The Review of Financial Studies

Credit Building or Credit Crumbling? A Credit Builder Loan's Effects on Consumer Behavior and Market Efficiency in the United States

Jeremy Burke
Center for Economic and Social Research, University of Southern California, USA

Julian Jamison
University of Exeter Business School, UK

Dean Karlan
Kellogg School of Management, Northwestern University, USA

Kata Mihaly
RAND Corporation, USA

Jonathan Zinman
Dartmouth College, USA

We thank the St. Louis Community Credit Union and especially Paul Woodruff for cooperation; the Consumer Financial Protection Bureau (CFPB), in particular Sarah Bainton Kahn and Daniel Dodd-Ramirez for their assistance and preparation of a policy brief based this study; participants at many conferences and seminars for comments; Innovations for Poverty Action, in particular Anna Cash, Lucia Goin, Nora Gregory, Ejin Kim, Peter Lugthart, and Kayla Wilding for research support. We gratefully acknowledge research funding provided by the Consumer Financial Protection Bureau [under competitive award CFP-12-Z-00020/0002]. The views expressed are those of the authors alone and are not necessarily shared by the CFPB or any other arm of the U.S. government. Institutional Review Board approval for human subjects protocols from Innovations for Poverty Action [no. 14January-001] and RAND Human Subjects Protection Committee [no. 2013-0660]. This study was registered with the AEA RCT Registry with the ID number AEARCTR-0000441. The credit bureau that provided data to us had the right to (a) review the paper to ensure that the analysis using credit scores was depersonalized, aggregated, and that the scores received the correct trademark attribution and (b) offer comments about the paper, which the authors agreed ex-ante to consider in good faith. The authors retained intellectual freedom to report the results of this study regardless of the outcomes. We obtained bureau data through "soft" credit pulls that do not impact consumer credit scores. FICO® is a registered trademark of the Fair Isaac Corporation. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Jonathan Zinman, jzinman@dartmouth.edu.

The Review of Financial Studies 36 (2023) 1585–1620

https://doi.org/10.1093/rfs/hhac060 Advance Access Publication September 2, 2022

A randomized encouragement design yields null average effects of a credit builder loan (CBL) on consumer credit scores. But machine learning algorithms indicate the nulls are due to stark, offsetting treatment effects depending on baseline installment credit activity. Delinquency on preexisting loan obligations drives the negative effects, suggesting that adding a CBL overextends some consumers and generates negative externalities on other lenders. More favorably for the market, CBL take-up generates positive selection on score improvements. Simple changes to CBL practice, particularly to provider screening and credit bureau reporting, could ameliorate the negative effects for consumers and the market. (*JEL* D12, G14, G21)

Received February 9, 2021; editorial decision July 14, 2022 by Editor Tarun Ramadorai. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

Consumer credit histories are important inputs to various markets. Lenders use them in determining willingness to ration or lend, and at what terms. Many landlords, insurers, and employers now use them when evaluating potential customers or employees (Bartik and Nelson 2021; Bos, Breza, and Liberman 2018; Dobbie et al. 2020). Yet a majority of credit users in the United States have below-prime credit scores (Brooks et al. 2015), and about 20% of the U.S. population is "credit invisible" due to thin or nonexistent credit bureau files (Brevoort, Grimm, and Kambara 2015). "Alternative data"—data beyond credit histories on standard loan products—can help price credit risk more accurately and reveals many of these consumers to be "invisible primes" (Brevoort and Kambara 2017; Di Maggio, Ratnadiwakara, and Carmichael 2021). One source of such data is payment behavior on "fresh start" or credit builder loans (CBLs).

CBLs are short-term installment contracts on small amounts in which the "lender" eliminates its credit risk by inverting the sequence of origination and repayment: "loan" proceeds are held in an escrow account and only released after one or all of the contracted payments, which include principal and an administrative fee, are made. The CBL thus operates less like a loan and more like either a costly commitment savings device (if individuals do not withdraw the funds until the end of the CBL term) or a costly sequence of deposits and withdrawals (if individuals choose to withdraw the funds immediately after making each payment). Nevertheless, and crucially, credit reporting treats CBLs as standard installment loans, per industry agreements between CBL providers and the three major credit bureaus. And as with standard loans, CBL providers report all CBL payment performance to the bureaus, both timely and late.

CBLs are widely available, and prominent financial self-help resources like NerdWallet and Credit Karma provide advice on how to access and manage them. Most CBL suppliers are credit unions like our partner or community banks, although many digital lenders and other "fintechs" have entered the market recently. Regulators are beginning to take an interest in the market, as evidenced by the Consumer Financial Protection Bureau commissioning and funding our study.

Like any credit building intervention, CBLs could affect consumers, providers, and markets alike. For consumers, CBLs could help them become credit visible, or shift their credit scores up or down. Our descriptive evidence suggests that both shifts likely occur with some frequency; for example, 40% of CBL users in our sample pay more than 30 days late on their CBL at some point. For providers, CBLs provide marginal customers a point of entry or reentry into the mainstream financial system, opening the possibility of cross-sells. For the market, via CBL providers reporting to credit bureaus, CBLs could help market efficiency, for example, if CBL take-up predicts downstream behavior in ways that are not fully captured by other observables. Or CBLs could harm market efficiency, if, for example, CBL behavior or reporting provides misleading signals or CBL usage creates negative spillovers by inducing delinquency on preexisting loans.

We start by estimating CBL treatment effects on consumers, and on lender cross-sells, using an encouragement design that randomizes take-up requirements. St. Louis Community Credit Union (SLCCU) has offered CBLs since 2009 and worked with the research team from September 2014 through February 2015 to identify a sample of over 1,500 SLCCU members who expressed interest in a CBL. As such our sample is drawn from a population of great interest for research, practice, and policy: consumers close to the margin of entering the market for credit building products. Nearly 20% of our sample lacked a FICO® Score at baseline, and scores are low overall among those who were scorable.

We then randomly assigned these individuals to one of two arms: a "CBL Arm" that followed SLCCU's standard enrollment process for a CBL, and an "Extra Step Arm" facing an additional "requirement" (that ended up being unenforced by staff) to complete five modules of online financial education, taking about 50–60 minutes in total, either onsite or offsite prior to opening a CBL. Only six individuals in the Extra Step Arm even started the online financial education, and thus the financial education itself should have not have a direct treatment effect. But the financial education requirement did contribute to a large take-up differential across the two arms: the CBL Arm had a take-up rate of 30% within 18 months of entering the study, while the take-up rate in the Extra Step Arm was only 12%. This first stage, and our sample of those interested in CBLs, identifies CBL treatment effects for marginal consumers, namely, those who are encouraged by SLCCU's marketing push and/or deterred by the extra requirement. We will discuss implications for external validity in Section 2.5.

We measure FICO® Scores and credit market behaviors using four data pulls obtained from one of the three major credit bureaus: one at baseline, and three more at endlines of roughly 6, 12, and 18 months post-random assignment. Our two main outcomes are whether the consumer has a FICO® Score, and their score conditional on having one at baseline. Having a credit score is an important step for consumers in becoming credit-visible and potentially signaling a positive credit history. It is also an important step for lenders and the market in the sense that a scoring company only reports a consumer's score when it has sufficient confidence in its predictive power. The numerical credit score itself is important, as discussed above, because of its widespread use in credit and other markets.

Averaging across the three endlines, we find a null average intent-to-treat (ITT) effect of the CBL on the likelihood of having a credit score. We also find a precisely estimated null average treatment effect on the credit score, among the subsample of individuals with a credit score at baseline.

These null average effects obscure important heterogeneous treatment effects (HTEs), most starkly by baseline installment credit activity. We are motivated to examine this margin of heterogeneity by theory, practice, and machine learning estimation designed to "let the data speak."

In theory, those with existing loans may benefit less from CBLs since they already have a recent credit history. Moreover, those with existing installment loans may struggle to manage their existing loan obligation(s) in tandem with a CBL if cash flows are tight, especially since recent evidence suggests that small expense shocks can trigger loan delinquency (Mello 2022; Wong 2020). Staying current on installment debt may be costly, in liquidity terms, relative to revolving loans and the CBL itself, since revolving credit offers more repayment flexibility, and the CBL has a small monthly repayment that can be refunded immediately after making it. On the other hand, successful CBL use should boost scores even for those with existing loans, and those with existing loans may have experience and/or better access to liquidity that helps them successfully manage the CBL.

In practice, baseline installment borrowing is prevalent, and readily observable. Should it drive treatment effects, any CBL provider could market and screen on it.

We let the data speak in two steps. First, we use a causal forest aggregate test for overall treatment effect heterogeneity. This test strongly rejects the hypothesis of homogeneous treatment effects on credit scores at the first endline and finds suggestive evidence of heterogeneity at the second endline. Second, we examine readily observable potential correlates of the causal forest's predicted conditional average treatment effect (CATE) for each consumer. These tests strongly reject the hypothesis of homogeneity with respect to baseline installment activity. Most strikingly, those in the bottom tercile of the distribution of installment credit activity at baseline have a mean CATE on their 6-month credit score of +15 points (SE 7 points), while those in the top tercile have a mean CATE of −17 points (SE 6 points). These effects are large enough to move someone across credit score bins that affect market access and terms.[1] We examine many other potential drivers of HTEs, but none is as robustly significantly correlated with CATEs in statistical or economic terms.

[1] Many of these bins span ranges of only 20 to 40 points. See, for example, https://www.myfico.com /credit-education/calculators/loan-savings-calculator/ (accessed January 1, 2022).

What, mechanically, produces the HTEs, including CBLs backfiring for some consumers? One possibility is differential firm behavior, with FICO® scoring the same behavior differently for people with different credit histories. We cannot test that hypothesis, as we are not privy to the proprietary model behind the FICO® Score. Another possibility is differential consumer behavior. And indeed we find HTEs on two categories of behaviors that factor into credit scoring: credit mix and repayment performance. The repayment performance results are the most striking, with no evidence of TEs on delinquency for those in the lower two terciles of baseline installment activity, but 0.22 SD more delinquency (SE 0.08 SD) for those in the CBL arm and the top tercile. The bulk of this effect is likely driven by *non*-CBL delinquency. Thus, even though the CBL studied here imposes minimal liquidity constraints in principle, adding a CBL to existing credit obligations seems too much for many borrowers to manage successfully in practice.[2]

Why, from a consumer decision-making standpoint, do CBLs backfire for some consumers? A behavioral model with limited attention to future liquidity constraints (Bronchetti et al. 2021) and/or overconfidence about making future payments could explain the pattern (Heidhues and Köszegi 2010). Alternative behavioral explanations do not make as clear predictions or fit our full pattern of results as nicely. For example, concepts of scarcity (Mullainathan and Shafir 2013) yield indeterminate predictions: default on CBLs and other loans could increase because of bandwidth constraints and/or decrease because of hyper-focused tunneling (Kaur et al. 2021; Lichand and Mani 2020; Ong, Theseira, and Ng 2019). Nor is consumer confusion about CBLs a likely explanation in our setting, where both marketing and high-touch interactions with staff likely provided accurate and reinforcing information, and our proxies for financial literacy and experience do not moderate CBL treatment effects (although future work might consider alternative measures of these constructs). General confusion also fails to explain why we find delinquency increases on other installment loans, but not revolving loans. But consumer confusion will be important to consider in other settings, given the prevalence of scams in credit repair and related product markets.

Turning to treatment effects on other SLCCU products (cross-sells), some evidence suggests that the CBL increases savings balances. This is consistent with some consumers using the CBL for what it is, functionally, aside from the credit reporting: a costly commitment to save. For other SLCCU outcomes, we find no evidence of effects on customer retention, and some evidence that non-CBL borrowing from SLCCU increases for those in the bottom tercile of baseline installment activity.

