# Delegating Dirty Work: Blame Shifting and the Boundary of the Firm*

Dong Huang†

November 1, 2026

Click here for the latest version

**Abstract**

This paper shows that delegating unpopular actions to an external agent can protect firms' customer relationships from the adverse consequences of such actions. This blame-shifting benefit of outsourcing is one that firms should consider when setting their organizational boundaries. I study an online consumer lender in China that randomly assigns delinquent borrowers to in-house or third-party debt collectors, who differ in collection harshness as measured by their propensity to call borrowers' social contacts. Harsher collection raises short-run repayment under both arrangements, but only in-house harshness deters borrowers—particularly those of higher credit quality—from returning to the lender after repayment. Call transcripts show that harsher collection comes with more threatening communication strategies in both arrangements, yet it provokes borrower anger only when the collector is in-house.

*I am deeply grateful to James J. Choi (Chair), William N. Goetzmann, Paul Goldsmith-Pinkham, and Nicholas C. Barberis for their invaluable guidance and continuous support. I also thank seminar participants at the Yale School of Management for their helpful comments and suggestions. All errors are my own.

†Yale University School of Management. Email: dong.huang@yale.edu.

Why do firms outsource customer-facing activities even when external agents have weaker incentives to protect customer relationships? The classic economic theory of outsourcing weighs its benefits, such as lower costs and specialized expertise, against agency costs, such as misaligned incentives and reduced control.[1] By this logic, customer-facing activities would seem especially likely to stay in-house, because external agents may not fully internalize the damage their conduct inflicts on the firm's customer relationships, particularly when their actions are directly visible to customers but difficult for the firm to monitor or control. Yet firms often outsource such activities.

This paper documents a countervailing benefit of delegation: blame shifting. When a firm delegates to an external agent, customers attribute less responsibility for the agent's conduct to the firm itself. As a result, a firm can take unpopular but profitable actions through an external agent while limiting the damage those actions would otherwise do to its customer relationships. Blame shifting is therefore a benefit of outsourcing that firms should consider when setting their organizational boundaries.

I investigate this blame-shifting mechanism in field data by examining how borrowers respond to harsh debt collection performed within different organizational boundaries. The analysis uses collection records and call transcripts from a leading online consumer lender in China that randomly assigns delinquent debts between in-house and third-party collectors and, within each group, across individual collectors who differ in collection harshness. I find that harsh collection activities accelerate short-run debt recovery under both arrangements, but only in-house collectors' harshness reduces borrowers' propensity to borrow again from the lender. In other words, the benefit of harshness transfers across the firm's boundary, while the cost to customer relationships does not.

Blame shifting is common in firm operations. Ex ante, a firm may find outsourcing unpopular actions attractive because it can distance itself from its agents' conduct. A troubled firm, for example, may appoint an interim chief restructuring officer (CRO) to make the hard decisions during insolvency.[2] Ex post, firms may likewise distance themselves from adverse outcomes that have occurred by attributing them to external providers. British Airways, for instance, blamed its third-party ground-handling agent for the theft of a customer's jewelry and asserted that the airline itself was not at fault.[3] Despite its preva-

[1] See Lafontaine and Slade (2007) for a survey of the evidence on firms' make-or-buy decisions.

[2] EPUNTO Interim Management Company advertises that its CRO can "take unpopular decisions [...] the founding team often do not dare to implement, acting as a reputational firewall" (https://epunto.es/en/empresas/servicios-interim-management/roles/cro/, last accessed September 1, 2026).

[3] See https://www.paddleyourownkanoo.com/2025/01/19/british-airways-is-suing-its-airport-ground-handler-in-chicago-for-1-2-after-workers-stole-high-value-jewelry-from-cargo-hold/. Similarly, Amazon has blamed its trucking contractors for their higher safety-violation rates

lence and intuitive appeal, however, blame shifting has received little empirical attention as a determinant of firm boundaries.

Debt collection provides a useful empirical setting for testing the blame-shifting mechanism because lenders routinely perform the same unpopular customer-facing activities both in-house and through third-party agencies. In the Consumer Financial Protection Bureau (CFPB) 2025 survey of major U.S. credit-card issuers, for example, all surveyed issuers conducted at least some collection activities in-house, while more than half also employed third-party agencies (CFPB, 2025a). Interactions with collectors are often unpleasant: in 2025, debt collection was the second most common product or service that consumers complain about (CFPB, 2026).[4] The setting is also economically important. Nearly one in five U.S. consumers with a credit report had at least one debt in collections as of the first quarter of 2023 (CFPB, 2025b).

This paper focuses on delinquent consumer loans before charge-off, which the originating lender typically still retains ownership of. Third-party agencies collect them on the lender's behalf and receive in return a commission proportional to the amount recovered. This arrangement, often known as "contingency-based collection," is common for collecting pre-charge-off delinquent debts in the U.S. (CFPB, 2023).

The Chinese lender I study follows a similar arrangement. It collects most debts in-house through day 25 of delinquency, after which about 62% of the remaining unpaid debts are outsourced to third-party collection agencies. For the majority of debts, this day-26 allocation is random because the company uses in-house performance as an internal benchmark to evaluate third-party agencies and to set and update their incentive schemes over time. The randomization creates comparable groups of debts handled by in-house and third-party collectors. This allows me to examine the impact of harsh collection under these two organizational arrangements. Because regulation requires third-party collectors to identify themselves as such at the start of each call, borrowers know whether an in-house or third-party collector is calling.

A particularly harsh collection practice is calling borrowers' social contacts. Such calls account for more than half of all calls placed in the first five days after assignment. These calls can improve recovery by imposing social pressure on borrowers, effectively leveraging borrowers' social capital as collateral.[5] At the same time, they can irritate

(https://www.cbsnews.com/news/amazon-trucking-contractors-have-higher-rates-of-safety-violations). URLs were last accessed on September 1, 2026.

[4]The CFPB received approximately 387,400 complaints about debt collection in 2025. They were 3.4 and 12.7 times as frequent as credit-card and mortgage complaints, respectively.

[5]See, for example, Besley and Coate (1995); Karlan et al. (2009); Perez-Truglia and Troiano (2018); Breza and Chandrasekhar (2019).

borrowers and damage customer relationships by exposing their financial difficulties to their social networks and subjecting them to shame (Chandrasekhar et al., 2018; Liao et al., 2026). This trade-off makes social-contact calling intensity a natural measure of collection harshness, which is effective but unpopular. If delegation shifts blame away from the lender, outsourcing may allow the lender to capture the recovery benefits of harsh collection while insulating its customer relationships from the costs.

The identification challenge is that collectors may be harsher toward borrowers who are harder to collect from, so a naive comparison would confound the effects of harshness with unobserved borrower characteristics. To identify the effects of harshness, I exploit a second layer of randomization. Every day, new day-26 delinquent debts are randomly assigned across all on-duty in-house callers and, separately, across all on-duty third-party callers. Since callers differ systematically in their propensity to call borrowers' social contacts, random assignment to callers generates exogenous variation in the intensity of social-contact calling that borrowers experience. This allows a judge leniency empirical design (Kling, 2006; Dobbie et al., 2021; Arnold et al., 2022), in which a caller's propensity to call the social contacts of *other* assigned borrowers serves as an instrument for the social-contact calling that a given borrower receives. Because the propensity is measured on other borrowers, it captures the caller's persistent style without being contaminated by the given borrower's unobserved characteristics. I estimate the effects using the unbiased jackknife instrumental variable estimator (UJIVE; Kolesár, 2013; Goldsmith-Pinkham et al., 2026).

On the benefit side, harsh collection, as proxied by the fraction of call attempts to a borrower's social contact ($\%social$), raises debt recovery by comparable magnitudes under both in-house and third-party collection. Specifically, a one-standard-deviation increase in harshness, that is, a 20-percentage-point increase in the fraction of social-contact calls, raises the day-30 recovery rate by about 0.7 percentage points under both arrangements, or 10–15% of the baseline recovery rate. The effect is concentrated among borrowers with more emergency contacts, consistent with the effect operating through social pressure. The recovery gains, however, fade at longer horizons. One interpretation is that social-contact calling mainly accelerates repayment rather than raising the amount ultimately repaid.[6]

On the cost side, I find that *only* social-contact calls made by in-house callers damage customer relationships, as measured by borrowers' propensity to return to the lender and apply for new loans after curing their current delinquent balance. In the first month after cure, a one-standard-deviation increase in $\%social$ by in-house callers reduces the propen-

[6]The fading may also reflect collector rotation: unpaid debts collected in-house (by third parties) are reassigned to other in-house (third-party) callers after a median of 11 (18) days, which dilutes the initial caller's influence at longer horizons.

sity to return by 3.1 percentage points from a 7% baseline, whereas the effect of third-party calls is small and statistically insignificant. The recovery gain from harsh collection thus crosses the firm boundary, while the damage to the lender's customer relationships largely stays in-house, consistent with blame shifting.

