)$ in terms of observables. By Lemma 1(iii) and iterated expectations,

$$\mathbb{E}(Y|Z=1) = \mathbb{E}(Y_1) = \mathbb{E}(Y_1|C=0)\mathbb{P}(C=0) + \mathbb{E}(Y_1|C=1)\mathbb{P}(C=1).$$

Re-arranging and substituting Lemma 1(i) and (v),

$$\begin{aligned}\mathbb{E}(Y_1|C=0) &= \frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y_1|C=1)\mathbb{P}(C=1)}{\mathbb{P}(D=0)}\\ &= \frac{\mathbb{E}(Y|Z=1) - \mathbb{E}(Y|Z=2, D=1)\mathbb{E}(D|Z=2)}{\mathbb{E}(1-D|Z=2)}.\end{aligned} \tag{14}$$

Part (ii) follows by combining (14) with Lemma 1(iv) and simplifying; part (iv) follows by combining (14) with Lemma 1(v) and simplifying. □

## E.2 Regression-based Estimation of TOT, TUT, ASG, ASL, and ASB

We now show how a collection of just-identified, linear IV regressions can be used to consistently estimate the ATE, TOT, and TUT effects, along with each of the ingredients needed to construct the ASL and ASB. These results are used below to provide a recipe for cluster-robust inference for the ASG, ASL, and ASB effects. The first result provides a regression-based approach to estimate the ATE and TOT.

**Proposition 2.** *Under Assumption 1,*

$$Y = \mathbb{E}(Y_0) + ATE \times Z_1 + TOT \times Z_2 D + U$$

*where $\mathbb{E}(U|Z) = 0$. Therefore, under standard regularity conditions, an IV regression of $Y$ on an intercept, $Z_1$ and $Z_2 D$ with instruments $(1, Z_0, Z_1)$ provides a consistent estimator of the ATE and TOT effects.*

*Proof.* By Assumption 1 (iii),

$$\begin{aligned}
Y &= Z_0 Y_0 + Z_1 Y_1 + Z_2[(1-C)Y_0 + CY_1] \\
&= (Z_0 + Z_2)Y_0 + Z_1 Y_1 + Z_2 C(Y_1 - Y_0) \\
&= (Z_0 + Z_1 + Z_2)Y_0 + Z_1(Y_1 - Y_0) + Z_2 D(Y_1 - Y_0) \\
&= Y_0 + Z_1(Y_1 - Y_0) + Z_2 D(Y_1 - Y_0).
\end{aligned}$$

since $Z_2 D = Z_2 C$ and $(Z_0 + Z_1 + Z_2) = 1$. Thus, defining

$$U \equiv [Y_0 - \mathbb{E}(Y_0)] + Z_1[(Y_1 - Y_0) - \text{ATE}] + Z_2 D[(Y_1 - Y_0) - \text{TOT}]$$

by construction we have

$$Y = \mathbb{E}(Y_0) + \text{ATE} \times Z_1 + \text{TOT} \times Z_2 D + U.$$

Now, since $Z_2 D = Z_2 C$ and $Z$ is independent of $(Y_1, Y_0)$ by Assumption 1 (i), we have

$$\begin{aligned}
\mathbb{E}(U|Z) &= [\mathbb{E}(Y_0|Z) - \mathbb{E}(Y_0)] + Z_1[\mathbb{E}(Y_1 - Y_0|Z) - \text{ATE}] + \mathbb{E}\left[Z_2 D \left\{(Y_1 - Y_0) - \text{TOT}\right\} |Z\right] \\
&= [\mathbb{E}(Y_0) - \mathbb{E}(Y_0)] + Z_1[\mathbb{E}(Y_1 - Y_0) - \text{ATE}] + Z_2 \mathbb{E}\left[C \left\{(Y_1 - Y_0) - \text{TOT}\right\} |Z\right] \\
&= Z_2 \mathbb{E}\left[C \left\{(Y_1 - Y_0) - \text{TOT}\right\} |Z\right].
\end{aligned}$$

Finally, by iterated expectations

$$\begin{aligned}
\mathbb{E}\left[C \left\{(Y_1 - Y_0) - \text{TOT}\right\} |Z\right] &= \mathbb{P}(C = 1|Z)\left[\mathbb{E}(Y_1 - Y_0|C = 1, Z) - \text{TOT}\right] \\
&= \mathbb{P}(C = 1)\left[\mathbb{E}(Y_1 - Y_0|C = 1) - \text{TOT}\right] = 0
\end{aligned}$$

since $Z$ is conditionally independent of $(Y_0, Y_1)$ given $C$, an implication of 1 (i). □

The next proposition provides regression-based estimates of the ATE and TUT.