[2] We attempted to engage participants in qualitative follow-up discussions to better understand participants' experiences with the CBL, particularly regarding cash flow management, but we were stymied by a low response rate.

Last, but not least, we examine impacts of the CBL on market information, using various predictive tests. Our main test focuses on self-selection: on whether CBL take-up reveals information about a consumer's future credit score. We find that CBL takers, relative to nontakers in the CBL Arm, show estimated credit score improvements of 11 points (SE 3 points). In theory, this upward trend is a combination of selection and the CBL average treatment effect. In practice, since the average ITT effect is a precisely estimated zero, the upward trend reveals strong positive (advantageous) selection: those who choose to open a CBL are improving irrespective of the CBL itself. This suggests that CBL take-up provides a valuable signal to lenders, and that credit bureaus should consider reporting CBLs as a distinct category rather than lumping them together with standard installment loans. We find little evidence of differential selection across our study arms, which strengthens external validity.

All told, we add to extant literatures in several respects. First, we use random variation to help separately identify CBL selection and treatment effects on credit behaviors and scores (see Liberman et al. (2021) for a similar approach to the U.K. payday loan market), adding evidence on a credit-building product to the literature on programmatic interventions (Kaiser et al. 2021). Second, and closely related to the first, our findings that a CBL with modest liquidity requirements causes delinquency on *non*-CBL loans, at least for those with preexisting installment debt, adds to work on default spillovers (De Giorgi, Drenik, and Seira forthcoming) and on consumer cash flow management and financial distress (e.g., Gelman et al. 2020; Olafsson and Pagel 2018; Dobbie and Song 2020). Third, we replicate and expand on the key finding from CBL industry reports—CBL usage is advantageously selected (Chenven 2014; Wolff 2016)—and infer that credit bureaus could better harness this information revelation by reporting CBLs as a distinct product category. We thereby build bridges to work on credit history as a public good that may lead for-profit firms to underinvest in information acquisition (e.g., Petersen and Rajan 1995), and on whether and how credit bureaus reduce asymmetric information and information costs (e.g., de Janvry, McIntosh, and Sadoulet 2010; Hertzberg, Liberti, and Paravisini 2011; Manso 2013; Garmaise and Natividad 2017). Fourth, our findings suggest that "product-linked" financial education requirements may be counterproductive, despite strong policy and programmatic interest in that approach (Askari 2009; Sledge, Gordon, and Kinsley 2011; Reyes et al. 2013). Fifth, our evidence on how CBLs backfire for some consumers adds to various strands of work on how credit market risk modeling technologies and practices produce disparate outcomes (e.g., Hurst et al. 2016; Fuster et al. 2022; Blattner and Nelson 2021).

In terms of practical takeaways, after appropriate caveats regarding external validity limitations, we discuss two key implications: (1) CBLs as currently constituted likely have a mix of positive and negative effects on consumers and market efficiency, and (2) simple changes in how providers target and credit bureaus report CBLs could produce more uniformly positive effects.

1. Study Setting and Design

1.1 Implementing partner and credit building product

We partnered with St. Louis Community Credit Union (SLCCU) to design and implement our study. SLCCU, a certified Community Development Financial Institution (CDFI), serves approximately 51,000 members who live or work in the greater St. Louis area. SLCCU has 11 branches (including three located within social service agencies), provides access to online financial education and phone-based credit counseling and education, and offers numerous financial products designed to improve members' financial stability. SLCCU has offered the "Credit Builder Loan" ("CBL") since 2009 and had originated approximately 4,400 CBLs at the onset of the study.

SLCCU markets and structures the CBL per credit union and CDFI industry standards. It markets the CBL as an opportunity to build credit history and improve credit scores (Figure 1 shows the marketing materials used by SLCCU, both in our study and routinely). The terms are such that no money changes hands at origination. Instead, the credit union places $600 in a restricted access savings account (an escrow account, basically). Borrowers then make 12 monthly payments of approximately $54 and the credit union releases $50 from the restricted savings account back to the consumer's regular savings account immediately upon receipt of payment each month.[3] As such, the payments portion of the CBL functions like a costly commitment savings account, yielding a certain and negative pecuniary return on saving; for example, if the consumer makes all 12 CBL payments and does not make any withdrawals, they will have invested $648 over the course of the year and yielded $600 at year's end. There are no other pecuniary costs for SLCCU CBLs, nor does SLCCU pull a credit report for CBL applicants or users.

CBL payments, both timely and late ("delinquent," in credit bureau parlance), are reported to each of the three major credit bureaus as a standard installment loan, using standard definitions of delinquency (e.g., a loan is considered timely if it is <30 days late, and first reported delinquent if ≥ 30 days late). As such, for 30-day delinquencies CBLs are reported just like any other loan. After that point, many providers, including SLCCU, will use the escrowed proceeds to close out the CBL. This prevents the CBL from incurring more serious delinquency—although the initial 30-day delinquency persists in credit reports for seven years, per U.S. regulations governing negative credit information.

[3] Credit unions tend to calibrate the CBL fee, $4 per month in our case, to cover the cost of the staff time required to administer the CBL, with the intent of generating returns downstream through cross-sells and/or helping their membership (credit unions are mutually owned and often operate like nonprofits).

Figure 1
CBL marketing materials

Approximately 40% of CBL users in our sample made at least one payment more than 30 days late (Figure 2). This high rate of delinquency indicates that CBLs could backfire, at least for some borrowers.

1.2 Data

We have three data sources: a baseline survey, SLCCU administrative account data, and FICO® Scores and credit report attributes from one of the three major credit bureaus. Surveyors administer the baseline survey as part of the CBL marketing process, as will be described below. The survey captures demographics and some aspects of financial status and attitudes. SLCCU administrative data are pulled monthly for everyone in our sample. These

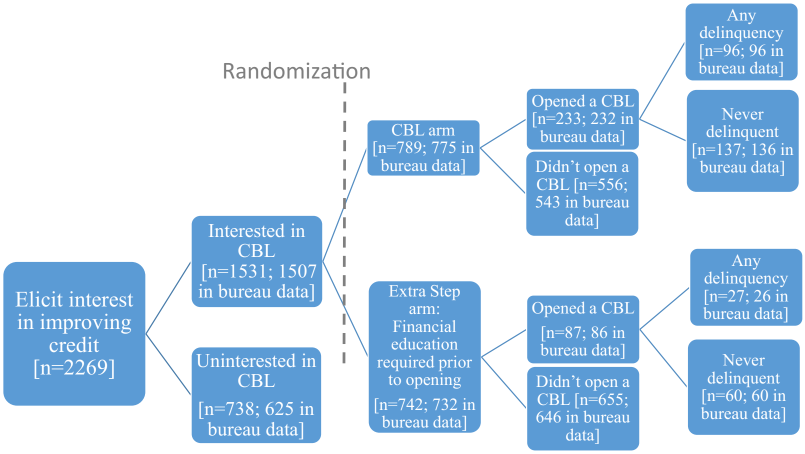

Figure 2
Sample construction, experimental design, and CBL payment performance
Sample sizes include only those individuals matched to the credit union's administrative data and hence inferred to be a credit union member at baseline. The sample sizes shown to be "in bureau data" refer to those in the study sample whom we were able to match to a credit report at baseline. CBL = Credit builder loan.

data capture CBL performance and the usage of other loan and deposit products.

The bureau data capture snapshots of borrowing and repayment activity and one widely used credit score, the FICO® Score. We obtain snapshots at baseline (on a biweekly rolling basis as participants entered the study), at approximately 6 and 12 months post-random assignment, and at $\geq$18 months post-assignment (with a maximum of 24 months, depending on assignment date). The credit bureau did not share loan-level data; for example, our measure of 30-day delinquency is the number of loans, include any CBL, on which the person is $\geq$30 days late. Some bureau variables are disaggregated to the person $\times$ loan type-level—for example, number and balance of installment or revolving loans—but not delinquency. CBLs are reported as installment loans, both in our data and in the credit reports visible to lenders and other firms.

1.3 Sampling and experimental design

Figure 2 illustrates our sampling and experimental design. Our goal for *survey sampling* was to create a sample frame of SLCCU members who are generally interested in improving their credit. Between October 2014 and February 2015, research staff ("surveyors") enrolled participants into the study at seven of the SLCCU branches. Surveyors approached individuals in the branch and first asked if they were generally interested in building their credit. Individuals responding affirmatively were escorted to a private office and asked for consent to participate in a "research study focused on credit markets and products."

In total, 2,310 individuals consented and started the short baseline survey. Of these 2,310, 2,269 were SLCCU members at baseline, as evidenced by a match to SLCCU administrative data.

Our goal for *the experiment* was to engineer variation in CBL take-up within a sample of SLCCU members who are interested in a CBL. After the survey, surveyors described the CBL and elicited participant interest in the CBL specifically (as distinct from credit building generally). We remove the 738 "Uninterested" individuals from the experiment sample: we do not randomly assign these individuals to an experimental arm. The remaining 1,531 expressed interest in the CBL and comprise the "experimental sample." Surveyors randomized these 1,531 participants, in real-time and at the individual level, into one of two arms.

Members in the "CBL Arm" were encouraged to open the CBL on the spot, per standard SLCCU procedures. As such, the CBL Arm got a standard encouragement treatment: an intense marketing push. Members in the "Extra Step Arm" were encouraged to open the CBL, but they were told they must first complete approximately 50 minutes of free online financial education prior to opening. As such, the Extra Step Arm received a mix of encouragement and discouragement, with the latter taking the form of a take-up friction. The financial education course is one of SLCCU's standard offerings and clients can complete it from a branch computer or any other web-connected device. Credit union staff could waive the financial education requirement for individuals in the Extra Step Arm, and they often did: only six individuals started the course, and only two completed it. The Extra Step nevertheless contributes to engineering the desired experimental variation in CBL take-up, as we document in Section 2.1.

For additional details on survey administration, marketing, randomization, and financial education content, see the Internet Appendix Section B-1.

We will discuss what is required under this design to disentangle selection from treatment effects in Section 2.4.1 and how to interpret treatment effects vis-à-vis external validity in Section 2.5. Interpreting results would be more straightforward under a design that *either* encouraged or discouraged some randomly assigned consumers, with the other arm being a more traditional control arm of being clearly business as usual, but we expected power constraints ex ante and thus prioritized a design that would maximize the take-up differential across the two arms.

1.4 Sample characteristics and randomization balance

Table 1 presents baseline summary statistics and randomization balance tests, on our experiment sample, for 17 key outcome variables and sources of potential heterogeneity. Columns 1 and 2 present descriptive statistics, separately for the Extra-Step (N = 742) and CBL (N = 789) Arms. Column 3 presents an estimate of the difference across the two arms for each variable. The overall pattern is

Table 1
Baseline characteristics and randomization balance for experiment sample

	(1)	(2)	(3)
	Mean (SD)		Univariate t-test diff:
Sample:	Extra Step Arm N=742	CBL Arm N=789	(2) - (1) (SE)
Age	43.823	42.475	−1.348
	(15.056)	(15.328)	(0.777)
Female	0.642	0.655	0.014
	(0.480)	(0.476)	(0.024)
Married	0.241	0.229	−0.012
	(0.428)	(0.421)	(0.022)
Number of adults in household	1.611	1.629	0.019
	(0.788)	(0.791)	(0.041)
Number of children in household	0.845	0.807	−0.038
	(1.237)	(1.229)	(0.064)
Race - Black	0.875	0.883	0.008
	(0.331)	(0.322)	(0.017)
College or more	0.264	0.253	−0.011
	(0.441)	(0.435)	(0.023)
Financial risk-taking scale (standardized)	0.000	0.039	0.039
	(1.000)	(1.008)	(0.052)
Self-control and credit knowledge index (standardized)	0.000	0.051	0.051
	(1.000)	(0.947)	(0.050)
Liquidity index (standardized)	0.000	−0.005	−0.005
	(1.000)	(0.928)	(0.049)
Delinquency index (standardized)	0.000	−0.074	−0.074
	(1.000)	(0.925)	(0.050)
1 = Higher than median of index of default outcomes	0.595	0.598	0.004
	(0.491)	(0.491)	(0.025)
1 = Scored on FICO®	0.840	0.810	−0.030
	(0.367)	(0.393)	(0.020)
Baseline FICO®Score	561.489	564.256	2.767
	(64.317)	(66.749)	(3.727)
Installment credit activity at baseline index (standardized)	0.000	−0.047	−0.047
	(1.000)	(1.000)	(0.052)
Revolving credit activity at baseline index (standardized)	0.000	0.006	0.006
	(1.000)	(1.026)	(0.052)
Number of prior loans, lifetime	7.773	7.220	−0.553
	(9.131)	(7.725)	(0.445)

Unit of observation is an individual. Index variables are standardized to the Extra Step Arm; see Internet Appendix B-2 for details on index components and construction. Sample size varies across variables due to missing observations.

consistent with a valid randomization: only one variable has a difference that is close to statistically significant at conventional cutoffs, and the difference on that variable (age) is economically small. A caveat is that many of the statistically null point estimates here have confidence intervals that include economically meaningful differences.