Moreover, in-house social-contact calls tend to drive away borrowers of higher credit quality: among those who return, loan performance declines with in-house $\%social$. For example, a one-standard-deviation increase in $\%social$ reduces the fulfillment rate of the first installment by 26 percentage points from a 60% baseline. Among returning borrowers handled by third-party collectors, the performance effects have the opposite sign, so third-party harshness tends to drive away riskier borrowers instead. The profitability of the new loans, which also accounts for loan pricing, shows the same pattern across the two arrangements, though less precisely estimated, suggesting that the selection operates on credit risk that the lender's pricing does not capture.

These selection patterns suggest that blame shifting unbundles two opposing effects of harsh collection on the borrower pool. On the one hand, harsh collection raises the cost of default for borrowers of unsecured loans and thereby screens out riskier borrowers, who default more often and therefore expect to face harsh collection more often. This effect does not depend on who makes the calls. On the other hand, safer borrowers have better outside options and are therefore more sensitive to harshness they attribute to the lender. Losing them hurts the lender when its pricing cannot fully reflect borrowers' privately known default risk. In-house harshness triggers both effects, whereas delegation deflects blame from the lender, retaining safer borrowers while preserving the screening, which makes harsh collection more profitable.

I find no supporting evidence for alternative mechanisms that can simultaneously explain the acceleration of recovery and the reduction in reborrowing. First, the reduction in reborrowing is not merely a compositional effect that arises because the marginal borrowers cured by harsh collection are intrinsically less likely to return. The reduction remains negative even when assuming the largest possible compositional effect.[7] Second, according to the lender's soft credit inquiries, an increase in in-house $\%social$ raises the number of applications cured borrowers submit to other lenders. This redirection of credit demand away from the lender is inconsistent with the two leading alternative channels—increased expectations of social reputation costs of default and activated informal family insurance—both of which predict a general decline in credit demand.

[7]To bound the treatment effect in the spirit of Lee (2009), I assume that every marginal borrower cured by more social-contact calls is an intrinsic never-returner, the largest possible compositional effect.

The alternatives' channel-specific predictions fail as well. Under the social reputation cost channel, applications for larger loans, which default more often, should decline more, inconsistent with what the data suggest. Under the informal insurance channel, effects should be stronger for borrowers whose social contacts include more kin, yet they do not vary with kin composition. Most importantly, all these mechanisms operate through the borrower's experience of social-contact calls *per se*, independent of who places the calls, so none of them can explain why the damage to customer relationships arises only in-house.

Since callers with different social-contact calling propensities may also collect differently in other respects, the IV estimates capture the effect of a bundle of collection practices. I characterize this bundle in two steps. First, I relate the callers' observable calling effort to $\%social$. High-$\%social$ callers reach borrowers more often in total, although individual conversations last no longer. They also tend to spread their calling effort more broadly across available numbers. For example, they make fewer repeated attempts per number, dial on fewer days, and, among in-house callers, are less likely to follow up with borrowers who have answered on the same day. These relationships between calling effort and $\%social$ are similar under in-house and third-party collection, so differences in calling effort are unlikely to explain why the relationship damage arises only in-house.

Second, I use a large language model on call transcripts to identify the collector's communication strategies and the borrower's responses and emotions. In the first conversation with the borrower, high-$\%social$ callers, whether in-house or third-party, rely more on threats and warnings of consequences, so harsher language is part of the bundle. For example, a one-standard-deviation increase in $\%social$ raises the probability of a legal threat by 15 percentage points among third-party collectors versus 5 in-house, and high-$\%social$ third-party collectors are rated as exerting more pressure and acting less professionally. Since this association is stronger among third-party collectors, differences in communication strategies alone cannot explain why the relationship damage arises only in-house. The borrower side of the same conversations, by contrast, is consistent with the relationship-damage gap: borrowers facing high-$\%social$ collectors push back more often only when handled in-house. Therefore, third-party calling also appears to attenuate the anger and resistance provoked by the communication harshness bundled with social-contact calling.

Finally, I perform a back-of-the-envelope analysis of social-contact calling's impact on the lender's overall profitability, summing the short-run recovery gains and the potential long-run profit loss from borrowers not returning. I cannot reject that in-house collectors already operate at the profit-maximizing level of $\%social$, while third-party collectors, who are harsher on average, are still on the upward-sloping part of the lender's profitability curve. This pattern suggests that third-party collectors may operate under binding con-

straints on social-contact calling, consistent with the regulatory focus on third-party debt collection.

Blame shifting also has implications for how firms may organize delegation. The gain from delegation should be larger where harshness is more valuable, consistent with the industry practice of outsourcing later-stage, harder-to-collect debts. Because less relationship cost follows delegated conduct back to the firm, organizational separation is compatible with higher-powered incentives for the agent, as third-party collectors' compensation is often proportional to the amount collected. The firm may also prefer to stay at arm's length from the agent and forgo control that would reduce conventional agency losses, if control makes the agent's conduct more attributable to the firm, as experimental evidence suggests (Hill, 2015).

The empirical findings also provide a rationale for regulating outsourced collection specifically, as in the U.S. Fair Debt Collection Practices Act of 1977 (FDCPA), which covers outside collectors but generally exempts creditors collecting their own debts.[8] When debts are collected in-house, the lender partly internalizes the externality of social-contact calling (e.g., shocks to borrowers' social networks and mental distress) through the damage to its customer relationships. Outsourcing weakens this internalization, so in the absence of regulation, third-party collectors would escalate social-contact calling beyond the social optimum.[9] Rules that make lenders accountable for their agencies' conduct, as in China, work through a different channel: they restore part of the relationship cost that delegation removes and counteract the lender's incentive to stay at arm's length.

This paper contributes to several strands of literature. First, this paper provides field evidence on the blame-shifting benefits of delegation, complementing theoretical and experimental work. Theoretically, delegation can make observed actions a noisier signal of the principal's intentions, allowing her to avoid reputational blame even under rational expectations, provided she cannot contractually align the agent's incentives perfectly with her own (Pei, 2015). Lab experiments find stronger results: principals who delegate an unfair decision are punished less than those who make it themselves (Bartling and Fischbacher, 2012), even when the agent merely passes the principal's choice through (Coffman, 2011; Oexl and Grossman, 2013).

[8]The FDCPA explicitly excludes "any officer or employee of a creditor while, in the name of the creditor, collecting debts for such creditor" from "debt collectors" covered by this act (15 U.S.C. § 1692a(6)).

[9]This echoes a frequently cited rationale for the FDCPA. For example, Senator Donald Riegle, chairman of the Senate Banking Committee's Subcommittee on Consumer Affairs, argued that "independent debt collectors, unlike creditors, are not restrained from excesses by the desire to protect their good will" (123 Cong. Rec. 6336 (1977), statement of Sen. Riegle; https://www.govinfo.gov/content/pkg/GPO-CRECB-1977-pt5/pdf/GPO-CRECB-1977-pt5-7.pdf). See also Hunt (2007) and Wilf-Townsend (2022).

The empirical findings also speak to the broader literature on the boundary of the firm. In agency theories of the make-or-buy decision, an outside agent who is rewarded based on measurable output, and who does not bear the damage its conduct does to the principal, neglects the outputs that are valuable to the principal but hard to measure (Holmström and Milgrom, 1991, 1994; Hart et al., 1997). Applied to collection, an outside collector who does not internalize the value of the lender's customer relationships is expected to collect too harshly. The findings in this paper qualify this view. When blame does not travel back to the principal, the harshness that misalignment produces can be valuable to the principal, who may therefore deliberately place the "dirty work" outside the firm's boundary rather than keeping it inside. In this sense, blame shifting can turn a canonical agency cost of outsourcing into part of its benefit.