**Proposition 3.** *Under Assumption 1*

$$Y = \mathbb{E}(Y_1) + ATE \times -Z_0 + TUT \times -Z_2(1-D) + V$$

*where $\mathbb{E}(V|Z) = 0$. Therefore, under standard regularity conditions, an IV regression of $Y$ on an intercept, $-Z_0$ and $-Z_2(1-D)$ with instruments $(1, Z_0, Z_1)$ provides consistent estimates of the ATE and TUT effects.*

*Proof.* By Assumption 1 (iii),

$$\begin{aligned}
Y &= Z_0Y_0 + Z_1Y_1 + Z_2[(1-C)Y_0 + CY_1] \\
&= Z_0Y_0 + Z_1Y_1 + Z_2[(1-C)(Y_0 - Y_1) + Y_1] \\
&= Z_0Y_0 + (Z_1 + Z_2)Y_1 + Z_2(1-C)(Y_0 - Y_1) \\
&= Z_0Y_0 + (1 - Z_0)Y_1 + Z_2(1-D)(Y_0 - Y_1) \\
&= Y_1 - Z_0(Y_1 - Y_0) - Z_2(1-D)(Y_1 - Y_0)
\end{aligned}$$

since $Z_2(1-C) = Z_2(1-D)$ and $(Z_1 + Z_2) = 1 - Z_0$. Thus, defining

$$V \equiv [Y_1 - \mathbb{E}(Y_1)] - Z_0[(Y_1 - Y_0) - \text{ATE}] - Z_2(1-D)[(Y_1 - Y_0) - \text{TUT}]$$

by construction we have

$$Y = \mathbb{E}(Y_1) + \text{ATE} \times -Z_0 + \text{TUT} \times -Z_2(1-D) + V.$$

Now, since $Z_2(1-D) = Z_2(1-C)$ and $Z$ is independent of $(Y_0, Y_1)$ by Assumption 1 (i),

$$\begin{aligned}
\mathbb{E}(V|Z) &= [\mathbb{E}(Y_1|Z) - \mathbb{E}(Y_1)] - Z_0[\mathbb{E}(Y_1 - Y_0|Z) - \text{ATE}] - \mathbb{E}[Z_2(1-D)\{(Y_1 - Y_0) - \text{TUT}\}|Z] \\
&= [\mathbb{E}(Y_1) - \mathbb{E}(Y_1)] - Z_0[\mathbb{E}(Y_1 - Y_0) - \text{ATE}] - Z_2\mathbb{E}[(1-C)\{(Y_1 - Y_0) - \text{TUT}\}|Z] \\
&= -Z_2\mathbb{E}[(1-C)\{(Y_1 - Y_0) - \text{TUT}\}|Z].
\end{aligned}$$

Finally, by iterated expectations,

$$\begin{aligned}
\mathbb{E}[(1-C)\{(Y_1 - Y_0) - \text{TUT}\}|Z] &= \mathbb{P}(C = 0|Z)[\mathbb{E}(Y_1 - Y_0|C = 0, Z) - \text{TUT}] \\
&= \mathbb{P}(C = 0|Z)[\mathbb{E}(Y_1 - Y_0|C = 0) - \text{TUT}] = 0
\end{aligned}$$

since $Z$ is conditionally independent of $(Y_0, Y_1)$ given $C$, an implication of Assumption 1 (i). □

Since ASG = TOT − TUT, the preceding two propositions provide consistent estimates of ASG effect. The ASB effect, $\mathbb{E}(Y_0|C = 1) - \mathbb{E}(Y_0|C = 0)$, can likewise be estimated by taking the difference of coefficients across two linear IV regressions, as shown in the following proposition.

**Proposition 4.** *Under Assumption 1*

$$(1-D)Y = \mathbb{E}(Y_0) \times Z_0 + \mathbb{E}(Y_0|C=0) \times (1-D)Z_2 + U_0 \tag{15}$$

$$(1-D)Y = \mathbb{E}(Y_0) \times (Z_0 + Z_2) + \mathbb{E}(Y_0|C=1) \times -DZ_2 + U_1 \tag{16}$$

*where* $\mathbb{E}(U_0|Z) = \mathbb{E}(U_1|Z) = 0$*. Thus, under standard regularity conditions, an IV regression of* $(1-D)Y$ *on* $Z_0$ *and* $(1-D)Z_2$ *with instruments* $(Z_0, Z_2)$ *and no intercept provides consistent estimates of* $\mathbb{E}(Y_0)$ *and* $\mathbb{E}(Y_0|C=0)$*. Similarly, an IV regression of* $(1-D)Y$ *on* $(Z_0+Z_2)$ *and* $-DZ_2$ *with instruments* $(Z_0, Z_2)$ *and no intercept provides consistent estimates of* $\mathbb{E}(Y_0)$ *and* $\mathbb{E}(Y_0|C=1)$*.*