The demographics of our experiment sample are predominantly female, unmarried, and Black. Only 25% of our sample has a college degree. The mean age is about 43, with a standard deviation of 15, and the support of its distribution spans most working ages. Credit bureau data tend to report limited, if any, demographic information because of data and legal limitations, but,

Table 2
Transition matrix for having a credit score

	(1)	(2)	(3)	(4)	(5)	(6)
	CBL Arm			Extra Step Arm		
		Have score at 18-month endline	No score at 18-month endline		Have score at 18-month endline	No score at 18-month endline
	N=	668	91	N=	632	85
Have score at baseline	622	97%	3%	609	95%	5%
No score at baseline	137	47%	53%	108	49%	51%

Unit of observation is an individual. Sample size is slightly reduced from baseline because here it is limited to persons with a credit report at our 18-month endline.

to the best of our knowledge, our sample is similar to the low-to-unscored population.[4]

In terms of credit history, a bit more than 80% of our sample has a FICO® Score at baseline. Table 2's transition matrices show that most movement on this variable goes in the direction of obtaining a score: nearly 50% of those unscored at baseline are scored at the 18-month endline, while only about 4% of those scored at baseline lack a score at the 18-month endline. A consumer can have a credit report with information on specific debts, without being scored, if FICO® cannot estimate risk with sufficient confidence.

Returning to Table 1, we see that scores are low on average among those with scores, albeit with substantial heterogeneity: the mean is about 560 and the standard deviation about 65. FICO® Scores can range from 300 to 850, with a national average of about 700, and most of our sample is well below common cutoffs for a "prime" borrower (usually 640 or 680). Subprime consumers typically face high prices and rationing (see, e.g., the evidence on utilization in the next paragraph). Many individuals have substantial past borrowing experience, with a mean and standard deviation of lifetime loans of about eight each. And many individuals have outstanding loans at baseline: over 60% have one or more installment loans, and over 45% have one or more revolving loans.[5] Nearly 50% of these borrowers have been delinquent during the past 12 months.

Focusing next on liquidity, we see that liquid asset holdings at SLCCU are low for most of the sample: 64% holds less than the required CBL monthly payment amount ($54) in their SLCCU deposit accounts at baseline. (The 1/0 variables for baseline borrowing activity, delinquency, and liquid assets are not shown in Table 1 because they are each part of broader indexes that are

4 Internet Appendix Table 1 compares our sample's demographics to a plausibly nationally representative sample of the left part of the creditworthiness distribution, specifically to the 17% of people in the 2018 National Financial Capability Survey who self-report their credit history as "Very bad" or "Bad." We see strong similarity on age, gender, income, and number of children. Our consumers much more likely to be non-white, substantially less likely to be married, and somewhat more educated. NFCS does not report the one other demographic we measure: total number of adults in the household.

5 The traditional credit bureaus have broad, but not entirely comprehensive, coverage of borrowing, so some people we classify as nonborrowers may in fact have an outstanding loan.

shown.) And among those with an open credit line at baseline, mean utilization is greater than 100%: the average person with a revolving credit line in our sample has exceeded their credit line(s). Together with prevalent low credit scores and delinquency, these patterns suggest that liquidity constraints bind for most of our sample.

2. Results

2.1 Average treatment effects

Table 3 presents ordinary least squares (OLS) average intention to treat (ITT) estimates for our key first- and second-stage outcomes. These are our main estimates of average treatment effects. ITT estimates tend to be preferred for policy purposes, since they provide the average impact on the target customers from offering a new product or service or policy. A rough estimate of the TOT would inflate the ITT coefficients by the reciprocal of the differential take-rate between the two experimental arms; as we shall see next, that is, $1/.18 \approx 5.5$, in our case.

Column 1, panels A and B, shows two estimates of the first stage. Our randomization induced large differences in CBL take-up across the CBL and Extra Step Arms, whether we count all take-up within 18 months post-offer (18 pp with a SE of 2 pp), or only take-up within the first 30 days (16 pp, SE 2 pp).[6] This strong first stage serves two purposes. The first is methodological: it enables us to estimate the causal effects of CBL access on downstream outcomes in columns 2 and 3 and subsequent tables. The second is substantive: it sheds light on the deterrent effect of financial education (and possibly other take-up frictions that add time or hassle costs), even when financial education is offered through a convenient delivery channel and at a seemingly opportune moment.[7]

Columns 2 and 3 show estimated effects on our main second-stage/downstream outcomes: having a credit score, and credit score conditional on having one at baseline. Here we use the four credit reports we have per-person, and our random assignment to either the CBL or Extra-Step Arm, to estimate intent-to-treat (ITT) effects using OLS equations of the following form:

$$Y_{it} = \alpha + \beta(CBL\ Arm \times Post_t) + \gamma Post_t + \sum_i \delta_i I_i + \varepsilon \tag{1}$$

[6] Approximately 53% of take-up occurred on the same day as the survey and offer; 68% occurred within the first 30 days; and 97% occurred within the first year. After the first 30 days, take-up is statistically indistinguishable across the two arms (see Section 2.4.2 for details). Internet Appendix Table 2 shows our key baseline characteristics do not have strong univariate correlations with take-up overall; for example, 26 of 28 p-values are $> .05$, and 25 are $> .10$ (columns 3 and 6), which is a pattern consistent with a lack of any true predictors of take-up, subject to the caveat that one potentially noteworthy exception is that takers in the CBL arm have lower credit scores than nontakers (-14 points, SE 6). Another is that takers in the Extra Step arm are more educated than nontakers (14 pp more likely to be college-educated, SE 5). We will discuss that result in Section 2.4.2.

[7] As detailed in Section 1.3, the financial education requirement for the Extra Step Arm deterred take-up even though it was not enforced: only six individuals even started the course. This implies that our treatment effect estimates need not account for the possibility that consumers in the Extra Step Arm benefitted from financial education.

Table 3
OLS average ITTs on CBL takeup and credit scores

	(1a)	(1b)	(2a)	(2b)	(3a)	(3b)
	First stage: CBL take-up		Key Second-stage outcomes: Credit scores			
Dependent variable:	1 = Within 18 months of offer	1 = Within 30 days of offer	1 = has FICO®Score 8		FICO®Score 8	
Sample:	Full		Full		Have score at baseline	
CBL Arm	0.178	0.161				
	(0.020)	(0.017)				
CBL Arm * Post			0.018		−1.888	
			(0.015)		(2.730)	
CBL Arm * 6-month endline				0.008		−2.428
				(0.014)		(2.615)
CBL Arm * 12-month endline				0.020		−1.267
				(0.017)		(3.262)
CBL Arm * 18-month endline				0.028		−1.981
				(0.020)		(3.745)
Observations	1531	1531	5977	5977	4865	4865
Individuals	1531	1531	1507	1507	1238	1238
Mean dependent variable in Extra Step Arm	0.117	0.059	0.873	0.873	567	567
SD dependent variable in Extra Step Arm	0.322	0.236	0.333	0.333	67	67
Mean dependent variable in Extra Step Arm at baseline	0	0	0.840	0.840	561	561
SD dependent variable in Extra Step Arm at baseline	0	0	0.367	0.367	64	64

OLS intention-to-treat estimates with standard errors (clustered on person in columns 2 and 3) in parentheses. Each column presents estimates from a regression of the variable described in the column heading on the variable(s) described in the row headings. Regressions in columns 2 and 3 also include a Post indicator, which takes the value of 1 if the observation is from an endline but not the baseline, and person fixed effects. Unit of observation for columns 2 and 3 is a person-credit report, with four observations for most persons: baseline, and three endlines at 6, 12, and 18 months post-treatment assignment. Number of observations is lower than the number of individuals x 4 credit reports in columns 2 and 3, because a small number of credit reports lack information on one or more dependent variables, including whether the person is scored.

Here, Y is a credit report variable for person i at time t, where t includes the baseline and the three endlines (pulled roughly 6, 12, and 18 months post-random assignment). *CBL Arm*=1 if i was randomly assigned to that arm; the Extra-Step Arm is the omitted category. The CBL interaction with *Post* identifies the average effect of CBL access across the three endlines. Because we have multiple observations per person we include person fixed effects I_i (thereby absorbing the main effect $CBL\,Arm_i$) and cluster standard errors at the person level (the unit of randomization).

The average treatment effect is null on each of the primary outcome variables. Column 2 shows a 1.8-pp estimate of the CBL ITT effect on the likelihood of having a FICO® Score, on a base 87% in the Extra Step Arm. The standard error of 1.5pp implies that the confidence interval includes meaningful but not large effects on the extensive margin of scoring, at least in ITT terms. Column 3 shows a −1.9-point estimate of CBL's effect on the FICO® Score, conditional on having a score at baseline, on a base of 567. The standard error of 2.7 points implies a rather precisely estimated zero in ITT terms. Column 2, panel A, and column 3, panel A, disaggregate the treatment effect by endline and show no strong evidence of differences or dynamics across endlines.

2.2 Heterogeneous treatment effects

2.2.1 Is there treatment effect heterogeneity? An omnibus test using a generalized random forest. The null average treatment effects mask important heterogeneity. To examine heterogeneity, we first chose an extensive set of model "inputs"—potential sources of HTEs—for a machine learning model to search across. In doing so we grouped correlated baseline variables into indexes, to reduce collinearity and preserve degrees of freedom. The legend to Table 4 details the inputs.

We then test for overall (sometimes referred to as "aggregate" or "omnibus") heterogeneity with a generalized random forest model (Wager and Athey 2018; Athey and Wager 2019; Athey et al. 2019). Using the forest prediction on held-out data, these tests compute the best linear fit with two regressors, the target estimand and the mean forest prediction. Table 4, panel A, reports the coefficient and p-value for each of the model's two key test statistics, separately for each outcome-endline combination. The Mean Forest Prediction tests whether the model predicts the outcome accurately. A substantial deviation from one is cause for concern, but we find no such evidence across any of the outcome-endline combinations: the p-values on the test of the null hypothesis of accurate prediction range from 0.973 to 0.994. The Differential Forest Prediction tests the null of no treatment effect heterogeneity, which we clearly reject for the continuous score outcome at 6 months ($p=0.002$). We find suggestive evidence of credit score HTEs at 12 months ($p=0.10$), but none at 18 months ($p=0.62$).[8] For the binary outcome of having a credit score, we only find suggestive evidence of HTEs at 6 months ($p=0.09$). The lack of HTEs on this outcome is likely a by-product of there being less variation to predict. Most people already have a credit score at baseline and then keep it over time (Table 2).