In addition, this paper contributes to the economics of debt collection. Existing work mainly examines how collection effectiveness and regulation affect recovery and the supply of unsecured credit.[10] I instead study the demand side, investigating how the collection conduct that borrowers experience affects their subsequent credit demand. Closest to this paper's organizational question, Fedaseyeu and Hunt (2018) explain the outsourcing of collection by competition among lenders.[11] In their framework, borrowers hold lenders fully accountable for their agents' practices, and the authors mention blame shifting only as a possible departure from rationality. Liao et al. (2026) also document borrower backlash to social-contact calling in the form of retaliatory default. In the setting I study, backlash instead operates through future borrowing relationships.[12]

Finally, this paper relates to the literature on social networks in consumer lending. Social ties can serve as collateral that sustains repayment when formal enforcement is weak (Besley and Coate, 1995; Karlan et al., 2009), and creditors use shaming to induce repayment, with effects ranging from higher compliance to backlash (Perez-Truglia and Troiano, 2018; Liao et al., 2026). At the same time, individuals may forgo help from their social networks: asking for help carries a stigma that deters people from seeking valuable advice (Chandrasekhar et al., 2018), and even cheap family financing carries shadow costs (Lee and Persson, 2016). By calling borrowers' social contacts, collectors forcibly disclose

[10] See, for example, Drozd and Serrano-Padial (2017); Bursztyn et al. (2019); Fedaseyeu (2020); Romeo and Sandler (2021); Fonseca (2023); Zhou (2024); Laudenbach and Siegel (2025); Choi et al. (2025).

[11] Specifically, when competing lenders delegate to the same agency, borrowers can no longer turn to a lender that collects more leniently, so harsh collection becomes sustainable.

[12] The difference is consistent with a change in the cost of retaliation. Their data (2015–2017) predate the extension of China's credit reporting system to online consumer loans, so delaying repayment cost borrowers little; in my sample (2021–2025), intentional delay raises future borrowing costs through worse credit records, making exit the cheaper way to express discontent.

difficulties that borrowers may prefer to keep private. My findings show both the costs and benefits of this forced use of social collateral.

The remainder of the paper proceeds as follows. Section 1 provides institutional background. Section 2 describes the data and sample. Section 3 explains the empirical strategy, and Section 4 reports the estimated effects of social-contact calling and discusses their interpretations. Section 5 examines how social-contact calling relates to other calling efforts and communication strategies. Section 6 discusses the implications of blame shifting for the lender's decision and for regulation. Section 7 concludes.

# 1 Institutional Background

## 1.1 Contingency-based third-party debt collection

The recovery of delinquent debts constitutes an important part of the consumer lending industry, affecting a substantial number of consumers and the ultimate losses borne by lenders. According to the Consumer Financial Protection Bureau (CFPB), in the U.S., the debt collection industry comprises about 2,500 collection agencies, earning $15.1 billion in annual revenue. Nearly one in five consumers with a credit report had at least one debt identified as being in collections as of the first quarter of 2023 (CFPB, 2025b).

Debt collection generally involves communicating with delinquent borrowers through phone calls, direct mail, emails, and text messages to induce repayment. These activities may be either performed internally by lenders or delegated to third-party debt collection agencies. The CFPB's survey of major credit-card issuers reports that all surveyed issuers conducted at least some pre-charge-off collection activities in-house, while more than half also employed third-party agencies before charge-off (CFPB, 2025a).

For lenders that employ third-party collection agencies, two economically distinct arrangements are particularly common. The first is contingency-based collection. Under this arrangement, the originating lender retains ownership of the loans but places the account with an outside collection agency, which attempts to recover the debt on behalf of the lender. Any recovered repayments go directly to the lender, and the agency is compensated with a percentage of the amount recovered. In the CFPB's 2023 survey of major credit-card issuers, agencies received an average contingency fee of 14% for pre-charge-off collections and 21% to as high as 47% for post-charge-off accounts, increasing in recovery difficulty (CFPB, 2023).

The second major arrangement is debt sale. Rather than retaining ownership and hiring an agency to collect, the creditor sells delinquent or charged-off receivables to a

debt buyer, typically at a substantial discount to their face value. The buyer becomes the owner of the receivables and collects the accounts itself. A debt buyer earns a profit from subsequent recoveries net of the acquisition price and the costs of collection.

Relative to debt sales, contingency-based arrangements allow the creditor to retain the residual upside from successful recovery and generally provide greater control over account treatment, settlement authority, and account assignments. These benefits come at a cost: the creditor continues to bear the risk of uncertain recoveries and must monitor the collection agency's conduct and incentives. In contrast, debt sales transfer much of the risk and operational burdens to the buyer and provide immediate liquidity, but usually at a large discount and with the loss of direct control over subsequent collection activities.

The CFPB's 2025 survey indicates that contingency-based arrangements are more prevalent than debt sales among large U.S. credit-card issuers (CFPB, 2025a). Fewer than half of surveyed issuers sold debt in 2023 and 2024. Over these two years, issuers that sold debt placed 16% of their charged-off balance inventory with third-party collection agencies but sold only 5% to debt buyers, and issuers that did not sell debt placed 10% with agencies.

## 1.2 The lending company and its business

The company I study is a leading online consumer finance service provider in China, the same lender studied in Choi et al. (2025). In 2025, the company originated 205 billion yuan (29.4 billion USD) of loans to approximately 8.2 million active users. The company's main business is originating point-of-sale installment loans to online consumers on its self-operated online shopping platform, as well as through partner online retailers and mobile payment service providers.[13]

The company focuses on young consumers in China aged 23 to 40 with shorter credit histories, aiming to capture their long-term consumption growth through lasting lending relationships. As a result, the company relies heavily on repeat borrowing to cultivate user loyalty and accumulate user data and credit history. Indeed, more than 85% of all active users on the company's platform were repeat users who had made at least one transaction in the same or the previous year.

The company provides both installment loans and revolving credit lines, similar to the setting studied by Aydin (2022). Installment loans mature in six months to three years and require equal monthly installments. A revolving credit line is similar to a credit card

[13]For example, when a consumer wants to pay with a mobile payment app (such as WeChat Pay or Alipay) but has an insufficient balance, the consumer can still pay the bill by taking out a loan from the company within the mobile payment app.

and lets consumers borrow from the company at the point of payment. "Credit card" loans have a default maturity of one month, but the borrower may request equal monthly payments over a longer maturity, subject to the company's credit screening. Borrowers generally use revolving credit lines for small daily payments and installment loans for more expensive purchases. Loan sizes range from only 8 yuan (1 USD, 10th percentile) to around 5,500 yuan (770 USD, 90th percentile).[14] Because the company's customers are typically riskier than the population average, most loans carry an annual interest rate of 24%, the upper limit allowed by Chinese regulators.

For each borrower, the monthly installments are due on the same day of each month until maturity. Borrowers who fail to pay in full by the due date are considered delinquent and enter the debt collection process, which the next section discusses in more detail. To fully resolve their delinquency, borrowers must repay the overdue installment, plus interest and late fees that accrue on the overdue amount during delinquency.[15] Borrowers who remain delinquent for three months are considered to have defaulted and are reported to external credit reporting agencies, limiting their ability to borrow not only from the company but also from other consumer finance companies and banks. Borrowers may also have difficulty making purchases even when no borrowing is involved, since their credit records may be shared with other businesses.[16]

## 1.3 Collection process

### 1.3.1 In-house vs. third-party debt assignment

Borrowers who fail to pay the monthly installment by the due date enter the debt collection process, during which the company calls them starting on day 2 and urges them to repay. The company uses different strategies depending on whether the debt is 2–25 days, 26–59 days, 60–85 days, or 86 or more days past due. One important variation in strategy is the fraction of outsourced delinquent debts, as illustrated in Appendix Figure B1. In the first 25 days, almost all debts are handled in-house. On day 26, 62% of debts are assigned

[14]Extremely small loans are mainly generated by "credit card" purchases since the company views each purchase, no matter how small it may be, as a standalone loan.

[15]The company may forgive part of the interest and late fees for borrowers over 10 days overdue as an incentive to repay. On average, delinquent borrowers can receive a 7.5% reduction in the overdue balance if they repay within three months of delinquency.

[16]For example, defaulted borrowers may not be able to book hotels without deposits. A borrower may be sued by the lender if the borrower intentionally refuses to repay despite having sufficient money. If the court rules in favor of the lender but the borrower still refuses to repay, the borrower will be added to the Supreme People's Court's blacklist of dishonest judgment debtors, an enforcement component of China's social credit system, and prohibited from expensive consumption and investments such as traveling by plane and purchasing real estate and luxury cars.

to third-party agencies for the first time. Once outsourced, the debts typically will not switch back in-house. About a month later, 15% more debts are outsourced on day 60 past due. After day 85, the company charges off the defaulted debts and gradually allocates almost all of the remaining debts to third-party collectors.[17] The following analysis focuses on the first notable outsourcing on day 26.