*Proof.* Assumption 1 (ii) implies $(1-D) = Z_0 + Z_2(1-C)$. Hence, by Assumption 1 (iii),

$$\begin{aligned}(1-D)Y &= [Z_0 + Z_2(1-C)]\{Z_0Y_0 + Z_1Y_1 + Z_2[(1-C)Y_0 + CY_1]\}\\ &= Z_0Y_0 + Z_2(1-C)[(1-C)Y_0 + CY_1]\\ &= Z_0Y_0 + Z_2(1-C)Y_0.\end{aligned}$$

because $Z_j^2 = Z_j$ for any $j$ and $Z_jZ_k = 0$ for any $j \neq k$ and, similarly, $(1-C)^2 = (1-C)$ and $C(1-C) = 0$. Therefore, since $Z_2(1-C) = Z_2(1-D)$,

$$\begin{aligned}(1-D)Y &= Z_0Y_0 + Z_2(1-D)Y_0\\ (1-D)Y &= (Z_0+Z_2)Y_0 + (-DZ_2)Y_0.\end{aligned}$$

Now, defining

$$\begin{aligned}U_0 &\equiv Z_0[Y_0 - \mathbb{E}(Y_0)] + Z_2(1-D)[Y_0 - \mathbb{E}(Y_0|C=0)]\\ U_1 &\equiv (Z_0+Z_2)[Y_0 - \mathbb{E}(Y_0)] + (-Z_2D)[Y_0 - \mathbb{E}(Y_0|C=1)]\end{aligned}$$

by construction, we have

$$\begin{aligned}Y &= \mathbb{E}(Y_0) \times Z_0 + \mathbb{E}(Y_0|C=0) \times Z_2(1-D) + U_0\\ Y &= \mathbb{E}(Y_0) \times (Z_0+Z_2) + \mathbb{E}(Y_0|C=1)(-Z_2D) + U_1.\end{aligned}$$

Finally, since $Z_2(1-D) = Z_2(1-C)$, and $Z$ is independent of $Y_0$,

$$\begin{aligned}\mathbb{E}(U_0|Z) &= Z_0[\mathbb{E}(Y_0|Z) - \mathbb{E}(Y_0)] + Z_2\mathbb{E}\{(1-C)[Y_0 - \mathbb{E}(Y_0|C=0)|Z\}\\ &= Z_2\mathbb{E}[Y_0 - \mathbb{E}(Y_0|C=0)|C=0, Z] = 0\end{aligned}$$

where the final two steps follow by iterated expectations and the fact that $Z$ is conditionally independent of $Y_0$ given $C$. Since $Z_2D = Z_2C$, a nearly identical argument gives

$$\begin{aligned}\mathbb{E}(U_1|Z) &= (Z_0 + Z_2)[Y_0 - \mathbb{E}(Y_0|Z)] - Z_2\mathbb{E}\{C[Y_0 - \mathbb{E}(Y_0|C = 1)|Z\}\\ &= -Z_2\mathbb{E}[Y_0 - \mathbb{E}(Y_0|C = 0)|C = 1, Z] = 0. \qquad \square\end{aligned}$$

The next results shows that the ASL effect, $\mathbb{E}(Y_1|C = 1) - \mathbb{E}(Y_1|C = 0)$, can be estimated as the difference of coefficients across two linear IV regressions, as shown in the following proposition.

**Proposition 5.** *Under Assumption 1,*

$$DY = \mathbb{E}(Y_1) \times (Z_1 + Z_2) + \mathbb{E}(Y_1|C = 0) \times (D - 1)Z_2 + V_0 \tag{17}$$

$$DY = \mathbb{E}(Y_1) \times Z_1 + \mathbb{E}(Y_1|C = 1) \times DZ_2 + V_1 \tag{18}$$

*where $\mathbb{E}(V_0|Z) = \mathbb{E}(V_1|Z) = 0$. Thus, under standard regularity conditions, an IV regression of $DY$ on $(Z_1 + Z_2)$ and $-Z_2(1 - D)$ with instruments $(Z_1, Z_2)$ and no intercept provides consistent estimates of $\mathbb{E}(Y_1)$ and $\mathbb{E}(Y_1|C = 1)$. Similarly, an IV regression of $DY$ on $Z_1$ and $DZ_2$ with instruments $(Z_1, Z_2)$ and no intercept provides consistent estimates of $\mathbb{E}(Y_1)$ and $\mathbb{E}(Y_1|C = 1)$.*