Figure 3 plots the generalized random forest's predicted conditional average treatment effect (CATE) for each outcome-endline combination for each consumer (these are also known as individualized treatment effects). The y-axis shows the estimated treatment effect magnitude, and the x-axis orders observations by that magnitude such that the curve is weakly increasing from left to right. Focusing on the continuous score, the range of CATEs illustrates considerable heterogeneity at 6 and 12 months; for example, the 27 or so point difference between the lowest and highest TEs is economically large, as we will discuss below in Section 2.2.3. Another key inference is that these person-specific CATEs fall fairly neatly into three bins: we see about one-third of the sample with a substantial negative TE, about one-third with close to zero, and about one-third with a substantial positive TE. As such we split the sample into

[8] Another manifestation of the lack of inferred HTEs on the 18-month credit score is that the CATE estimates are not monotonically increasing across CATE terciles in Table 4, panel B, column 6.

Table 4
Causal forest aggregate tests for CBL treatment effect heterogeneity

		(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:		1 = has FICO®Score 8			FICO®Score 8		
	Endline:	6 mo	12 mo	18 mo	6 mo	12 mo	18 mo
A. Aggregate test for treatment effect heterogeneity							
Mean forest prediction	Coefficient β_1:	0.967	0.926	0.983	1.022	1.048	1.031
	SE :	(1.863)	(5.992)	(1.109)	(0.639)	(3.353)	(4.164)
	p-value ($\beta_1 = 1$) :	0.986	0.990	0.988	0.973	0.989	0.994
Differential forest prediction	Coefficient β_2 :	0.943	−1.057	−2.290	1.388	0.665	−0.263
	SE :	(0.690)	(1.027)	(0.959)	(0.492)	(0.523)	(0.876)
	p-value ($\beta_2 \leq 0$) :	0.086	0.848	0.991	0.002	0.102	0.618
B. Average treatment effect by terciles of conditional average treatment effect							
Bottom tercile of CATE		−0.03	0.00	0.01	−9.98	−6.14	7.02
		(0.03)	(0.03)	(0.01)	(4.00)	(5.27)	(6.14)
Top tercile of CATE		0.01	−0.01	0.02	6.83	3.19	−5.50
		(0.02)	(0.03)	(0.04)	(3.93)	(4.36)	(5.07)
Difference of top tercile - bottom tercile		0.03	−0.01	0.01	16.80	9.33	−12.52
95% confidence interval range (+/−)		0.07	0.08	0.08	10.98	13.40	15.61
Number of observations		1413	1374	1330	1164	1126	1073

Unit of observation is a person-endline. For each column in this table– each outcome-endline combination– we ran a casual forest using the GRF package in R (Athey et al. 2019; R version 1.0.1, grf version 0.10.4) to predict the outcome listed in the column heading and obtain the CBL's conditional average treatment effects (CATE) on it. The p-values in panel A show the probability that model is well-calibrated ($\beta 1 = 1$) and identifies homogeneous CATEs across observations ($\beta 2 \leq 0$). Panel B uses the predicted CATE for each observation to divide observations into CATE terciles (see Figure 3 and its discussion in the text for why terciles are warranted) and then estimates the OLS ITT separately for each tercile. The right-hand-side variables included in the causal forest for the binary outcome "1 = Has FICO®Score 8" are: age; number of adults in the household; number of children in the household; standardized risk taking score; number of open trade lines; savings balance and combined savings and checking balance (both in hundreds of dollars, winsorized at 95th percentile); dummies equal to one if baseline survey is missing, credit report is missing, the participant is female, the participant's race is Black, the participant is married, the participant has attended college, the participant's household income is less than $30k, the participant is still an SLCCU member, and the participant has a non-CBL loan; and standardized indexes of insecurity, self-control, attention to credit status, credit process knowledge, delinquency, new credit, and lack of liquidity. The right-hand-side variables included in the causal forest for the continuous outcome of FICO®Score 8 are those listed above, with the addition of baseline FICO®Score and a standardized index of the amount that the respondent owes based on account balances. Sample sizes are lower here (than e.g., the number of individuals with data for each outcome in Table 3) because we are doing each outcome-endline combination separately, and because of missing values on input variables.

the top and bottom CATE terciles in Table 4, panel B, and find further evidence of economically meaningful heterogeneity for the 6-month credit score, with an estimated difference in treatment effects between the top and bottom CATE terciles of 16.80 points (±10.98).[9]

Table 4 suggests that there are HTEs but reveals nothing about *who* benefits the most or least from CBLs. We consider the "who" question next.

2.2.2 Treatment effect heterogeneity for whom? Understanding who benefits most or least from CBLs could deliver important insights regarding

[9] Here, the procedure is simply to estimate the OLS ITT, separately by each CATE tercile. The apparent monotonicity violation in column 6, where the point estimate for the top CATE tercile is lower than that for the bottom, is due to imprecision and the CATEs not identifying true heterogeneity for 18-month credit scores (per the differential forest prediction's clear failure to reject homogeneity in panel A).

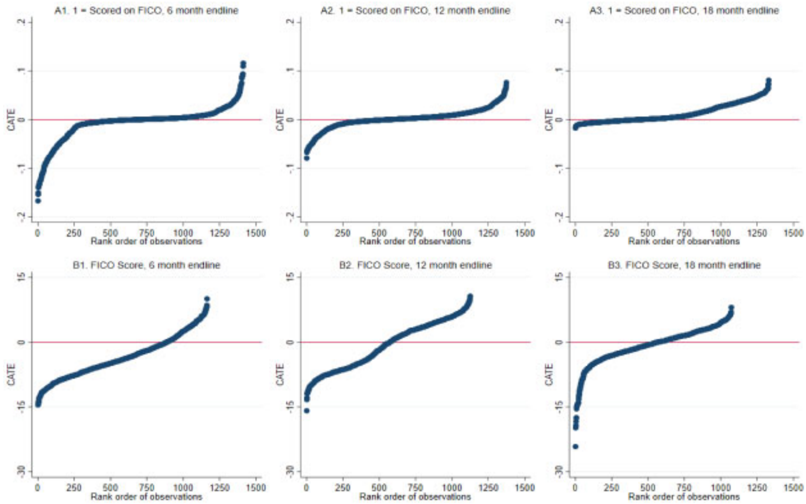

Figure 3
CATE plots for each outcome at each endline
Predicted CATEs from the generalized random forests estimated in Table 4. Panels A1-A3 are estimated on the full sample. Panels B1-B3 are estimated on the subsample of individuals who had a FICO®Score at both baseline and endline.

consumer decision-making, product development, and/or policy design. Table 5 explores which, if any, observable consumer characteristics moderate CBL treatment effects, albeit with three concessions for the sake of brevity and focus. First, we focus on the continuous credit score outcome instead of the extensive margin outcome because the omnibus test in Table 4 finds more evidence of HTEs on the former. Second, we consider only the 6- and 12-month endlines because the omnibus test does not find evidence of heterogeneity at 18 months. Third, we present results for a subset of potential moderators—that is, of model inputs to the casual forest—likely to be of greatest interest for theory, practice, and policy.

Table 5 uses four complementary approaches to statistical inference re: treatment effect moderators. The first two approaches use the causal forest results to test for correlations between a potential moderator and a treatment effect. The second two use a LASSO model of treatment effects to test the extent to which a potential moderator is statistically important from a model selection perspective.

Our first approach in Table 5 examines univariate correlates of the causal forest's predicted person-specific CATEs plotted in Figure 3 and discussed above. Columns 1 and 2 report each potential *moderator's* mean for individuals in bottom and top CATE terciles (columns 1 and 2). Column 3 reports the p-value on the difference between those means. Column 4 presents the q-value from a familywise multiple hypotheses correction (Benjamini and Hochberg 1995; Anderson 2008), where we define three families of tests: (1) demographic

Table 5
Potential sources of CBL treatment efffect heterogeneity on FICO®score

	(1) Mean (SE) of row variable for observations in lowest tercile CATE	(2) Mean (SE) of row variable for observations in highest tercile CATE	(3) p-value (1) = (2)	(4) q-value (1) = (2)	(5) ATE (SE) for observations in lowest tercile of row variable	(6) ATE (SE) for observations in highest tercile of row variable	(7) p-value (5) = (6)	(8) q-value (5) = (6)	(9) Order the variables were added to the model by LASSO	(10) LASSO stability selection
A. 6-month endline										
Age	36.74	50.03	0.00	0.00	−12.13	12.68	0.01	0.06	2	100
	(0.65)	(0.65)			(6.27)	(7.57)				
Female	0.67	0.65	0.50	0.69	−6.45	−0.29	0.46	0.62	T-8	86
	(0.02)	(0.02)			(7.05)	(4.66)				
Married	0.26	0.23	0.36	0.62	−2.16	−0.65	0.87	0.87	9	85
	(0.02)	(0.02)			(4.52)	(7.59)				
Number of adults in household	1.56	1.55	0.84	0.84	−0.51	−17.97	0.19	0.51	10	62
	(0.04)	(0.04)			(5.16)	(13.17)				
Number of children in household	1.02	0.67	0.00	0.00	1.45	−10.48	0.22	0.51	7	91
	(0.06)	(0.06)			(5.39)	(7.79)				
Race - Black	0.88	0.88	0.82	0.84	−10.41	−1.04	0.43	0.62	11	64
	(0.02)	(0.02)			(13.99)	(3.95)				
College or more	0.35	0.19	0.00	0.00	−0.54	−5.85	0.53	0.62	T-8	93
	(0.02)	(0.02)			(4.39)	(7.73)				
Financial risk-taking scale (standardized)	0.12	−0.04	0.02	0.04	6.65	5.23	0.92	0.92	12	63
	(0.05)	(0.05)			(6.3)	(11.16)				
Self-control and credit knowledge index (standardized)	0.45	−0.14	0.00	0.00	4.65	−2.53	0.45	0.72	T-5	100
	(0.05)	(0.05)			(6.44)	(6.93)				
Liquidity index (standardized)	0.05	−0.09	0.05	0.05	−3.66	2.87	0.48	0.72	6	99
	(0.05)	(0.05)			(5.72)	(7.52)				
Baseline FICO®Score	577.53	543.52	0.00	0.00	2.98	−1.58	0.51	0.68	3	100
	(3.07)	(3.07)			(4.05)	(6.45)				
Installment credit activity at baseline index (standardized)	0.74	−0.52	0.00	0.00	15.21	−16.58	0.00	0.00	1	100
	(0.04)	(0.04)			(6.52)	(6.09)				
Revolving credit activity at baseline index (standardized)	0.33	−0.07	0.00	0.00	−4.12	2.90	0.43	0.68	T-5	100
	(0.05)	(0.05)			(4.72)	(7.94)				
Number of prior loans, lifetime	10.41	5.15	0.00	0.00	5.29	2.48	0.77	0.77	4	100
	(0.40)	(0.40)			(6.55)	(7.17)				

(*Continued*)

Table 5
(continued)