The allocation of debt on day 26 between in-house and third-party collection (hereafter, the in-house and third-party *sectors*) is random for a majority of debts. In particular, debts with at least 8,000 yuan (1,100 USD) of remaining loan principal (5,000 yuan, or 700 USD, before 2022), which account for 77% of overdue balances at day 26, are randomized between the two sectors. In addition, within each sector, day-26 debts are randomly assigned across all callers on duty every day.

Regulations require third-party callers to disclose their identity as a third party to the borrower during conversations. Typically, they start a conversation with an opening like "Hello, this is a *delegated representative* of <the lender>." In contrast, in-house callers simply say they are from the lender.

### 1.3.2 Collector rotation

To incentivize early effort and diversify the caller styles borrowers experience, a caller can usually keep a newly assigned debt for only a median of 11 (in-house) to 18 (third-party) days before it is randomly reassigned to another caller if it remains open. The actual rotation timing is caller-independent, varies over time, depending on overall caller capacity, company policies, month-ends, and holidays, and is generally known to callers at the time of assignments.[18] Appendix Figure B2 shows the fraction of debts that remain with the initial day-26 caller over the subsequent delinquent days by collector type. More than 83% of in-house-collected borrowers are still with the initial callers by day 30; third-party initial callers retain an even larger fraction of 97%. The retention rate declines over time, and by day 59, nearly all active debts have been reassigned.

Despite the rotation, a significant fraction of collection efforts and outcomes are front-loaded and from the initial collector, as reported in Appendix Table B1. For instance, among in-house-collected debts, the first five days generate 33% of all phone calls (40% of answered calls and 46% of call duration) over days 26–59; among them, 93% is from the initial callers, who alone account for 57% of calls over this period. Third-party collection

[17] The company only keeps a few large loans that are worth further action, such as court settlement, and loans whose borrowers have filed a complaint.

[18] For example, the lending company usually reallocates all open debts on the first business day of a month or the first day after national holidays, so the number of days that the previous caller can hold a debt is truncated.

is less front-loaded but similarly concentrated on the initial caller (21% of calls arrive in the first five days, and 54% come from the initial caller).

## 1.4 Social contact calls

The random assignment of debts between collector types and across collectors provides exogenous variation in collection strategies and styles. One important strategy that is central to collection in the first two months of delinquency is calling social contacts, which can impose social reputation pressure on borrowers. The lending company controls the overall frequency of social calling by adjusting which phone numbers appear on the lists that callers receive, and in what order, but callers also have discretion in who to call. Figure 1 shows the overall shares of calls to borrowers' own phone numbers and to borrowers' social contact numbers as a function of days past due. Callers call only the borrower's own number in the first three days, but start calling their self-registered emergency contacts on day 4 with 20–30% likelihood.[19] Another escalation occurs on day 11 when the fraction of social contact calls jumps to around 40%.

On day 26, collection activities escalate again as social-contact calls surpass calls to borrowers' own numbers. In addition, callers start calling frequent contacts extracted from borrowers' cell phones.[20] Following the escalation, there is a transitory rise in repayment as a fraction of the outstanding overdue balance, as shown in Appendix Figure B4. Later, the shares of the two types of calls remain about equal to each other through day 59.[21]

## 1.5 Borrower returns to the lender

Delinquent borrowers who have settled the current overdue installment and exited the debt collection process ("cured borrowers") are eligible to *apply* for new loans.

Figure 2 presents the distribution of the timing of the *first* return application of each cured borrower, measured as the number of days after the overdue balance closing date. Appendix Figure B5 supplements it by showing the timing of all return applications and their rejection rates by days since cure. There are basically two waves of returning borrowers. The first and largest wave occurs immediately after the cure day. 18% of

[19]Borrowers are asked to report 2 to 4 emergency contacts with the lender at the time of registration and consent to call emergency contacts in the case of delinquency.

[20]The lender's smartphone app used to ask for permission to access borrowers' phone logs and phone books to extract close contacts and analyze social networks. However, as shown in Appendix Figure B3, along with the tightening of personal information regulation and the rise of privacy concerns, such practices have become harder since 2020 and were completely banned in mid-2024.

[21]Around day 32, calls to borrowers' own numbers temporarily recapture the majority, as the lender asks callers to contact borrowers' own numbers more at the one-month-of-delinquency milestone.

returning curers first return within 30 days after cure. These are impatient borrowers who seek new liquidity immediately after repayment, although nearly all are rejected, as shown in Appendix Figure B5.

The second wave of applications arrives at 180 days. This wave is primarily driven by a relaxation of screening standards, as shown by a sharp decline in rejection rates in Appendix Figure B5. The lending company uses an indicator of any delinquency in the past six months as an input to its underwriting model, so delinquent borrowers' access to credit is relaxed once the six months have passed.

# 2 Data and Sample

The data include comprehensive information on the debt collection process at the lending company between October 2021 and December 2025. To ensure that I can observe borrowers' returning applications for at least one year after day 26 and the performance of loans originated to returners, I restrict the sample to borrowers whose debt remains open on day 26 of delinquency between October 2021 and December 2023.

I focus on *first-time* delinquent borrowers, who have not been through the company's debt collection process before, so that their responses are not affected by prior experiences. Consistent with the company's practice, multiple delinquent debts of an individual borrower are merged into one entry during collection. Therefore, each borrower corresponds to only one "bulk" delinquent debt in the data.

I apply additional filters to select the working sample. First, I drop less than 1% of debts with extremely large overdue balances (over 50,000 yuan, or 7,000 USD) or remaining loan principal (over 100,000 yuan, or 14,000 USD), as well as debts without an assigned collector on day 26. These are typically nonstandard contracts originated to specific customers who are treated differently in collection. I also require at least one call placed during days 26–30, so that the social-contact calling share is defined. I do not trim the left tail of the size distribution: debts with remaining principal below the randomization threshold (8,000 yuan, or 1,100 USD; 5,000 yuan, or 700 USD, before 2022) follow a different assignment rule (Section 1.3.1), so they are excluded from the sample by construction.

In addition, to ensure that the analyzed collectors are active callers who exert a reasonable amount of effort, I exclude those who join or leave the company mid-month or are absent for extended periods. To do so, each month I include only collectors who are on duty for at least 19 days, place at least 50 calls on days 26–30 of the debts they handle,

and serve a total workload of at least 1,500 borrower-days (including both first-time and repeat delinquent borrowers).[22]

Finally, as is standard in judge leniency designs, I exclude collector-months with relatively few debts to obtain a less noisy estimate of social-contact calling propensity at the collector-month level. In the main analysis, I require each collector-month to have at least 15 *first-time* delinquent borrowers, which is around the 10th percentile of the distribution. However, as the results section and the appendix show, the findings are quantitatively similar under alternative lower and higher thresholds, and even when this filter is removed.

The final sample includes 126,935 first-time delinquent borrowers (debts), with 66,664 collected in-house and 60,271 collected by the third-party agency. Table 1 reports the summary statistics. Panel A summarizes the loan and borrower characteristics in the data. The average overdue balance is around 2,600 yuan (370 USD), and the average remaining principal is around 19,000 yuan (2,700 USD). Before day 26, borrowers repay less than 2% of the overdue balance on average, which is why they face more aggressive collection from day 26 on. The internal credit score is based on the probability of default estimated by the company and is converted to an integer score according to the decile position among all delinquent borrowers on day 1, with 1 being the highest decile of default probability. Borrowers are 31 years old on average, with 70% male, 10% with a bachelor's degree or above, and 5% currently enrolled in college. The variable distributions are very close between in-house and third-party, confirming the random assignment between the two sectors.

As shown in Panel B, collectors vary in age and in how long they have been collecting for the lender. Junior collectors may have worked for just 3–4 months, while the most senior collectors have tenures of 2–3 years. On average, the working sample contains 138 (in-house) to 177 (third-party) first-time delinquent borrowers per collector. At the collector-month level, each collector handles around 24 first-time delinquent borrowers in both sectors. In-house collectors place about five times as many calls as third-party collectors during days 26–30 of the debts they handle, while both work nearly full months after the days-on-duty filter. In each calendar month, the sample includes more than 2,000 cases and around 100 active collectors in each sector.