*Proof.* By Assumption 1, $D = Z_1 + Z_2C$. Hence, by Assumption 1 (iii),

$$\begin{aligned}DY &= (Z_1 + Z_2C)\{Z_0Y_0 + Z_1Y_1 + Z_2[(1 - C)Y_0 + CY_1]\}\\ &= Z_1Y_1 + Z_2C[(1 - C)Y_0 + CY_1]\\ &= Z_1Y_1 + Z_2CY_1\end{aligned}$$

because $Z_j^2 = Z_j$ for any $j$ and $Z_jZ_k = 0$ for any $j \neq k$ and, similarly, $(1 - C)^2 = (1 - C)$ and $C(1 - C) = 0$. Therefore, since $Z_2(1 - C) = Z_2(1 - D)$,

$$\begin{aligned}DY &= (Z_1 + Z_2)Y_1 + Z_2(D - 1)Y_1\\ DY &= Z_1Y_1 + Z_2DY_1.\end{aligned}$$

Now, defining

$$\begin{aligned}V_0 &= (Z_1 + Z_2)[Y_1 - \mathbb{E}(Y_1)] + Z_2(D - 1)[Y_1 - \mathbb{E}(Y_1|C = 0)]\\ V_1 &= Z_1[Y_1 - \mathbb{E}(Y_1)] + Z_2D[Y_1 - \mathbb{E}(Y_1|C = 1)]\end{aligned}$$

by construction we have

$$\begin{aligned}Y &= \mathbb{E}(Y_1) \times (Z_1 + Z_2) + \mathbb{E}(Y_1|C = 0) \times Z_2(D - 1) + V_0\\ Y &= \mathbb{E}(Y_1) \times Z_1 + \mathbb{E}(Y_1|C = 1) \times Z_2D + V_1.\end{aligned}$$

Finally, since $Z_2(1-D) = Z_2(1-C)$ and $Z$ is independent of $Y_1$,

$$\begin{aligned}\mathbb{E}(V_0|Z) &= (Z_1+Z_2)[\mathbb{E}(Y_1|Z) - \mathbb{E}(Y_1)] - Z_2\mathbb{E}\{(1-C)[Y_1 - \mathbb{E}(Y_1|C=0)|Z\} \\ &= -Z_2\mathbb{E}[Y_1 - \mathbb{E}(Y_1|C=0)|C=0, Z] = 0\end{aligned}$$

where the final two steps follow by iterated expectations and the fact that $Z$ is conditionally independent of $Y_1$ given $C$. Since $Z_2 D = Z_2 C$, a nearly identical argument gives

$$\begin{aligned}\mathbb{E}(V_1|Z) &= Z_1[\mathbb{E}(Y_1|Z) - \mathbb{E}(Y_1)] + Z_2\mathbb{E}\{C[Y_1 - \mathbb{E}(Y_1|C=1)|Z\} \\ &= Z_2\mathbb{E}[Y_1 - \mathbb{E}(Y_1|C=1)|C=1, Z] = 0.\end{aligned}$$

□

### E.3 Inference for ASG, ASB, and ASL

We now explain how to carry out cluster-robust inference for the ASG, ASB, and ASL effects, as implemented in our companion STATA package. Each of these effects can be expressed as a difference of coefficients from two just-identified linear IV regressions. The ASG effect is the difference of the TOT effect from Proposition 2 and the TUT effect from Proposition 3. Similarly, the ASB effect is the difference of $\mathbb{E}(Y_0|C=1)$ and $\mathbb{E}(Y_0|C=0)$ from Proposition 4 while the ASL effect is the difference of $\mathbb{E}(Y_1|C=1)$ and $\mathbb{E}(Y_1|C=0)$ from Proposition 5. Within each pair of IV regressions the outcome variable and instrument set is identical; only the regressors differ. Since our estimators of all three effects share the same structure, our discussion of inference abstracts from the specific regressors and instruments used in each case to avoid duplication. To implement these results in practice without relying on our STATA package, simply substitute supply the regressors and instruments specified in the relevant propositions.

Let $g = 1, ..., G$ index clusters and $i = 1, ..., N_g$ index individuals within a particular cluster $g$. In our experiment, a cluster is a branch-day combination and the experimentally-assigned treatment (control, forced, or choice arm) is assigned at the cluster level. We assume that observations are iid across clusters but potentially correlated within cluster. Now consider a pair of just-identified linear IV regressions given by

$$\begin{aligned}Y_{ig} &= \boldsymbol{X}'_{1,ig}\boldsymbol{\theta}_0 + U_{ig} \\ Y_{ig} &= \boldsymbol{X}'_{0,ig}\boldsymbol{\theta}_1 + V_{ig}\end{aligned}$$

with common instrument vector $\boldsymbol{W}_{ig}$. Stacking observations in the usual manner, e.g.