	(1) Mean (SE) of row variable for observations in lowest tercile CATE	(2) Mean (SE) of row variable for observations in highest tercile CATE	(3) p-value (1) = (2)	(4) q-value (1) = (2)	(5) ATE (SE) for observations in lowest tercile of row variable	(6) ATE (SE) for observations in highest tercile of row variable	(7) p-value (5) = (6)	(8) q-value (5) = (6)	(9) Order the variables were added to the model by LASSO	(10) LASSO stability selection
B. 12-month endline										
Age	47.68	43.37	0.00	0.00	−2.38	9.97	0.21	0.36	4	65
	(0.78)	(0.79)			(6.72)	(7.64)				
Female	0.62	0.70	0.03	0.05	−9.40	5.10	0.10	0.22	5	53
	(0.02)	(0.02)			(7.60)	(4.79)				
Married	0.27	0.24	0.37	0.43	3.29	−6.51	0.29	0.41	–	28
	(0.02)	(0.02)			(4.69)	(8.13)				
Number of adults in household	1.61	1.61	0.95	0.95	5.04	−24.26	0.03	0.22	8	52
	(0.04)	(0.04)			(5.51)	(13.29)				
Number of children in household	0.58	0.91	0.00	0.00	3.41	−14.02	0.08	0.22	11	30
	(0.06)	(0.06)			(5.67)	(7.95)				
Race - Black	0.83	0.87	0.10	0.15	−5.72	1.47	0.56	0.65	13	31
	(0.02)	(0.02)			(14.21)	(4.15)				
College or more	0.35	0.23	0.00	0.00	0.71	−0.10	0.93	0.93	3	77
	(0.02)	(0.02)			(4.65)	(7.90)				
Financial risk-taking scale (standardized)	−0.04	0.07	0.11	0.17	6.48	2.63	0.80	0.94	10	38
	(0.05)	(0.05)			(6.44)	(12.11)				
Self-control and credit knowledge index (standardized)	0.11	0.10	0.86	0.86	0.52	2.47	0.85	0.94	7	57
	(0.05)	(0.05)			(6.85)	(7.32)				
Liquidity index (standardized)	0.38	−0.14	0.00	0.00	1.50	0.71	0.94	0.94	12	30
	(0.05)	(0.05)			(5.79)	(8.06)				
Baseline FICO®Score	632.51	510.73	0.00	0.00	8.40	−7.89	0.04	0.08	1	100
	(2.07)	(2.08)			(4.37)	(7.10)				
Installment credit activity at baseline index (standardized)	0.28	−0.04	0.00	0.00	13.09	−15.45	0.00	0.01	2	82
	(0.05)	(0.05)			(6.71)	(6.51)				
Revolving credit activity at baseline index (standardized)	0.60	−0.15	0.00	0.00	−1.76	5.29	0.45	0.47	6	58
	(0.05)	(0.05)			(5.20)	(8.17)				
Number of prior loans, lifetime	8.85	7.49	0.02	0.02	4.11	11.38	0.47	0.47	9	42
	(0.42)	(0.42)			(6.92)	(7.50)				

For columns 1 and 2, each row shows the results of an OLS regression of the row variable on indicators for each of the CATE terciles defined in Table 4. For columns 5 and 6, each row shows the results of an OLS regression of FICO Score®on treatment, for those observations classified in the row variable's tercile listed in the column header (for binary row variables, observations with the value of 0 (1) constitute the lowest (highest) tercile for defining the regression sample in columns 5 and 6). The q-values in columns 4 and 8 are the minimum false discovery rate at which the null hypothesis would be rejected, using the Benjamini-Hockberg (1995) step-up method as in Anderson (2008) to correct for the multiple hypotheses within each family. We define three families of tests: demographic (variables age through college), qualitative financial behaviors and preferences (variables financial risk-taking through liquidity), and prior credit history (rest of the variables). Column 9 shows the order the variable was added to the model by a tenfold CV LASSO. Column 10 applies the Meinshausen and Buhlmann (2010) test of variable selection stability using the same LASSO model, iteratively drawing 50% subsamples 100 times and summing the number of times each variables are selected. T indicates ties in ranking, meaning they were added to the model together. Dash indicates that the variable was never added to the model. Internet Appendix B-2 contains details on index components and construction. See Internet Appendix Table 5 for results on examining index components as potential sources of treatment effect heterogeneity.

characteristics, (2) financial behaviors and preferences measured using our baseline survey, and (3) prior credit history (drawn from credit bureau data). Columns 3 and 4 permit inference about whether a particular variable correlates with the CATE. But this inference does not reveal whether any correlation is economically important or spans both positive and negative predicted treatment effects. For that, we turn to columns 5–8, where we compare *CATE* estimates across *moderator* (i.e., row variable) terciles.

A moderator—a source of HTEs—should satisfy two criteria: (1) an economically important and statistically significant difference in the input across bottom and top CATE terciles in columns 1–4, and (2) an economically important and statistically significant difference in the CATE across the top and bottom input terciles in columns 5–8. That is, a model input drives the identified variation in person-specific predicted CATEs if and only if it covaries strongly with those CATEs.

The only correlate satisfying each of those criteria, at both endlines, is the installment activity index calculated from baseline credit reports. Thus, we focus on this margin of HTEs in the rest of our analyses.[10] As detailed in Internet Appendix Section B-2, this index is comprised of three components: number of open installment loans, any open installment loan, and the number of new credit inquiries during the previous 12 months. The latter component covers inquiries for revolving as well as installment loans, but we include it in the installment index because it is strongly correlated with the other installment index components and not with the revolving index components.

Table 5, columns 1–4, show large differences in the baseline installment activity index across the top and bottom terciles of predicted treatment effects, with 1.26-SD less activity (p-value $=.00$, q-value $=.00$) for those in the top TE tercile at the 6-month endline, and 0.32-SD less activity at the 12-month endline (p-value $=.00$, q-value $=.00$). Columns 5 and 6 suggest that those with less baseline installment activity have large positive treatment effects at each endline (15 points and 13 points, with a SE of 7 points), while those with more baseline installment activity have negative treatment effects at each endline (-17 points and -15 points, with a SE of 6 and 7 points). The estimated CATE difference of 32 points at the 6-month endline has a p-value and q-value of .00, and the estimated difference of 28 points at the 12-month endline has a p-value of .00 and a q-value of .01 (columns 7 and 8).[11]

As a further investigation into the importance of baseline installment activity for heterogeneity in predicted treatment effects, we show the order of variable

[10] Because this is our key margin of heterogeneity, Internet Appendix Table 3, panels A and B, repeat Table 1's full sample descriptive statistics and balance checks within the top and bottom terciles of baseline installment activity. Internet Appendix Table 4, panel A, shows that we cannot reject equal first stages across baseline installment activity terciles.

[11] Internet Appendix Table 5, panel A and B, presents estimates that mirror Table 5, but for a generalized random forest that uses index component variables in places of indexes as model inputs. These tables and Internet Appendix Table 4, panel B, suggest that the extensive margin of baseline installment borrowing is especially key.

entry in a LASSO model of treatment effects (column 9) and, following Meinshausen and Bühlmann (2010), examine the stability of variable selection by iteratively drawing 50% subsamples 100 times and counting the number of times each potential moderator is selected (column 10). For the 6-month endline, baseline installment credit activity is the first variable added to the model, and it is selected 100% of the time. For the 12-month endline, installment activity is the second variable to added to the model, and it is selected 82 out of 100 times.

Our takeaways from Table 5 are that consumers with less installment activity at baseline fare well with CBLs, and that those with more installment activity fare worse relatively speaking, and poorly absolutely speaking. Furthermore, the causal forest does not find loading on other covariates, such as our proxies for financial literacy or experience, to suggest that baseline installment activity is merely a proxy for other characteristics or behaviors.

2.2.3 Implications of treatment effect heterogeneity for consumers and providers. A practical implication of our results thus far is that CBL providers could secure higher average treatment effects, and more uniformly positive treatment effects, with two simple and complementary strategies. First, target-market to consumers with less installment activity. Second, screen out consumers with more installment activity, or at least discourage them from taking up a CBL. The results summarized in footnote 11 suggest that a simple screening or targeting rule, based only on whether a consumer has outstanding installment debt or not, could be effective.

Our results thus far also suggest economically important treatment effects for consumers who do experience a score change. Recall that our treatment effects estimates are intention-to-treat, and that the CBL take-up differential across study arms is about 18%, which suggest inflating the ITT estimates roughly 5.5-fold to get a sense of treatment-on-the-treated (ToT) effects. This implies, for example, that someone in the lowest tercile of baseline installment credit activity would experience a score increase of roughly 70 to 80 points (per Table 5, column 5). Such an increase is clearly enough to produce substantial benefits in the form of increased access to credit and/or decreased costs. Consider, for example, a 48-month new car loan, which is plausibly a marginal product for many consumers in our sample, and note that the mean baseline credit score for those in the bottom tercile of installment loan activity is about 535, with a standard deviation of about 70 (Internet Appendix Table 3, panel B). Per the myFICO Loan Savings Calculator, moving up one SD—which, coincidentally, is roughly the size of the implied ToT effect—would reduce the APR by roughly 150 basis points from a baseline APR of 16.1% for the 500–589 score range.[12] A slightly larger score increase would further move

[12] See https://www.myfico.com/credit-education/calculators/loan-savings-calculator/ (accessed December 7, 2021) for “National” market and $10,000 principal amount.

the consumer from the 590–619 to 620–659 range, and further decrease the average APR to 10.2%, almost a 600-basis-point decrease from baseline.[13]

These gross returns are akin to massive increases in alpha, and on their own would be extremely valuable to consumers in our sample: the prevalence of liquidity constraints and low incomes implies that the marginal value of a dollar saved on borrowing costs is quite high for consumers. Estimating net returns requires an adjustment for the liquidity cost of making CBL payments and for the risk that the liquidity cost is unexpectedly high (especially if that high cost leads to credit behavior that decreases the credit score). As detailed in Section 1.1, the CBL here is designed to pose only very modest liquidity demands, and hence the requisite cost and risk adjustments would be minimal in classical models of financial decision-making.

2.2.4 HTEs on credit behaviors. The results in Table 5 raise the question of whether differences in treatment effects are due to differences in CBL-induced credit behaviors—specifically, in factors used as inputs to the FICO® scoring model. A leading alternative hypothesis is that those with different baseline installment credit activity respond similarly to the CBL, but that their similar behavior is scored differently by the model. This alternative hypothesis is viable given the limited modeling information that Fair-Isaac publicly reveals: "The importance of these categories may vary from one person to another ...."[14]

Table 6 uses variants of Equation (1) to estimate CBL treatment effects on credit behaviors. Columns 1–5 present estimates for behavior indexes measuring four of the five behavior factors FICO® states it uses in its scoring model: "New Credit," "Payment History" (delinquency), "Amounts Owed" (which includes both "Balances" and a "Utilization" measure), and "Credit Mix." (We lack a direct measure of the fifth factor behind the FICO® Score, "Length of Credit History.")[15] Columns 6 and 7 present additional results, on CBL delinquency, which is not broken out separately in the bureau (because, as discussed above, the delinquency measure in column 2 includes CBLs, because of reporting and data limitations) but is tracked by our partner credit union. For each measure of each factor we present average treatment effects in panel A, but focus on panel B, where we present HTEs by baseline installment activity. We view each credit behavior in this table as an outcome family unto itself and emphasize one key hypothesis test per family: testing the null of equal treatment effects between the bottom and top terciles of baseline installment

[13] Internet Appendix Table 4, panel B, columns 3 and 4, uses our simplest model of HTEs to estimate effects on crossing these key score thresholds and finds a pattern consistent with our results on our main credit score outcomes, albeit with less precise inferences.

[14] https://www.myfico.com/credit-education/whats-in-your-credit-score (accessed February 6, 2021).

[15] For each of New Credit and Delinquency, Table 6 presents a second version of the outcome index that drops the 6-month endline (column 1, panel B, and column 2, panel b), because each of these indexes contains one or more components with a 12-month lookback.

Table 6
CBL treatment effect heterogeneity on credit behaviors

FICO®Score 8 Factor:	(1a) New credit	(1b) New credit	(2a) Delinquency	(2b) Delinquency	(3) Amounts owed	(4)	(5) Credit mix	(6) CBL delinquency	(7)
Dependent variable index includes:	Inquiries, number of accounts	Inquiries, number of accounts	10 measures of delinquency, collections, & derogatories (higher values = less timely repmt). Includes CBL delinquency.	10 measures of delinquency, collections, & derogatories (higher values = less timely repmt). Includes CBL delinquency.	Balances: Revolving, auto loans, other installment	Utilization: 4 discrete measures of credit limit usage and outstanding balances; # open installment loans	1=(Open installment and open revolving loan)	1=Currently Delinquent on CBL	1=Ever Delinquent on CBL
Sample:	Full	Drop 6-month endline	Full	Drop 6-month endline	Full	Full	Full	No baseline	No baseline
Source:	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	SLCCU Admin	SLCCU Admin
A. Main effects									
CBL Arm								0.014	0.086
								(0.004)	(0.014)
CBL Arm * Post	0.004	−0.011	0.081	0.087	−0.058	−0.002	−0.015		
	(0.036)	(0.047)	(0.039)	(0.050)	(0.041)	(0.041)	(0.021)		
Observations	5981	4482	5981	4482	5488	5981	5981	4558	1525
Individuals	1507	1505	1507	1505	1425	1507	1507	1525	1525

(*Coninued*)

Table 6
Continued.