[22]Nineteen on-duty days is around the 5th percentile of the distribution and the smallest value with a nontrivial density of collectors, corresponding to a full February net of the Chinese New Year holiday. Fifty calls and 1,500 borrower-days are around the 10th percentile of the respective distributions. The three filters overlap substantially and exclude about 10% of observations in total.

# 3 Empirical Strategy

## 3.1 Treatment and jackknife instrumental variable

### 3.1.1 Social calling as treatment

This paper asks how harsh collection practices affect repayment and the customer relationship, and whether these effects differ between in-house and third-party collectors. The harshness of the collection activities can be multi-dimensional, but as discussed in Section 1.4, calling social contacts is central to the collection process, and its intensity serves as a natural "index" of harshness. I therefore choose the fraction of call attempts that go to borrowers' social contacts (%*social*) as the treatment variable.[23]

I choose days 26 to 30, the first five days of the assignment, as the evaluation window for callers' tendency to call social contacts. This evaluation window gives a reliable measure of caller styles because, as discussed in Section 1.3.2, about 90% of the collection efforts and outcomes in this window can be attributed to the initial caller despite caller rotation. Moreover, since caller rotation is random and more frequent in-house, any variation introduced by subsequent callers is measurement error that biases against significant findings, and the attenuation bias should be larger in-house, if anything. As Table 1 Panel C shows, social-contact calls average 64% of call attempts under third-party collection, compared with 51% in-house.

### 3.1.2 Identification

The OLS estimator of regressing recovery or customer relationship measure on %*social* gives a biased estimate of the treatment effect of calling social contacts due to unobserved confounders.[24] To address the endogeneity problem, I utilize the random assignment of debts across callers on day 26 in a judge leniency empirical design.

The design relies on callers differing in their propensity to call borrowers' social contacts. Because debts are randomly assigned to callers within a calendar month, the assigned caller's propensity that a borrower receives is uncorrelated with the borrower's characteristics, creating exogenous, caller-specific variation in the actual treatment that the borrower

[23]Sections 3.1.4 and 5 discuss in greater detail how the multi-dimensional nature of strategies affects the interpretation of the estimated treatment effects of %*social* and how %*social* is associated with other aspects of collection strategies, respectively.

[24]For example, borrowers in a worse financial position repay less and also draw harsher collection practices intended to pressure them to repay, which biases the recovery estimate downward; callers may instead escalate to social contacts when they expect it to work, biasing the estimate upward. The direction of the net bias is ambiguous ex ante and can differ across outcomes and sectors.

experiences. I therefore instrument a borrower's actual %*social* with the assigned caller's propensity in that month, measured by the average %*social* of the *other* borrowers assigned to the same caller in the same month. Leaving the borrower out of the instrument is essential: an average that included the borrower's own treatment would carry the borrower's confounders into the instrument and bias the estimate toward OLS. This is the jackknife instrumental variable estimator (JIVE), whose first stage uses caller-by-month dummies as instruments but predicts each borrower's treatment without that borrower's own observation.

In particular, I use the unbiased jackknife instrumental variable estimator (UJIVE) proposed by Kolesár (2013) and discussed in more detail in Goldsmith-Pinkham et al. (2026). UJIVE internally defines the instrumental variable in the following way: for borrower $i$ treated by caller $c$ in month $t$, their treatment variable $X_{ict}$ (%*social*) is instrumented by

$$\hat{X}_{ict} = \left(1 - \frac{1}{N_2 + 1}\right)\left[\frac{1}{N_1} \sum_{\substack{j \neq i \\ c'=c,\ t'=t}} X_{jc't'} - \frac{1}{N_2} \sum_{\substack{j \neq i \\ t'=t,\ \text{any } c'}} X_{jc't'}\right], \tag{1}$$

where $N_1$ is the number of other borrowers treated by the same caller $c$ in month $t$ as borrower $i$, and $N_2$ is the number of other borrowers treated by any callers in the same month $t(i)$ as borrower $i$.[25] In other words, the instrument $\hat{X}_{ict}$ captures caller $c$'s propensity to call borrowers' social contacts in month $t$ relative to the month-$t$ average tendency, after leaving borrower $i$ out in both averages.

Given that the debt assignment is effectively random across both types of callers in the same pool, the month-level leave-out average (the second term in the bracket of Equation (1)) is calculated over both types of callers. Robustness checks confirm that the reported results are robust to instruments constructed separately within in-house and third-party callers and to alternative jackknife IV constructions.[26] Indeed, the alternative jackknife estimators differ from UJIVE essentially in whether borrower $i$ is also left out of the month-level average. With around 2,000 borrowers in every month, one borrower can barely move that average, so the alternative constructions produce virtually identical estimates (Appendix A.1.3).

[25] The factor $1 - 1/(N_2 + 1)$, constant within a month and about 0.9995 in the data, is the leverage adjustment of the jackknife derived in Appendix A.1.1.

[26] These robustness checks are discussed in more detail in the later sections that report the corresponding results. Alternative JIVEs include JIVE1 (Angrist et al., 1999), Improved JIVE (IJIVE; Ackerberg and Devereux, 2009), and Shrunken JIVE (SJIVE; Frandsen et al., 2026).

### 3.1.3 Instrument validity

*Balance test.* Appendix Table B2 verifies the orthogonality of the instrument with respect to observable pre-treatment loan and borrower characteristics, including the overdue amount, remaining loan principal, internal credit score deciles, age, gender, education, and repayment before day 26. Specifically, following Goldsmith-Pinkham et al. (2026), I regress each characteristic on the treatment using UJIVE, and insignificant coefficients constitute evidence of instrument exogeneity. All coefficients are small and statistically insignificant, with the exception of a marginal imbalance in the overdue amount among in-house-collected debts, which is economically negligible.[27]

*Average monotonicity.* The average monotonicity condition ensures that the IV estimator is a weighted average of potentially heterogeneous individual treatment effects with non-negative weights and, therefore, possesses a causal interpretation. It requires that, for each borrower, the counterfactual treatment intensity they would have received had they been assigned to any caller be non-negatively correlated with the caller's social-calling tendency. As suggested by Goldsmith-Pinkham et al. (2026), the "non-negative IV weights" implication of average monotonicity can be tested by using $V_i \times X_i$ as the outcome in the UJIVE specification, where $V_i$ is a binary indicator for borrowers in some characteristic group, so that the UJIVE estimate is a weighted average of the 0/1 indicators with the same IV weights as in the treatment effect estimations. Appendix Table B3 reports the results. The UJIVE estimates are well within the $[0, 1]$ interval, suggesting no violation of the assumption.

*First-stage regressions.* Appendix Figure B6 shows the distribution of the UJIVE instrumental variable and its relationship with the actual treatment intensity using a local linear regression, separately for in-house (Panel A) and third-party (Panel B) collectors. The distributions suggest that the IV provides meaningful exogenous variation in callers' tendency to call social contacts. The local linear regressions show a strong, monotonically increasing first-stage correlation, with narrow 95% confidence intervals across the full distribution, indicating a strong first stage. A small first-stage $F$-statistic does not bias UJIVE, which remains approximately unbiased even with many weak instruments (Goldsmith-Pinkham et al., 2026), and screening on the $F$-statistic can itself distort inference (Angrist and Kolesár, 2024; Chyn et al., 2025).

[27] A one-standard-deviation increase in $\%social$ corresponds to an overdue amount lower by 58 yuan (8 USD), or 2.2% of the sample mean.

### 3.1.4 Interpretation of the IV estimates

Collection harshness is multidimensional. Callers who differ in their propensity to call social contacts may also differ in other collection practices, such as the tone of their calls, the use of threats, how quickly they escalate collection, or how persistently they follow up. The IV estimate therefore does not isolate the effect of increasing the fraction of social-contact calls while holding everything else fixed. Instead, it captures the effect of a caller-induced increase in collection harshness, expressed per unit increase in $\%social$, an observable index of a broader collection-harshness style. Other collection practices that systematically co-move with callers' social-calling propensity and constitute dimensions of harshness are part of this treatment bundle, not violations of the exclusion restriction.