$$\mathbf{W}_g \equiv \begin{bmatrix} \boldsymbol{W}'_{1g} \\ \vdots \\ \boldsymbol{W}'_{N_g g}\end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_G \end{bmatrix},$$

we can write the preceding equations in matrix form as

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\theta}_{\mathbf{1}} + \mathbf{U}$$
$$\mathbf{Y} = \mathbf{X}_0\boldsymbol{\theta}_{\mathbf{0}} + \mathbf{V}$$

with instrument matrix $\mathbf{W}$. Now, the IV estimators for $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_0$ can be expressed as

$$\widehat{\boldsymbol{\theta}}_1 = \left(\mathbf{W}'\mathbf{X}_1\right)^{-1}\mathbf{W}'\mathbf{Y} = \boldsymbol{\theta}_1 + \left(\mathbf{W}'\mathbf{X}_1\right)^{-1}\mathbf{W}'\mathbf{U}$$
$$\widehat{\boldsymbol{\theta}}_0 = \left(\mathbf{W}'\mathbf{X}_0\right)^{-1}\mathbf{W}'\mathbf{Y} = \boldsymbol{\theta}_0 + \left(\mathbf{W}'\mathbf{X}_0\right)^{-1}\mathbf{W}'\mathbf{V}.$$

Our parameter of interest is an element of the difference $(\theta_1 - \theta_0)$, so it suffices to calculate the asymptotic distribution of $(\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_0)$.

Re-arranging the preceding two equations, we have

$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1\right) = \left(\frac{\mathbf{W}'\mathbf{X}_1}{G}\right)^{-1}\left(\frac{\mathbf{W}'\mathbf{U}}{\sqrt{G}}\right) = \left(\frac{1}{G}\sum_{g=1}^{G}\mathbf{W}_g'\mathbf{X}_{1,g}\right)^{-1}\left(\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\mathbf{W}_g'\mathbf{U}_g\right)$$
$$\sqrt{G}\left(\widehat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0\right) = \left(\frac{\mathbf{W}'\mathbf{X}_0}{G}\right)^{-1}\left(\frac{\mathbf{W}'\mathbf{V}}{\sqrt{G}}\right) = \left(\frac{1}{G}\sum_{g=1}^{G}\mathbf{W}_g'\mathbf{X}_{0,g}\right)^{-1}\left(\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\mathbf{W}_g'\mathbf{V}_g\right).$$

Now, define $\mathbf{Q}_0 \equiv \mathbb{E}[\mathbf{W}_g'\mathbf{X}_{0g}]$ and $\mathbf{Q}_1 \equiv \mathbb{E}[\mathbf{W}_g'\mathbf{X}_{1g}]$ and suppose that these expectations exist and that the matrices $\mathbf{Q}_0$ and $\mathbf{Q}_1$ are invertible. Then, as $G \to \infty$

$$\left(\frac{1}{G}\sum_{g=1}^{G}\mathbf{W}_g'\mathbf{X}_{0,g}\right)^{-1} \to_p \mathbf{Q}_0^{-1}, \quad \left(\frac{1}{G}\sum_{g=1}^{G}\mathbf{W}_g'\mathbf{X}_{1,g}\right)^{-1} \to_p \mathbf{Q}_1^{-1}$$

by the continuous mapping theorem. Now, by our experimental design and exclusion restriction, $\mathbf{W}_{ig}$ is independent of $U_{ig}$ both unconditionally and conditional on cluster size. Therefore,

$$\begin{aligned}\mathbb{E}[\mathbf{W}_g'\mathbf{U}_g] &= \mathbb{E}\left[\sum_{i=1}^{N_g}\mathbf{W}_{ig}U_{ig}\right] = \mathbb{E}_{N_g}\left[\mathbb{E}\left\{\sum_{i=1}^{N_g}\mathbf{W}_{ig}U_{ig}\,\middle|\, N_g\right\}\right] \\ &= \mathbb{E}_{N_g}\left[\sum_{i=1}^{N_g}\mathbb{E}(\mathbf{W}_{ig}U_{ig}|N_g)\right] = \mathbb{E}_{N_g}\left[\sum_{i=1}^{N_g}\mathbb{E}(\mathbf{W}_{ig}U_{ig})\right] = 0\end{aligned}$$

by iterated expectations, and similarly smilarly, $\mathbb{E}[\mathbf{W}_g'\mathbf{V}_g] = \mathbf{0}$. Thus, under mild regularity conditions (e.g. finite fourth moments), as $G \to \infty$ we have

$$\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\mathbf{W}_g' \otimes \begin{bmatrix}\mathbf{U}_g \\ \mathbf{V}_g\end{bmatrix} \to_d \mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$$

where the heteroskedasticity– and cluster–robust variance-covariance matrix $\mathbf{\Omega}$ is defined as

$$\mathbf{\Omega} \equiv \mathbb{E}\left[\left(\mathbf{W}_g\mathbf{W}_g'\right) \otimes \begin{pmatrix} \mathbf{U}_g\mathbf{U}_g' & \mathbf{U}_g\mathbf{V}_g' \\ \mathbf{V}_g\mathbf{U}_g' & \mathbf{V}_g\mathbf{V}_g' \end{pmatrix}\right] \equiv \begin{bmatrix} \Omega_{UU} & \Omega_{UV} \\ \Omega_{VU} & \Omega_{VV} \end{bmatrix}.$$