FICO®Score 8 Factor:	(1a) New credit	(1b) New credit	(2a) Delinquency	(2b) Delinquency	(3) Amounts owed	(4) Amounts owed	(5) Credit mix	(6) CBL delinquency	(7) CBL delinquency
Dependent variable index includes:	Inquiries, number of accounts	Inquiries, number of accounts	10 measures of delinquency, collections, & derogatories (higher values = less timely repmt). Includes CBL delinquency.	10 measures of delinquency, collections, & derogatories (higher values = less timely repmt). Includes CBL delinquency.	Balances: Revolving, auto loans, other installment	Utilization: 4 discrete measures of credit limit usage and outstanding balances; # open installment loans	1=(Open installment and open revolving loan)	1=Currently Delinquent on CBL	1=Ever Delinquent on CBL
Sample:	Full	Drop 6-month endline	Full	Drop 6-month endline	Full	Full	Full	No baseline	No baseline
Source:	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	Credit Bureau	SLCCU Admin	SLCCU Admin
B. Heterogeneity by baseline credit access									
(1) CBL Arm * Bottom tercile of installment credit activity at baseline index								0.017	0.100
								(0.006)	(0.022)
(2) CBL Arm * Middle tercile of installment credit activity at baseline index								0.006	0.076
								(0.005)	(0.021)
(3) CBL Arm * Top tercile of installment credit activity at baseline index								0.019	0.091
								(0.008)	(0.022)
(4) CBL Arm * Post * Bottom tercile of installment credit activity at baseline index	0.013	0.017	0.026	0.040	0.021	0.096	0.056		
	(0.040)	(0.051)	(0.060)	(0.076)	(0.065)	(0.069)	(0.027)		
(5) CBL Arm * Post * Middle tercile of installment credit activity at baseline index	−0.046	−0.064	0.001	−0.011	−0.109	−0.041	−0.077		
	(0.047)	(0.061)	(0.062)	(0.078)	(0.079)	(0.070)	(0.039)		
(6) CBL Arm * Post * Top tercile of installment credit activity at baseline index	0.023	−0.014	0.224	0.242	−0.068	−0.082	−0.041		
	(0.084)	(0.108)	(0.081)	(0.101)	(0.063)	(0.074)	(0.040)		
p-value of (1) = (2) or (4) = (5)	0.335	0.306	0.775	0.638	0.205	0.163	0.005	0.120	0.421
p-value of (2) = (3) or (5) = (6)	0.474	0.688	0.028	0.047	0.689	0.687	0.522	0.120	0.622
p-value of (1) = (3) or (4) = (6)	0.919	0.793	0.049	0.110	0.321	0.078	0.042	0.838	0.768
Observations	5970	4476	5970	4476	5482	5970	5970	4471	1496
Individuals	1502	1502	1502	1502	1423	1502	1502	1496	1496
Mean dependent variable in Extra Step Arm, baseline	0.000	0.000	0.000	0.000	0.000	0.000	0.370	NA	NA
Mean dependent variable in Extra Step Arm, Post	−0.095	−0.111	−0.036	−0.018	0.099	0.100	0.443	0.005	0.036

OLS intention-to-treat estimates with standard errors (clustered on person in columns 1-6) in parentheses. Each panel-column presents estimates from a regression of the variable described in the column heading on the variable(s) described in the row headings, with regressions in panel A Columns 1–5 also including person fixed effects and Post, and the regressions in panel B, columns 1–5 also including person fixed effects and Post interacted with each of the baseline installment credit activity terciles. In columns 1–5, unit of observation is a person-credit report, with at most four observations for most persons: baseline, and three endlines at 6, 12, and 18 months post-treatment assignment, which are included in the Post indicator for the experiment period. Number of observations is lower than the number of individuals x 3 or 4 credit reports because some credit reports lack information on one or more dependent variables. Unit of observation in column 6 is a person-SLCCU data snapshot, with those snapshots timed to coincide roughly with the credit report endlines. Columns 6 and 7 use endline data only, because no one in our sample had a CBL at baseline. Those who did not open a CBL are coded as zero in columns 6 and 7. Index variables are standardized to be mean zero and standard deviation one in the Extra Step Arm at baseline; see Internet Appendix B-2 for details on index components and construction.

activity. We present the p-value on that test in row "p-value of (i) = (iii) or (iv) = (vi)," showing p-values on the other tercile treatment effect comparisons for completeness.

We find little evidence of HTEs for new credit activity (column 1, panels A and B), but do find some evidence of heterogeneity on delinquency (column 2, panels A and B), with a large (0.20 SD) difference in TEs on delinquency between the top tercile of baseline installment activity and the other terciles that has a p-value of .049 or .110, depending on how we measure the outcome. Here, higher values indicate more delinquency and default, and so the HTE is driven by a deterioration in performance for the high-installment tercile (e.g., the 0.22-SD increase in poor performance in column 2, panel A, SE 0.08 SD). Columns 6 and 7 suggest the pattern in column 2 is not driven by the CBL itself, for two reasons. First, we do not see HTEs on CBL delinquency; in particular, there is little evidence that those in the highest tercile of baseline installment activity have higher CBL delinquency. Some CBL users may prioritize CBL payments over installment loan repayments because, although they are treated identically by the credit bureaus, the CBL payment presents substantially lesser demands on liquidity (Section 2-A). Second, column 6, which uses three endline snapshots of SLCCU data to measure delinquency and thereby mirrors our credit bureau data structure, shows that the magnitude of any treatment effect on CBL delinquency measured at any single point in time is small. This is because any CBL delinquency only appears "on the books" for less than a month, due to SLCCU's practice of curing any 30-day CBL delinquency with the remaining escrow balance and then immediately closing the CBL account (Section 1.1). Column 7 confirms that measuring CBL delinquency across multiple SLCCU data snapshots produces average TEs on delinquency that are closer to what one would expect given the 18% take-up differential between the CBL and Extra-Step arms.

Turning to the "Amounts Owed" factor, we see that panel B, columns 3 and 4, show suggestive evidence of larger TEs on the bottom tercile than the others. Even if this pattern were statistically stronger, its implication for scoring would less clear than for the other factors. High utilization is scored negatively, but there may be nonmonotonicity; for example, some middle range of utilization may be scored more favorably than none.

Panel B, column 5, suggests large differences in TEs on credit mix between the bottom tercile and the others. For those in the bottom tercile, CBL access increases the likelihood of having both an installment and revolving loan open by 0.056 pp (SE 0.027). The point estimates for the other terciles are substantially different and negative (p-values on the difference from the bottom tercile of .01 and .04). Since having both loan types open is scored positively, this heterogeneity in credit mix could be contributing to the installment activity HTEs on credit scores.

Altogether, these results are consistent with the CBL inducing differential responses in credit mix and delinquency that drive the HTEs by baseline installment activity in Table 5.[16]

2.3 Impacts on usage of other SLCCU products

Table 7 examines CBL treatment effects on the usage of other SLCCU products, using the same specifications we use for Table 6. These results help round out the picture of how consumer financial behavior changes as creditworthiness builds or deteriorates, on whether the CBL helps individuals build savings (SLCCU does not focus on this extensively in its marketing, but other CBL providers do), and on the bottom-line viability of CBLs from the supply-side perspective. Odd-numbered columns estimate average treatment effects for the full sample across the three endlines, and even-numbered columns estimate treatment effects separately by baseline installment credit activity terciles.

Columns 1 and 2 show no evidence of treatment effects on membership retention (e.g., −1 pp with SE 1 pp in column 1), although the confidence intervals do not rule out economically meaningful effects on attrition given that only 7% of the full sample is no longer an SLCCU member by the 18-month endline. Columns 3 and 4 show no treatment effect of the CBL on non-CBL borrowing from SLCCU on average (1 pp, SE 2 pp, control mean 0.32), but with suggestive evidence of heterogeneity: the TE on those in the bottom tercile of baseline installment credit activity is an estimated 4.9 pp (SE 2.7 pp) increase, while the TEs on those in other terciles are imprecisely estimated nulls (−0.1 pp with a SE of 4.0 and 3.1 pp).

Columns 5–8 examine treatment effects on deposit account balances. These are key outcomes for understanding whether there is a flypaper effect of CBL proceeds. Positive treatment effects on balances would be consistent with members using CBL for what it is, mechanically, aside from its credit reporting feature: a costly commitment savings device. We see some evidence that CBL increases the level of savings balances, with the full sample result in column 5 ($248, SE $121) perhaps being driven by those in the upper terciles of installment credit activity at baseline in column 6. But Internet Appendix Table 7, columns 1–3, shows this pattern is not entirely robust to alternative functional forms of savings balances. We add checking account balances together with savings in Table 7, columns 7 and 8, and Internet Appendix Table 7, columns 4–6, finding imprecisely estimated null TEs on balances in these specifications. Overall, our estimates are too imprecise to yield sharp inferences on CBL effects on deposit account balances.

Summarizing Table 7, we find little evidence that the CBL backfires from the provider's perspective, and some statistically weak hints of benefits.

[16] Internet Appendix Table 6 shows similar results when we limit the sample to those with a credit score at baseline.

Table 7
CBL treatment effects on usage of other SLCCU products

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	1 = Remain an SLCCU member		1 = Any non-CBL loan with SLCCU outstanding		Balances of all savings accounts ($ hundreds)		Balances of all savings + checking accounts ($ hundreds)	
Sample:	Full							
A. Main effects								
CBL Arm * Post	−0.008		0.009		2.476		1.297	
	(0.011)		(0.019)		(1.214)		(1.669)	
B. Heterogeneity by baseline credit access								
(1) CBL Arm * Post * Bottom tercile of installment credit activity at baseline index		−0.010		0.049		0.149		0.077
		(0.021)		(0.027)		(1.355)		(1.590)
(2) CBL Arm * Post * Middle tercile of installment credit activity at baseline index		−0.012		−0.011		5.519		5.456
		(0.021)		(0.040)		(3.064)		(3.164)
(3) CBL Arm * Post * Top tercile of installment credit activity at baseline index		−0.001		−0.011		2.122		−1.479
		(0.015)		(0.031)		(1.727)		(3.890)
p-value of (1) = (2)		0.730		0.148		0.369		0.711
p-value of (2) = (3)		0.685		0.997		0.334		0.167
p-value of (1) = (3)		0.960		0.217		0.109		0.129
Observations	6124	6008	6124	6008	6124	6008	6124	6008
Individuals	1531	1502	1531	1502	1531	1502	1531	1502
Mean dependent variable in Extra Step Arm at baseline	1.000	1.000	0.322	0.327	4.987	5.040	7.435	7.536

Unit of observation is a person-SLCCU data snapshot, with four observations for most persons at roughly the same timing as our credit report pulls: baseline, and three endlines at 6, 12, and 18 months post-treatment assignment, all three of which are included in the Post indicator for the experiment period. Standard errors, in parentheses, are clustered at the person-level. Each column presents results from a single OLS regression of the dependent variable described in the column heading on the variable(s) shown in the rows, Post and person fixed effects (odd columns), with even columns including Post * Bottom tercile of credit access at baseline index, Post * Middle tercile of credit access at baseline index, and Post * Top tercile of credit access at baseline index instead of the *Post* indicator. All outcome variables here are calculated from SLCCU administrative data. Balances are recorded as zero for those who leave the credit union.