Under this interpretation, the strict exclusion restriction requires that assigned callers affect borrower outcomes only through collection harshness, and it is violated if callers also matter through dimensions of behavior unrelated to harshness, such as the information they provide about the debt and how to repay it, or leniency in negotiating repayment terms. Frandsen et al. (2023) show that a weaker "average exclusion" condition suffices: such non-harshness channels leave the estimate consistent as long as their direct effects on the outcome are uncorrelated with callers' $\%social$ tendency, and otherwise bias it upward if the correlation is positive and downward if it is negative.

I address this concern in two ways. First, as shown in Section 4.1, the recovery effect of $\%social$ is concentrated among borrowers with more emergency contacts, so a meaningful fraction of the recovery effect operates through the social-network channel itself rather than through caller traits that would apply to every borrower alike.

Second, Section 5 examines directly how the $\%social$ instrument relates to other observable dimensions of caller behavior, including call frequency, call duration, follow-up, and the communication strategies and styles revealed by call transcripts. Some dimensions vary little with $\%social$. For example, information-related communication strategies show small, similar estimates in both sectors, and the payment reduction actually granted to borrowers is uncorrelated with $\%social$. These channels are therefore unlikely to bias the estimates.

Other dimensions co-move with $\%social$ as part of the bundle. For instance, because high-$\%social$ callers call the borrower's own number less persistently, the estimates capture the net effect of more pressure through social contacts and less pressure through repeated calls to the borrower. Since persistent calling likely raises recovery, the recovery estimate, if anything, understates the effect of social-contact pressure alone; since repeated calls are themselves unpleasant, the same is likely true in the estimated relationship damage.

This trade-off is similar for in-house and third-party collectors, so differences in calling persistence are unlikely to explain why the relationship damage arises only in-house.

## 3.2 Outcome variables

### 3.2.1 Recovery: NPV of money collected

I evaluate the recovery performance of collectors by the net present value of money collected over a fixed horizon, discounted to day 26. The discount rate is set to 24% per annum, which is the interest rate cap that the lending company can legally charge and can be viewed as the opportunity cost of the uncollected money that could have been lent to other borrowers.

To match the evaluation window of the treatment variable, the main specifications also focus on the money collected between days 26 and 30, the first five days of the assignment. As reported in Table 1 Panel C, the day-30 recovery rate is 6.8% in-house and 4.5% under third-party collection.[28]

### 3.2.2 Customer relationship: Reborrowing

The tendency of cured borrowers to return to the lender provides a natural measure of their relationship with the lender. The analysis focuses on the borrowers who resolve the overdue balance before charge-off, i.e., within 60 days after the day-26 assignment. These borrowers account for about 15% of the sample, and around 18% of them return and apply for at least one new loan within six months after resolving the overdue balance (Table 1 Panel C). Guided by the two return waves documented in Section 1.5, I measure cured borrowers' return over two types of cure-anchored windows. The first set of outcomes consists of indicators of any loan application within $[0, 30]$ days of cure and wider windows up to $[0, 180]$, and the second set consists of indicators of entry at the 180-day approval relaxation gate, distinguishing first-time from repeat applications in the $[180, 210]$ window.

Appendix Table B4 supports the customer relationship interpretation of returning: among in-house-collected borrowers (Panels A and B), reborrowing is negatively associated with borrower complaints at every horizon, and increasingly so in the long run. Among third-party-collected borrowers (Panels C and D), the negative association also appears at most horizons, though much more weakly than in-house. Borrower complaints are perhaps

[28] A large part of the *gross* recovery gap is offset by differences in collection costs: the same panel shows that in-house collectors place about five times as many call attempts as third-party collectors and spend about five times as much time talking to borrowers. See Section 4.1 and footnote 29 for details. The recovery rates of the two sectors converge to around 13% by day 59, before the debts move to the next stage of the collection process.

a more direct short-run measure of damage to the customer relationship. However, using them as the outcome does not provide sufficient statistical power given that complaints are rare in the sample. Only 1.6% of borrowers ever filed a complaint, likely because doing so can be more costly than simply walking away from the lender.

# 4 Effects of Social Calling by Organizational Form

This section documents the causal effects of social calling intensity on the recovery of delinquent debts and borrowers' tendency to return after repayment, as well as how these effects differ between in-house and third-party collectors.

## 4.1 Recovery

Figure 3 visualizes the reduced-form relationship between the cumulative collected NPV by day 30 and callers' propensity to call social contacts, measured by the UJIVE instrument constructed in Section 3.1, shown separately for in-house and third-party callers, and overlaid with the distribution of the instrument's value. Both variables are first residualized by sector-by-month fixed effects and then recentered to the sector-specific means of the outcome and of $\%social$, respectively.

Figure 3 shows that the positive relationship between recovery and assigned social calling propensity is monotone, close to linear, and comparable in both sectors. The background histograms indicate that the two propensity distributions share a common support, with the third party's density peaking to the right of the in-house density, consistent with the mean difference we have seen in the summary statistics.

The figure also shows that third-party recovery is consistently lower than in-house recovery, by about 2.3 percentage points over the first five days of the assignment. A back-of-the-envelope calculation suggests that collection costs offset much of this *gross* gap. In-house collectors place five times as many call attempts as third-party collectors (Table 1, Panel C), so each yuan they recover costs the lender considerably more: about 33 cents per yuan collected, against about 12 cents for the third party.[29] Net of these costs, the level gap narrows to about 0.2 percentage points. The upward-sloping relationships

[29] I compute collection costs as follows. Based on in-house collectors' paycheck records, variable labor costs are about 3 cents per yuan collected, and the lender pays about 1 yuan (roughly 8 USD per hour) for every minute a collector spends on the phone. In the 2017 U.S. Economic Census, payroll accounts for 45% of the total expenses of collection agencies (2017 NAICS 561440). Since no comparable figure is available for China, I use the U.S. share, dividing labor costs by 0.45 to obtain total costs, and apply the same unit costs to both sectors.

remain intact (Appendix Figure B7) because collection costs vary little with the social-contact calling propensity. The level gap matters for the lender's eventual make-or-buy decision, but since the focus of this paper is borrowers' responses to harsh collection and their differences across the firm boundary, it is the slopes rather than the levels that matter for the analysis.

The recovery effect of social-contact calling is formally tested in Table 2. Column 2 reports the UJIVE estimates with only calendar-month fixed effects — the necessary control conditional on which debt assignment is random across callers. For in-house callers, a 20-percentage-point (roughly 1 s.d.) increase in the fraction of calls to social contacts raises the day-30 recovery by 0.68 percentage points, or a 10% increase from the baseline average recovery (0.068). The treatment effect of $\%social$ persists among third-party callers with almost identical magnitudes, as indicated by the insignificant $z$-statistic of the sector difference in Panel C, suggesting that the *benefit* of calling social contacts transfers across the organizational boundary, while the *cost* side does not, as we will see in the next section. The treatment effect is robust to the inclusion of additional pre-assignment borrower and loan characteristics, as reported in column 4. Columns 1 and 3 show that the OLS estimates differ substantially from the UJIVE estimates due to selection bias, which is resolved by the judge leniency design. As reported in Appendix Table B5, which re-estimates the specifications in columns 2 and 4 of Table 2, the estimates are also stable across alternative jackknife instrumental variable constructions (Panel A) and across various collector-month caseload filters of the working sample (Panel B).[30]

Because the UJIVE aggregates heterogeneous individual treatment effects, a natural question is which borrowers' effects the estimates in Table 2 represent. Following Goldsmith-Pinkham et al. (2026), replacing the dependent variable with $W_i \times \%social$ for a pre-treatment characteristic $W_i$ in the UJIVE specifications recovers a weighted average of $W_i$ with exactly the same weights with which the UJIVE aggregates individual treatment effects. Appendix Table B6 reports this characterization for the no-covariate specification of Table 2 column 2. On every characteristic examined, including overdue balance, credit score, repayment before day-26 assignment, age, gender, and education, the design-weighted mean tracks the sample mean within 0.1 standard deviations in both sectors, so the borrowers who carry the identifying variation look like the estimation sample and the UJIVE estimates are close to representatively weighted average effects. Moreover,

[30]The JIVE family comprises JIVE1 (Angrist et al., 1999), IJIVE (Ackerberg and Devereux, 2009), SJIVE (Frandsen et al., 2026), and, as a deliberately biased benchmark, 2SLS using caller-by-month dummies directly as instruments. The caseload floor is lowered to 10 cases, raised to 20 cases, or removed.

the weighted averages are close between the in-house and third-party samples, so the two sectors' UJIVE estimates aggregate treatment effects over similar borrower populations.