Therefore, the joint limiting distribution of $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_0$ is given by

$$\begin{bmatrix} \sqrt{G}(\widehat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_1) \\ \sqrt{G}(\widehat{\boldsymbol{\theta}}_0 - \boldsymbol{\theta}_0) \end{bmatrix} \to_d \begin{bmatrix} \mathbf{Q}_1^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_0^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_U \\ \boldsymbol{\xi}_V \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{\xi}_U \\ \boldsymbol{\xi}_V \end{bmatrix} \sim \mathrm{N}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \Omega_{UU} & \Omega_{UV} \\ \Omega_{VU} & \Omega_{VV} \end{bmatrix}\right)$$

from which it follows by the continuous mapping theorem that

$$\sqrt{G}\left[(\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_0) - (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0)\right] \to_d \begin{bmatrix} \mathbf{Q}_1^{-1} & -\mathbf{Q}_0^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_U \\ \boldsymbol{\xi}_V \end{bmatrix}.$$

To use this result in practice, we estimate the standard errors of $(\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_0)$ as the square root of the diagonal elements of the matrix

$$\widehat{\mathrm{Avar}}(\widehat{\boldsymbol{\theta}}_1 - \widehat{\boldsymbol{\theta}}_0) = \begin{bmatrix} (\mathbf{W}'\mathbf{X}_1)^{-1} & -(\mathbf{W}'\mathbf{X}_0)^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{UU} & \mathbf{S}_{UV} \\ \mathbf{S}_{VU} & \mathbf{S}_{VV} \end{bmatrix} \begin{bmatrix} (\mathbf{X}_1'\mathbf{W})^{-1} \\ -(\mathbf{X}_0'\mathbf{W})^{-1} \end{bmatrix}$$

where

$$\begin{aligned} \mathbf{S}_{UU} &\equiv \sum_{g=1}^{G} \mathbf{W}_g'\widehat{\mathbf{U}}_g\widehat{\mathbf{U}}_g'\mathbf{W}_g \\ \mathbf{S}_{UV} &\equiv \sum_{g=1}^{G} \mathbf{W}_g'\widehat{\mathbf{U}}_g\widehat{\mathbf{V}}_g'\mathbf{W}_g \\ \mathbf{S}_{VU} &\equiv \mathbf{S}_{UV}' \\ \mathbf{S}_{VV} &\equiv \sum_{g=1}^{G} \mathbf{W}_g'\widehat{\mathbf{V}}_g\widehat{\mathbf{V}}_g'\mathbf{W}_g \end{aligned}$$

and we define the IV residuals

$$\begin{aligned} \widehat{\mathbf{U}}_g &\equiv \mathbf{Y}_g - \mathbf{X}_{1,g}\widehat{\boldsymbol{\theta}}_1 \\ \widehat{\mathbf{V}}_g &\equiv \mathbf{Y}_g - \mathbf{X}_{0,g}\widehat{\boldsymbol{\theta}}_0. \end{aligned}$$

In our application the number of clusters, $G$, is large. If desired, an *ad hoc* degrees of freedom correction can be applied by multiplying the standard errors by $\sqrt{G/(G-1)}$.

## Appendix F. Testable Implications of the Exclusion Restriction

This section describes the testable implications of our exclusion restrictions: (3) and (4) from Section 5.2. To consider possible violations of these conditions, it is helpful to introduce some additional notation. As above, let $Y_0 \equiv Y(d = 0, z = 0)$ and $Y_1 \equiv Y(d = 1, z = 1)$ denote the potential outcomes under *forced treatment*: $Y_0$ is the potential outcome when forced into the status quo contract and $Y_1$ when forced into the commitment contract. Further let $Y_{0,2} \equiv Y(d = 0, z = 2)$ and $Y_{1,2} \equiv Y(d = 1, z = 2)$ denote the potential outcomes under *free choice of treatment*: $Y_{0,2}$ is the potential outcome when choosing the status quo contract and $Y_{1,2}$ when choosing the commitment contract. Using this notation, (3) becomes $Y_0 = Y_{0,2}$ and while (4) becomes $Y_1 = Y_{1,2}$. If we do not impose this pair of equalities, part (iii) of Assumption 1 becomes