2.4 Effects on market information and tests for differential self-selection

Next, we investigate how the CBL affects the quality of information available to the market.

We have already provided some evidence on this question, with our estimates of treatment effects on the likelihood that a consumer is scored. As discussed in the Introduction, this is an important margin for lenders and the market in the sense that a scoring company only reports a consumer's score when it has sufficient confidence in its predictive power.

Our additional tests here focus on the question of whether CBL take-up reveals information. We also consider the related issue of whether our experimental design induces differential selection across study arms.[17]

2.4.1 Does CBL take-up reveal information? Evidence regarding self-selection on unobservables.. Our first analysis, in Table 8, tests for self-selection: does CBL take-up help predict someone's future credit score? The idea here is that a consumer's CBL take-up decision may reveal something about their credit risk trajectory that otherwise would be unobserved to lenders. We implement selection tests that predict each of our two main credit score outcomes by replacing the random assignment indicator in Equation (1) with an indicator for whether someone took up a CBL. Normally this "naïve" specification would capture an unidentifiable combination of treatment and selection effects, but given a null for average treatment effects (Table 3) the naïve specification identifies selection in the full sample. Even-numbered columns control for baseline levels and trends that can vary with the baseline score level (we use Post Double Selection LASSO to select which *Post*Baseline score bin* terms to include)—controlling for these more sharply focuses on selection on consumer attributes that are unobserved by lenders.

Table 8 shows strong evidence of positive (i.e., advantageous) selection on CBL take-up, for both outcomes and specifications. For example, column 2 shows that CBL takers are 11 pp (SE 1 pp) more likely to have a credit score in the endline period than nontakers, and column 4 shows that CBL takers who enter the sample with a credit score have scores that are 11 points (SE 3 points) higher during the endline period.

In all, Table 8 implies that CBLs attract consumers who are on an upward trajectory that is not fully captured by baseline observables. This has market implications: lenders can use CBLs to identify consumers whose creditworthiness is about to start improving. We speculate that credit bureaus could facilitate even stronger advantageous self-selection by distinguishing CBLs from standard installment loans in their data.

---

[17] Internet Appendix C considers another key question regarding effects on market information: whether CBLs improve or weaken the predictive power of credit scores. This is an important question, especially since CBLs are reported to the credit bureaus as installment loans rather than as their own category. Data and institutional constraints limit our ability to draw firm inferences, but the approach we take in the Internet Appendix should prove useful for other studies.

Table 8
Selection into CBL

	(1)	(2)	(3)	(4)
Dependent variable:	1 = Has FICO®Score 8		FICO®Score 8	
Sample:	Full		Participants who have score at baseline	
Took up CBL * Post	0.112	0.105	12.261	10.739
	(0.022)	(0.014)	(3.508)	(3.157)
Controls for baseline variables * Post	No	Yes	No	Yes
Number of people in sample that took up a CBL	318	318	257	257
Observations	5977	5977	4865	4865
Individuals	1507	1507	1238	1238
Mean dependent variable at baseline	0.824	0.824	563	563

Unit of observation is a person-credit report, with four observations for most persons: baseline, and three endlines at 6, 12, and 18 months post-treatment assignment, all three of which are included in Post indicator for the experiment period. Standard errors, in parentheses, are clustered at the person-level. Each column presents results from a single OLS regression of the dependent variable described in the column heading on the row variables described in the table, Post, and person fixed effects. Even-numbered columns include Post interactions with baseline credit score variables selected by Post Double Selection LASSO: baseline FICO®Score 8, 1 = baseline FICO®Score 8 in the 400s, 1 = baseline FICO®Score 8 in the 500s, 1 = baseline FICO®Score 8 in the 600s, and indicator variables for missing values. These bin indicators are all zero for consumers without a score at baseline. Heterogeneous treatment effects by baseline installment activity (Table 5) imply that we cannot identify a pure selection effect separately for those subgroups, and so we only estimate average selection effects here.

2.4.2 Differential self-selection across study arms?. Next, we consider four sets of tests for differential selection across arms. Differential selection would affect the interpretation of our treatment effects and hence external validity. For example, if consumers in the Extra Step Arm expect to incur relatively high costs of take-up—in the form of the added time cost of completing financial education, or the added hassle cost of requesting a waiver—perhaps they only take-up if they will reap relatively large benefits. In that case our design would underestimate CBL benefits for consumers relative to a design that featured only encouragement and no discouragement.

Our first test, in Table 9, panel A, takes our preferred specification from Table 8 and examines whether the conditional correlation between future scores and CBL take-up differs across our two study arms, for our two key measures of future scores. Columns 1 and 3 shows that the estimated correlation between take-up and having a score at endlines is 0.104 (SE 0.016) in the CBL Arm and 0.116 (SE 0.021) in the Extra Step Arm. Column 5 shows that the p-value for these two correlations being equal is .68. Columns 6, 8, and 10 repeat this test for endline credit scores, showing estimated correlations of 12.9 (SE 4.1) for the CBL Arm and 8.1 (SE 4.5) for the Extra Step Arm, with a p-value of .36. These tests show little evidence of differential selection.

Our second set of tests examines the possibility of differential selection by take-up timing. Table 9, panel B, provides some motivating evidence for these tests by suggesting that the Extra-Step requirement pushes some take-up mass from same-day to the next few weeks. Returning to panel A, rows 2–4, repeat our self-selection test separately for three mutually exclusive take-up timing

Table 9
Selection into CBL by arm and timing

A. Selection by take-up timing	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dependent variable:	1 = Has FICO®Score 8				*p*-value (1) = (3) or (2) = (4)	FICO®Score 8				*p*-value (6) = (8) or (7) = (9)
Sample:	CBL Arm		Extra Step Arm			CBL Arm		Extra Step Arm		
(1) Took up CBL * Post	0.104		0.117		0.68	12.949		8.123		0.36
	(0.016)		(0.031)			(4.140)		(4.507)		
(2) Took up CBL at same day * Post		0.116		0.133	0.71		8.161		3.364	0.52
		(0.021)		(0.048)			(5.316)		(6.867)	
(3) Took up CBL between days 2 and 30 * Post		0.068		0.084	0.72		21.789		16.131	0.56
		(0.028)		(0.046)			(7.351)		(8.469)	
(4) Took up CBL > 30 days * Post		0.095		0.119	0.62		19.671		8.166	0.14
		(0.024)		(0.049)			(6.207)		(6.702)	
p-value of row (2) = row (3) = row (4)		0.30		0.74			0.16		0.48	
Controls for baseline variables * Post	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes	
Number of people in sample that took up a CBL	232	232	86	86		191	191	66	66	
Observations	3072	3072	2905	2905		2466	2466	2399	2399	
Individuals	775	775	732	732		625	625	613	613	
Mean dependent variable at baseline	0.810	0.810	0.840	0.840		564	564	561	561	

B. Take-up timing

	All took up	Took up on same day as offer	Took up between days 2-30	Took up > 30 days	
N =	320		169	49	102
Take up in CBL Arm	30%	18%	4%	7%	
Take up in Extra Step Arm	12%	4%	2%	6%	

For panel A, unit of observation is a person-credit report, with four observations for most persons: baseline, and three endlines at 6, 12, and 18 months post-treatment assignment, all three of which are included in Post indicator for the experiment period. Standard errors, in parentheses, are clustered at the person-level. Each column presents OLS estimates from a regression of the variable described in the column heading for columns 1–4 or 5–8 on Post, person fixed effects, and the variables described in the row labels. The baseline variables*Post row refers to variables selected by Post Double Selection LASSO, which are Post interacted with: baseline FICO®Score 8, 1 = baseline FICO®Score 8 in the 400s,1 = baseline FICO®Score 8 in the 500s, 1 = baseline FICO®Score 8 in the 600s and indicator variables for missing values. These bin indicators are all zero for consumers without a score at baseline. Columns 5 and 10 show the results of the Wald test that coefficient for the CBL arm is equal to the coefficient the Extra Step Arm. Heterogeneous treatment effects by baseline installment activity (Table 5) imply that we cannot identify a pure selection effect separately for those sub-groups, and so we only estimate average selection effects here.

bins: those who take-up on the same day as the offer, within the first 30 days of the offer, and >30 days. Reading down these rows, one sees some hint that those who take-up after the first day are more positively selected on their future credit score (columns 7 and 9), but the "p-value of (ii)=(iii)=(iv)" row shows no rejection of the hypothesis that coefficients are equal across the three take-up timing bins.

The key test here, in terms of implications for identifying treatment effects and selection, is for differential selection within take-up timing across arms. Columns 5 and 10 report the p-values on these tests for each of our credit score outcomes and take-up timing bins, and we find little evidence of differences, subject to power constraints of course: five of the six p-values range from .52 to .72. The other p-value is .14, with a difference in point estimates suggesting that, within the latest takers, those in the CBL arm may be more positively selected.

The previous two tests focus on selection on unobservables, because they condition on baseline observables. For selection on observables, we return to Internet Appendix Table 2 and focus on its test for whether baseline observables predict take-up differently across the two arms (column 7). A conservative interpretation of these tests is that there is no evidence of differential prediction. We run 14 tests in column 7 and obtain only one p-value $< .10$.[18] However, the one exception is noteworthy because it fits with the prior that takers in the Extra Step Arm will be relatively savvy, specifically here in the form of higher educational attainment.

The three sets of tests thus far yield little evidence of differential selection. The two potential exceptions to that pattern point in different directions: we see a hint from the take-up timing analysis that the CBL arm is more positively selected on unobservables (Table 9, column 10, row 4), and a hint from the selection on observables analysis that the Extra Step arm is more positively selected on education (Internet Appendix Table 2, column 7).

None of the tests above directly confronts the question of greatest interest for assessing the external validity of our results: does self-selection, under our design, produce treatment effect estimates that differ from what one would expect to find from a market expansion or contraction? To confront this question, we conduct a fourth set of tests, for whether takers have different predicted treatment effects across the two arms. We do this by averaging the generalized random forest's predicted CATEs across endlines for each person-outcome combination, and then comparing that average across arms. Table 10, columns 1–3, does this for CBL takers only[19] and finds no difference across arms for

[18] This is unsurprising, given the lack of evidence that baseline observables predict take-up in either arm in column 3 or 6.

[19] Internet Appendix Table 8 confirms that we find no differences in CATEs across arms in the full sample, as one expects under random assignment. Note that the CATE levels for 1 = scored outcome are an order of magnitude smaller than our main OLS estimate in Table 3, column 2. Several explanations are possible. One

Table 10
Examining differential selection: Mean CATEs by treatment arm for CBL takers

	(1) Extra Step Arm	(2) CBL Arm	(3) p-value (1) = (2)
Sample:	CBL takers		
	Dependent variable: 1 = has FICO®Score 8		
CATE mean (se)	0.0021	0.0039	0.446
	(0.0019)	(0.0012)	
N	82	222	
	Dependent variable: FICO®Score 8		
CATE mean (se)	−2.258	−1.319	0.091
	(0.429)	(0.289)	
N	63	183	

For each outcome, we take the average of the predicted CATEs across three endlines for each individual.

being scored at endline: the average CATE is 0.002 (SE 0.002) for takers in the Extra Step Arm, and 0.004 (SE 0.001) for takers in the CBL arm. The p-value of .45 that does not reject that these two small average treatment effects are equal. There is some suggestion of a small difference for the credit score, with those in the CBL arm having weakly higher (less negative) CATE, −1.3 points (SE 0.29), than those in the Extra Step Arm (−2.3, SE of 0.43) and a p-value of .09 on the difference.

Overall, we find little evidence of economically important differential selection across these four sets of tests.