Columns 5 and 6 split borrowers by the number of emergency contacts they have. For both in-house and third-party collectors, the effect concentrates among borrowers with at least three emergency contacts in terms of both the magnitude and statistical significance. This suggests that at least a meaningful portion of the treatment effect documented here works through the social-network channel rather than through generic collector aggressiveness, consistent with the effect of social reputation pressure generated by social-contact calling in debt collection.

The day-30 recovery gains, however, do not persist through longer horizons. Appendix Table B7 re-estimates the covariate-adjusted UJIVE specification of Table 2 column 4 with the recovery horizon extended in five-day steps from day 35 through day 59. The third-party effect falls by a quarter by day 35 and disappears entirely from day 40 onward. The in-house point estimates remain close to the day-30 magnitude at every horizon but are no longer statistically significant at conventional levels from day 35 on. This fading pattern may reflect social-contact calling accelerating repayment rather than improving ultimate repayment, since calling borrowers' social contacts is unlikely to resolve their fundamental financial difficulties. Nevertheless, because the day-30 delinquency ratio is an important metric of the lender's loan performance that could affect regulatory compliance and future funding costs, the lender still has incentives to use harsh collection practices despite weak long-run recovery impact. On the other hand, the pattern is also consistent with caller rotation attenuating heterogeneity across initially assigned callers over time, as discussed in Section 1.3.2.

## 4.2 Borrower return propensity

A potential cost of social-contact calling is the damage to the customer relationship as delinquent borrowers are shamed and irritated by such aggressiveness (Liao et al., 2026). As discussed in Sections 1.5 and 3.2.2, I measure the quality of the customer relationship by the reborrowing propensity of cured borrowers who resolve the overdue balance before charge-off, i.e., within 60 days of the day-26 assignment. I estimate the same OLS and UJIVE specifications as in the previous section but now replace the recovery outcome variable with indicators of any *application* in the 0–30 days (the first reborrowing wave), 0–180 days, and 180–210 days since cure. These outcomes capture borrowers' willingness to return to the lender independent of its approval decisions. For the 180–210-day window, which captures the second reborrowing wave after the delinquency exits the six-month lookback

window of the screening system, I further distinguish applications by whether they are from first-time returners, who never applied before day 180, or from repeat returners.

Table 3 reports the results. Column 2 of Panel A shows that a cured borrower randomly assigned to an in-house caller who calls social contacts 20 percentage points (roughly 1 s.d.) more often is 3.1 percentage points less likely to return in the first month right after the account cure. The magnitude is economically significant since it accounts for nearly half of the baseline return rate of 7% in the first month after cure. In contrast, if the borrower is contacted by the same type of caller but from a third-party agency, their propensity to return is barely affected by social-contact calling: the third-party coefficient, $-0.026$ ($t = -0.80$), is one-sixth of the in-house one and statistically indistinguishable from zero. The test of the sector difference in Panel C shows that the difference between in-house and third-party callers is significant at the 5% level. The significantly positive OLS coefficient in column 1 illustrates the severity of the selection problem: callers escalate to social contacts precisely when facing less responsive borrowers in deeper distress, who are also quicker to seek new liquidity after cure.

The first-month reduction in reborrowing, however, does not persist. As reported in column 4, the UJIVE estimates in the wider $[0, 180]$ window are null in both sectors. This is plausible because these early applications are dominated by the first wave of impatient borrowers who look for new liquidity immediately after cure, and their deteriorating financial conditions eventually override whatever factor suppresses reborrowing. Therefore, the null results in column 4 may simply reflect liquidity-need dynamics and are uninformative about the persistence of the relationship damage.

The persistence question is better evaluated in the $[180, 210]$-day window, the second reborrowing wave, among borrowers who never applied before day 180. Borrowers' access to credit is restored on day 180, which is a natural re-entry point for a preserved relationship. Column 6 shows an economically larger reduction on this margin among in-house-collected borrowers: a 1 s.d. more aggressive in-house caller reduces the propensity of first-time return by 1.5 percentage points, or 45% of the 3.4% baseline rate. Again, such a reduction does not carry over to the third party, whose coefficient is null and significantly different from that of its in-house counterpart.[31] Both margins are also stable across alternative jackknife estimators and various caseload filters, as reported in Appendix Table B8, which re-estimates the column-2 and column-6 specifications.

The in-house results point to damage to the lender's customer relationship: social-contact calls impose social reputational pressure on borrowers, which helps recovery but

[31] In contrast, the screening relaxation has no significant effects on return propensity for borrowers who have tried to apply before day 180, as reported in column 8.

also shames them in front of their relatives, colleagues, and friends by disclosing their weak financial conditions. Angry borrowers blame and "punish" the lender for such collection aggressiveness by walking away.

However, the difference in effects between in-house and third-party collectors indicates that the lender is blamed *only* when the calls are made in-house. Borrowers do not attribute the aggressiveness to the originating lender when reached by a third-party agency. Such a disparity implies that outsourcing shifts the blame for collection aggressiveness away from the lender, a commonly observed strategy in many firms' daily operations, as in the opening examples involving the BBC, Amazon's trucking contractors, and British Airways. Combined with the recovery results in Section 4.1, this asymmetry gives outsourcing a clear economic rationale: delegating the collection of delinquent debt keeps the full recovery benefit of social-contact calling while shedding its customer-relationship cost.

Some other mechanisms may also explain social-contact calling's recovery enhancement and reborrowing reduction simultaneously, such as the alienation of marginal borrowers cured by $\%social$, updated expectations of the social reputation costs of default, or informal family insurance activated by social calling, which will be discussed in more detail in Sections 4.4 and 4.5. Most importantly, these channels all operate through the borrower's experience of social-contact calls, which is the same treatment under in-house and third-party collection; only the caller's organizational affiliation differs. None of them therefore predicts a difference in effects based on the caller's organizational affiliation.

## 4.3 Performance of returning loans

Are the borrowers who keep returning despite aggressive collection worse credits? Table 4 examines the realized performance of the originated returning loans evaluated by three different metrics. The bad-loan rate is the principal-weighted share of loans in which at least one installment scheduled within the evaluation horizon becomes 30 or more days past due. The payment fulfillment rate is the share of the payments scheduled within the horizon that the borrower actually makes. The ROA (return on assets) combines repayment performance with loan pricing to measure realized profitability: it is the payments collected within the horizon minus the principal scheduled for repayment in the same horizon, divided by the average outstanding principal and expressed as a monthly rate. In other words, it is the monthly excess return on the capital at stake.[32] For each loan origi-

[32]Equivalently, the numerator is the interest, fee, and penalty payments collected within the horizon (gross revenues collected), net of the scheduled principal that is not collected by its due date, which is assumed to be completely written off. The denominator is the outstanding book principal cumulated over the horizon in yuan-months, with written-off principal removed from the book at its due date.

nated in the panel's application window, the three metrics are evaluated after the first and the third scheduled installments and over the full loan term. The UJIVE specifications are at the borrower level, so each metric is further aggregated to the borrower level. As before, I consider application windows both before and after the day-180 screening gate. However, since the loan approval rate is low before day 180 of cure, I replace the pre-gate $[0, 30]$ window emphasized earlier with the wider $[0, 180]$ window to obtain a larger sample of originated loans and improve statistical power.

Loans originated to early returners within 180 days after cure perform quite differently between the two sectors. For third-party callers, the borrowers who return despite more social-contact calling take out loans that perform *better* on every margin. After one month of origination, for every 1 s.d. (20-percentage-point) increase in $\%social$, the bad-loan rate falls by 13.2 percentage points ($t = -2.44$), fulfillment rises by 6.9 percentage points ($t = 1.08$), and the ROA point estimate rises by 4.5 percentage points per month of capital exposure ($t = 1.19$), though the last two effects are estimated imprecisely. The pattern is qualitatively similar at the three-month and full-term horizons. In contrast, the pattern reverses for in-house callers, and the sector difference is significant in eight of the nine comparisons in columns 7–9. The performance gap between sectors suggests an intensive margin of the relationship damage that stays in-house: social-contact calls from in-house collectors, if anything, tend to drive away borrowers with higher credit quality relative to the lender's screening standards, while the same calls from third-party collectors effectively screen the pool by scaring away borrowers with adverse private information. Section 4.6 discusses the intuition behind these opposite selection patterns.