$$Y = \mathbb{1}(Z = 0)Y_0 + \mathbb{1}(Z = 1)Y_1 + \mathbb{1}(Z = 2)\left[(1 - C)Y_{0,2} + CY_{1,2}\right]$$

but parts (i) and (ii) continue to hold. Accordingly, parts (i)–(iii) of Lemma 1 are unchanged, while parts (iv) and (v) become

$$\mathbb{E}(Y|D = 0, Z = 2) = \mathbb{E}(Y_{0,2}|C = 0), \quad \mathbb{E}(Y|D = 1, Z = 2) = \mathbb{E}(Y_{0,1}|C = 1).$$

Using these expressions, the testable restrictions we consider here are as follows:

$$\mathbb{E}(Y_0|C = 0) = \mathbb{E}(Y_{0,2}|C = 0) \tag{19}$$

$$\mathbb{E}(Y_1|C = 1) = \mathbb{E}(Y_{1,2}|C = 1). \tag{20}$$

Equation 19 is a restriction on the average potential outcomes of non-choosers that we use to point identify the TUT effect. It says that someone who *would* choose the control condition when given a choice, experiences the same potential outcome, on average, when *assigned* to the control condition. Equation 20 is a restriction on the average potential outcomes of choosers that we use to point identify the TOT effect. It says that someone who *would* choose the commitment contract when given a choice, experiences the same potential outcome, on average, when *assigned* to this condition. Because they refer to different groups of people, either of (19) and (20) could hold when the other is violated. For this reason we consider each in turn. Our approach is closely related arguments from Huber & Mellace (2015) and DiTraglia & Garcia-Jimeno (2019), among others. As such, the following is a heuristic explanation rather than a fully-rigorous proof. See the aforementioned papers for more details of how to make this argument fully rigorous.

Consider first (19). Let $p \equiv \mathbb{P}(C = 1) = \mathbb{P}(D = 1|Z = 2)$ denote the share of choosers in the population. This value is point identified regardless of whether the exclusion restriction holds. Because $Z$ was randomly assigned, a fraction $p$ of borrowers with $Z = 0$ are choosers while the remaining $(1 - p)$ are non-choosers. It follows that, regardless of whether the exclusion restriction holds, the observed distribution of $Y|Z=0$ is a mixture of $Y_0|C=0$ and $Y_0|C=1$ with mixing weights $(1-p)$ and $p$. This allows us to construct a pair of bounds for $\mathbb{E}(Y_0|C=0)$ as follows. The non-choosers must lie *somewhere* in the distribution of $Y|Z=0$. Consider the two most extreme possibilities: they could occupy the bottom $(1-p)\times 100\%$ of the distribution or the top $(1-p)\times 100\%$ of the distribution. For this reason, computing the average of the the *truncated* distribution of $Y|Z=0$, cutting out the top $p\times 100\%$, provides a lower bound for the average of $Y_0$ among non-choosers. Similarly, cutting out the bottom $p\times 100\%$ provides an upper bound. Let $y^0_{1-p}$ denote the $(1-p)$ quantile of $Y|Z=0$ and $y^0_p$ denote the $p$ quantile of the same distribution. Using this notation, the bounds are given by

$$\mathbb{E}\left(Y|Z=0, Y\leq y^0_{1-p}\right)\leq \mathbb{E}(Y_0|C=0)\leq \mathbb{E}\left(Y|Z=0, Y\leq y^0_p\right)$$

These bounds do not rely on the exclusion restriction. Under Equation 19, however, we know that $\mathbb{E}(Y_0|C=0)=\mathbb{E}(Y|D=0,Z=2)$. Therefore, if the exclusion restriction for non-choosers holds, we must have

$$\mathbb{E}\left(Y|Z=0, Y\leq y^0_{1-p}\right)\leq \mathbb{E}(Y|D=0,Z=2)\leq \mathbb{E}\left(Y|Z=0, Y\leq y^0_p\right). \tag{21}$$

Equation 21 provides a pair of testable implications of (19). If either inequality is violated, then the exclusion restriction for non-choosers fails.