2.5 External validity?

Now we consider external validity, and in particular the extent to which our results are indicative of what would happen if the CBL market were to expand or contract in response to business innovation or policy intervention. We consider three key aspects.

One key aspect of external validity is the extent to which our sample is representative of consumers who are close to the margin of participating in the CBL market, since those consumers are most likely to be drawn in by an expansion or pushed out by a contraction. With that in mind, we sampled consumers who were interested in a CBL, but not yet using one, since those consumers should be close to the margin (Section 1.3). One might also consider a broader definition of marginal consumers, namely, those with thin or poor credit histories, and our sample looks similar to the limited available comparable data on that population (Section 1.4).

Another key aspect is the extent to which our experiment mimics how a market change would affect marginal consumers. The lack of evidence for differential selection into CBL take-up across our two study arms, as

is the imprecision of the OLS ITT result: the CATEs are basically zero, and the OLS estimate contains zero in its confidence interval. Another is functional form; for example, estimating the ITT with probit instead of OLS yields a marginal effect point estimate that is still not statistically different than zero, of -0.009 (0.017), but more similar in magnitude to the CATEs.

documented in the previous subsection, provides some reassurance here. Those results are consistent with the interpretation that consumers induced to take a CBL by the intensive marketing in our CBL arm, and/or discouraged from taking a CBL by the nominal financial education requirement in the Extra Step arm, are similarly marginal. It may be that the CBL arm acted as an encouragement, in the form of high-touch marketing, and that the Extra Step arm on balance was something close to business as usual, in the sense that the high-touch marketing and nominal financial education requirement had offsetting effects on CBL demand.

A third important aspect is how our results could change in the long-run and general equilibrium. Anything that changes consumer demand for CBLs—business innovations, policy interventions, consumer learning over time, etc.—could change the signaling value of CBLs. We view this as the biggest open question regarding the external validity of the results in this paper.

Overall, it seems likely to us that our study reveals useful insights about CBLs and the market for credit building products more broadly. But we recognize that external validity is in the eye of the beholder, and emphasize that the best approach to assessing external validity is with more studies.

2.6 Implications for policy and practice

With respect to overall efficiency, our estimates of the CBL's effects on consumers, providers, and the market suggest that CBLs could be efficient, and perhaps Pareto-improving, with some modest design changes. Credit bureaus should consider reporting CBLs as a distinct category rather than as a traditional installment loan (as they do with distinct categories for unsecured vs. secured credit cards). General equilibrium effects will be important to monitor, and could reinforce or counteract the partial equilibrium results in our study; for example, anything that increases consumer demand for CBLs—design changes, or consumer learning over time, etc.—could change their signaling value. Providers should consider remediating or screening out those with preexisting installment debt.

Expanding a bit on implications for providers, we see three potential product/program design implications to explore going forward. First, trying to build consumers' financial knowledge with "product-linked" financial education may be counterproductive. We find that a modest financial education requirement decreases product (CBL) take-up by nearly 20 percentage points, even among our sample of consumers that had expressed interest in credit building generally and the CBL specifically. Second, providers should test various approaches to dealing with the possibility that CBLs backfire for those with preexisting installment debt. Possibilities include screening out existing borrowers; offering or requiring a scaffolded approach that focuses first on timely repayment of existing obligations and then segues into another traditional loan or CBL; offering or requiring help with cash flow management; informing and/or reminding users that they need only part with $54 for a few minutes on the payment due date, as $50 of each payment is available to be returned to the customer upon demand. Third, automation of marketing, screening, and payment functions is likely essential for CBL providers to operate at scale, as the small deal sizes required to meet consumer needs and constraints imply a high ratio of fixed costs to potential revenues. The recent emergence of fintech lenders, including ones that screen based on ability-to-pay analysis of checking account data, is encouraging in this regard, and it will be interesting to see whether credit unions and other providers with strong digital operations follow suit.

3. Conclusion

We use a randomized encouragement design and predictive modeling to examine impacts of a credit-builder loan (CBL) on borrowers, providers, and credit market information. The results are mixed, but promising, subject to the external validity caveats discussed in the previous section. They also highlight several opportunities for research and development on CBLs and household finance more broadly.

The CBL studied here has null average treatment effects on consumer credit scores, but these average effects obscure important heterogeneity on a readily observable margin: baseline installment borrowing. Those with more activity at baseline experience large credit score drops from the CBL, while those with less obtain the intended large credit score increase.

Perhaps most strikingly, our results suggest that the CBL increases overall *non*-CBL delinquency among borrowers with higher levels of baseline installment activity. For supply-side consideration, this implies some negative externalities to other lenders in the form of default spillovers. For consumer-side consideration, together with high delinquency rates on the CBL itself (approximately 40%), this suggests that adding CBL's seemingly modest liquidity requirement is too much for many CBL users to manage.

We also find that CBL takers are substantially more likely to obtain or improve their credit scores over the next 6–18 months on average, conditional on their baseline score, implying that lenders can use CBLs to advantageously select borrowers who are on an upward trajectory. (As such our results also illustrate how merely comparing outcomes before versus after product take-up, a common advertising strategy of CBL providers, is misleading.) But credit bureau reporting of CBLs as standard installment loans jams the positive signal of CBL take-up for potential lenders other than the CBL provider.[20]

Altogether, our results highlight some key questions for future research and policy and product development. For research, we need to better understand how to model the decision-making of very resource-constrained consumers.

[20] In contrast, take-up of a new standard loan is often considered predictive of increased credit risk.

For policy and product development, efforts to help consumers improve their credit market outcomes should consider how to target more effectively and how such efforts affect the information environment and market efficiency.

Testing CBL design changes, together with testing whether our results replicate, offers exciting possibilities for revealing insights into fundamental aspects of consumer decision-making. The differential effects we find on baseline installment debt activity beg for particular scrutiny. Is coming up with a short-term outlay of $54 really so disruptive to customers with preexisting installment loans, and if so . . . why? And why don't consumers with preexisting installment loans anticipate this disruption and simply decline the CBL?

References

Anderson, M. 2008. Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and early training projects. *Journal of the American Statistical Association* 103:1481–95.

Askari, A. 2009. Banks and financial education: integrating practice, products, and partnerships. *Community Investments* 21:23–26.

Athey, S., J. Tibshirani, and S. Wager. 2019. Generalized random forests. *Annals of Statistics* 47:1148–78.

Athey, S., and S. Wager. 2019. Estimating treatment effects with causal forests: An application. *Observational Studies* 5:36–51.

Bartik, A., and S. Nelson. 2021. Deleting a signal: Evidence from pre-employment credit checks. Working Paper, Univesity of Illinois at Urbana-Champaign.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 57:289–300.

Blattner, L., and S. Nelson. 2021. How costly is noise? Data and disparities in the U.S. mortgage market. arXiv, preprint, https://arxiv.org/abs/2105.07554a.

Bos, M., E. Breza, and A. Liberman. 2018. The labor market effects of credit market information. *Review of Financial Studies* 31:2005–37.

Brevoort, K. P., P. Grimm, and M. Kambara. 2015. Data point: Credit invisibles. Consumer Financial Protection Bureau.

Brevoort, K. P., and M. Kambara. 2017. Data point: Becoming credit visibles. Consumer Financial Protection Bureau.

Bronchetti, E. T., J. B. Kessler, E. B. Magenheim, D. Taubinsky, and E. Zwick. 2021. Is attention produced optimally? Working Paper, Swarthmore College.

Brooks, J., K. Wiedrich, L. Sims, and S. Rice. 2015. Excluded from the financial mainstream: How the economic recovery is bypassing millions of Americans. Corporation for Enterprise Development.

Chenven, S. 2014. The power of credit building: Credit building strategies for funders. Asset Funders Network.

De Giorgi, G., A. Drenik, and E. Seira. Forthcoming. The extension of credit with non-exclusive contracts and sequential banking externalities. *American Economic Journal: Economic Policy*. https://doi.org/10.1257/pol.20200220.

de Janvry, A., C. McIntosh, and E. Sadoulet. 2010. The supply and demand side impacts of credit market information. *Journal of Development Economics* 93:173–88.

Di Maggio, M., D. Ratnadiwakara, and D. Carmichael. 2021. Invisible primes: Fintech lending with alternative data. Working Paper, Harvard Business School.

Dobbie, W., P. Goldsmith-Pinkham, N. Mahoney, and J. Song. 2020. Bad credit, no problem? Credit and labor market consequences of bad credit reports. *Journal of Finance* 75:2377–419.

Dobbie, W., and J. Song. 2020. Targeted debt relief and the origins of financial distress: Experimental evidence from distressed credit card borrowers. *American Economic Review* 110:984–1018.

Fuster, A., P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther. 2022. Predictably unequal? The effects of machine learning on credit markets. *Journal of Finance* 77:5–47.

Garmaise, M. J., and G. Natividad. 2017. Consumer default, credit reporting, and borrowing constraints. *Journal of Finance* 72:2331–68.

Gelman, M., S. Kariv, M. D. Shapiro, D. Silverman, and S. Tadelis. 2020. How individuals respond to a liquidity shock: Evidence from the 2013 government shutdown. *Journal of Public Economics* 189:103917.

Heidhues, P., and B. Köszegi. 2010. Exploiting naïvete about self-control in the credit market. *American Economic Review* 100:2279–303.

Hertzberg, A., J. Liberti, and D. Paravisini. 2011. Public information and coordination: Evidence from a credit registry expansion. *Journal of Finance* 66:379–412.

Hurst, E., B. J. Keys, A. Seru, and J. Vavra. 2016. Regional redistribution through the US mortgage market. *American Economic Review* 106:2982–3028.

Kaiser, T., A. Lusardi, L. Menkhoff, and C. Urban. 2021. Financial education affects financial knowledge and downstream behaviors. *Journal of Financial Economics* 145:255–72.

Kaur, S., S. Mullainathan, F. Schilbach, and S. Oh. 2021. Do financial concerns make workers less productive? Working Paper, University of California, Berkeley.

Liberman, A., D. Paravisini, and V. Pathania. 2021. High-cost debt and perceived creditworthiness: Evidence from the U.K. *Journal of Financial Economics* 142:719–36.

Lichand, G., and A. Mani. 2020. Cognitive droughts. Working Paper, Harvard.

Manso, G. 2013. Feedback effects of credit ratings. *Journal of Financial Economics* 109:535–48.

Meinshausen, N., and P. Bühlmann. 2010. Stability selection. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72:417–73.

Mello, S. 2022. Fines and financial wellbeing. Working Paper, Dartmouth College.

Mullainathan, S., and E. Shafir. 2013. *Scarcity: Why having too little means so much*. New York: Times Books, Henry Holt and Company.

Olafsson, A., and M. Pagel. 2018. The liquid hand-to-mouth: Evidence from personal finance management software. *Review of Financial Studies* 31:4398–446.

Ong, Q., W. Theseira, and I. Y. H. Ng. 2019. Reducing debt improves psychological functioning and changes decision-making in the poor. *Proceedings of the National Academy of Sciences* 116:7244–49.

Petersen, M. A., and R. G. Rajan. 1995. The effect of credit market competition on lending relationships. *Quarterly Journal of Economics* 110:407–43.

Reyes, B., E. Lopez, S. Phillips, and K. Schroeder. 2013. Building credit for the underbanked: Social lending as a tool for credit improvement. Report, San Francisco State University.

Sledge, J., S. Gordon, and M. Kinsley. 2011. Making the shift from financial education to financial capability: Evidence from the financial capability innovation fund. Center for Financial Services Innovation.

Wager, S., and S. Athey. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113:1228–42.

Wolff, S. 2016. Providing a fresh start: An analysis of Self-Help Federal Credit Union's Fresh Start product. Self-Help Credit Union.

Wong, F. 2020. Mad as hell: Property taxes and financial distress. Working Paper, NBER.