We can use an analogous approach to construct testable implications for 20. Because $Z$ was randomly assigned, a fraction $p$ of participants with $Z=1$ are choosers so the distribution of $Y|Z=1$ is a mixture of $Y_1|C=1$ and $Y_1|C=0$ with mixing weights $p$ and $1-p$. The participants with $C=1$ must lie *somewhere* in the distribution of $Y_0|Z=1$ so again consider the two most extreme cases: they could occupy the bottom or the top $p\times 100\%$ of the distribution. Hence,

$$\mathbb{E}\left(Y|Z=1, Y\leq y^1_p\right)\leq \mathbb{E}(Y_1|C=1)\leq \mathbb{E}\left(Y|Z=1, Y\geq y^1_{1-p}\right)$$

where $y^1_p$ and $y^1_{1-p}$ are the $p$ and $1-p$ quantiles of the distribution of $Y|Z=1$. As above, these bounds do not rely on the exclusion restriction. If Equation 20 holds, however, we know that $\mathbb{E}(Y|D=1,Z=2)=\mathbb{E}(Y_1|C=1)$, yielding the following testable implications for choosers

$$\mathbb{E}\left(Y|Z=1, Y\leq y^1_p\right)\leq \mathbb{E}(Y|D=1,Z=2)\leq \mathbb{E}\left(Y|Z=1, Y\geq y^1_{1-p}\right). \tag{22}$$

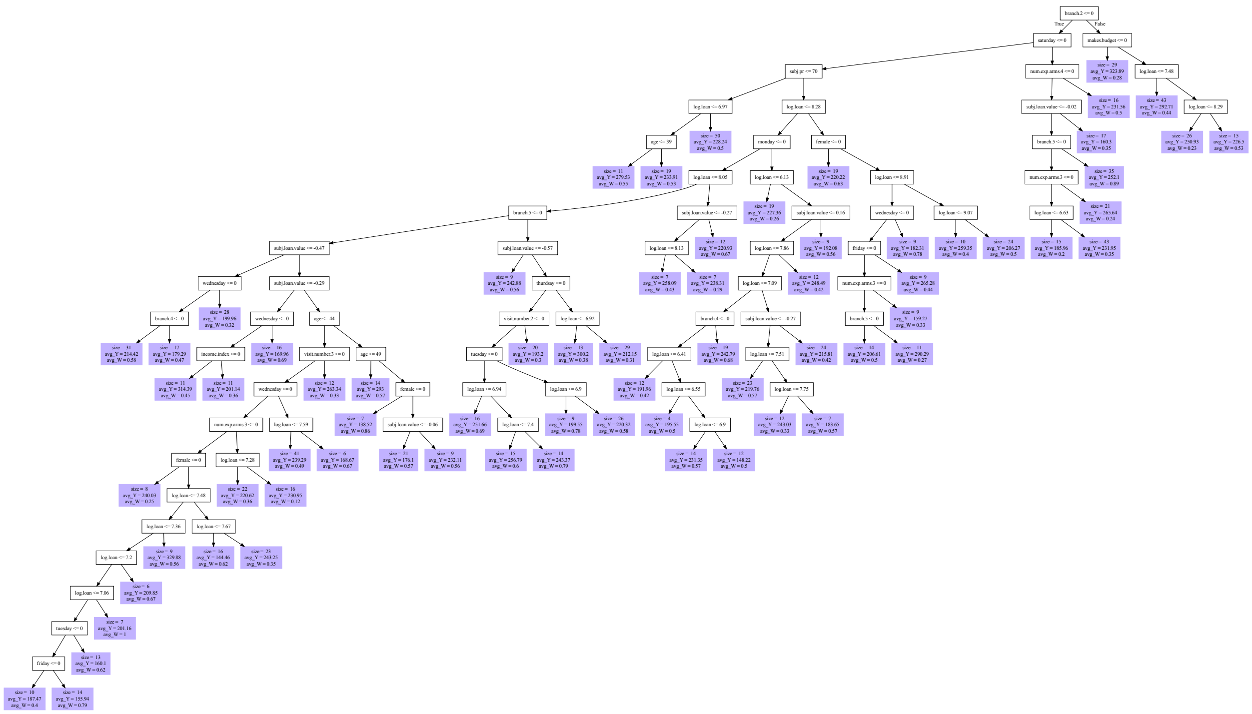
If either inequality is violated, then the exclusion restriction from Equation 20 fails.

The Huber & Mellace (2015) bounds for the TOT & TUT APR are respectively: $[-366, 490]$, and $[-19, 109]$. Since both TUT (24) and ToT (122) estimates from the LATE approach lies inside these bounds, we cannot falsify the exclusion restriction. Moreover, the p-values associated with testing violation of inequalities 21, and 22 is equal to one in both cases[33].

[33] We include the implementation of the validity test for the exclusion restriction for choosers and non-choosers in

# Appendix G. Causal Random Forest, HTE, and 'mistakes'

Figure OA-13: Example Causal Tree



This figure shows an example of one of the regression trees used to compute the causal forest estimates described in Section 6.2. Each tree is constructed from a sequence of recursive binary partitions of the covariate space, with a splitting rule that is targeted to the causal quantity of interest, following the methodology of Athey *et al.* (2019).

our STATA package.