Among first-time returners at the $[180, 210]$ screening relaxation gate (Panel B), the effects are attenuated and none is statistically significant, but the same signs appear on every metric and horizon, with the sole exception of third-party payment fulfillment over the full term. The in-house point estimates indicate worse performance with more social-contact calls, while the third-party estimates indicate better performance, with sector gaps of the same sign as in Panel A.

Finally, loans originated to returners within 180 days after cure generally generate a negative ROA of around $-16\%$, suggesting that approved borrowers are of lower credit quality than the screening system expects, corroborating the earlier characterization of the early applications as a first wave of impatient borrowers in deteriorating financial condition. In comparison, the average ROA is close to zero for loans originated to first-time returners in the $[180, 210]$ window.

## 4.4 Compositional vs. causal effects

The analysis of reborrowing is restricted to borrowers who cure within 60 days of the day-26 assignment because uncured borrowers are generally unable to apply for new loans from the platform, as discussed in Section 1.5. This raises a concern that the reduction in reborrowing documented in Section 4.2 is only compositional. In other words, the reduction may simply reflect the fact that marginal borrowers cured by social calling are intrinsically less likely to return, and social calling may actually have no causal damage on the company's goodwill with any given borrowers. This type of concern is common in program evaluation: for example, economists are usually interested in the wage effect of a job training program, but wages are observed only when the participant is employed, a margin that is also affected by the program (Lee, 2009).

Following Lee (2009), I bound the *causal* damage of social-contact calls to the lender relationship among borrowers who would cure within 60 days regardless of such calls. To make the intuition concrete, I assume in this section that the treatment is binary: borrowers are treated if they are randomly assigned to a caller whose propensity to call social contacts is above the median in that calendar month. The smallest damage in magnitude, the upper Lee bound on the effect, obtains when the compositional effect is maximal: all "compliers," those who cure only when treated, have the lowest possible return propensity, namely zero. Conversely, the largest damage, the lower bound, obtains when all compliers have a return propensity of one. The details of the Lee bounds construction can be found in Appendix A.2.

Figure 4 reports the Lee bounds for the effect of the above binary treatment, along with Imbens and Manski (2004) 95% confidence intervals based on bootstrapped standard errors. The UJIVE estimates under the continuous treatment, multiplied by the first-stage effect of the binary treatment on $\%social$, are also plotted for reference. For the in-house damage over the $[0, 30]$ window and the $[180, 210]$ first-time-returner window, both Lee bounds are well below zero, with the Imbens–Manski 95% confidence intervals marginally crossing zero, suggesting that the compositional effect, if any, does not overturn the results documented in Section 4.2.

## 4.5 Alternative mechanisms

This section investigates two alternative mechanisms that could simultaneously explain the improvement in recovery and the reduction in reborrowing: increased expectations of the social reputation costs of default, and the activation of informal family insurance. Section 4.5.1 examines a prediction common to both mechanisms: that borrowers'

credit demand with alternative lenders also declines with %*social*. Sections 4.5.2 and 4.5.3 then test the implications specific to each channel. I find no supporting evidence for either channel. More fundamentally, both channels operate through the borrower's experience of social-contact calls *per se*, independent of who places the calls; neither, therefore, can account for the sharp differences by caller affiliation documented above.

### 4.5.1 Credit demand with alternative lenders

The decline in borrowers' propensity to return to the lender could reflect a general reduction in credit demand—driven by updated beliefs about the costs of delinquency, additional liquidity from social contacts, or other channels—rather than lender-specific damage to the customer relationship. The two explanations make opposite predictions for borrowing elsewhere: a general reduction in credit demand should lower applications to other lenders as well, whereas lender-specific damage should leave them unchanged or even raise them as borrowers redirect their demand. I test these predictions using soft credit inquiry records that capture borrowers' activity at other lenders.

The company runs monthly soft inquiries on the external credit reports of a set of active users for routine account reviews and marketing. One of these reports comes from a credit aggregator that partners with other lending companies and banks to compile borrowers' borrowing activity across platforms and institutions. The report includes the number of credit applications, either approved or rejected, that the borrower submitted to the aggregator's partner lenders over lookback periods ranging from one week to two years before the inquiry date.

Table 5 examines the relationship between social-contact calling and loan applications to other lenders over the one-week lookback before each inquiry; extending the lookback to one month yields qualitatively similar results. I restrict the sample to the same cured borrowers as in the return propensity analysis and use inquiries made within 30 or 180 days after cure.[33] As columns (1) and (2) show, neither the probability of being queried within 30 days after cure nor the timing of the inquiries responds to %*social* in either sector, and both are nearly identical across sectors (13% of cured borrowers are queried, on day 9 on average). Nevertheless, columns (3)–(6) control for inquiry timing to account for the time trend in borrowing activity after cure.

Column (3) regresses the number of applications to other lenders on %*social* with the same UJIVE specification as before. Among cured borrowers with at least one inquiry

[33]Because the analysis is at the borrower level, for a borrower with multiple inquiries within the window, the dependent variable is the average application count across these inquiries. I do not use the [180, 210] window as in the return propensity analysis since the 180-day credit access relaxation is likely lender-specific.

within 30 days after cure, a 20-percentage-point (roughly 1 s.d.) increase in %*social* from in-house callers raises the number of applications to other lenders by 0.14, or 12% of the mean. Because these borrowers are also less likely to return, the result suggests that they divert part of their credit demand away from the lender, consistent with lender-specific relationship damage rather than a general withdrawal from borrowing. In contrast, the third-party estimate is negative, as a general reduction in credit demand would predict, but insignificant. The sector difference in Panel C is significant at the 5% level.

Restricting the sample to non-returners yields a slightly larger in-house effect (column 5), so the increase does not come from borrowers who also return to the lender. Extending the inquiry window to 180 days after cure yields estimates with the same signs that are smaller in magnitude and statistically insignificant (columns 4 and 6), mirroring the null return effect over the $[0, 180]$ window in Section 4.2.

The in-house reduction in returning therefore does not reflect a general withdrawal from credit, contrary to the prediction shared by the reputation-cost and informal-insurance channels. The next two subsections test the implications specific to each channel.

### 4.5.2 Expected social reputation costs of default

One potential channel involves the social reputation costs of default. Social-contact calls impose social pressure on borrowers by reaching their relatives, colleagues, and friends, so borrowers repay delinquent debts early, trading off liquidity needs against damage to their social image. This resembles social collateral (Besley and Coate, 1995; Karlan et al., 2009; Perez-Truglia and Troiano, 2018; Breza and Chandrasekhar, 2019). Meanwhile, the experience also leads borrowers to revise upward the social reputation cost they expect to incur upon a future default, thereby suppressing their subsequent demand for credit. Since the expected cost of default is the cost given default multiplied by the probability of default, the deterrent should be strongest where a future default is most likely.

In particular, the reduction in reborrowing should be flat or increasing in loan size, and never confined to small loans: whatever cost a borrower comes to expect upon default, larger loans are more likely to default.[34] Table 6 characterizes the loan-size profile of returning by replacing the any-application indicators in the baseline regressions with indicators that the cumulative applied-for amount exceeds successively higher thresholds.

[34] Among the loans originated to returning borrowers, the share that becomes at least 30 days delinquent over the full loan term is 16.7%, 19.8%, and 27.4% in the bottom, middle, and top terciles of loan size, respectively. A similar positive relationship between loan size and default risk is documented for subprime auto loans by Adams et al. (2009) and, through the selection of riskier borrowers into larger loans, for payday loans by Dobbie and Skiba (2013). If the reputation cost is instead fixed per default and weighed against the benefit of the loan, it is relatively larger for small loans (Perez-Truglia and Troiano, 2018). The outside-borrowing evidence in Section 4.5.1 does not depend on this distinction.

I find no supporting evidence, at least in-house. Among in-house-collected borrowers, the reduction is concentrated among smaller applications, in both statistical significance and relative magnitude. The expected reputation cost channel, if any, appears only among third-party-collected borrowers, where coefficients are larger and significant for applications above 5,000 yuan (700 USD) in the first month after cure (Panel C). Moreover, the channel predicts less borrowing from all lenders, whereas in-house-collected borrowers apply more to other lenders (Section 4.5.1).

### 4.5.3 Informal family insurance

Another candidate mechanism is that social-contact calls activate informal family insurance. Borrowers tend to underutilize informal family support, as seeking help carries a stigma (Chandrasekhar et al., 2018) and family financing carries shadow costs (Lee and Persson, 2016). By disclosing the delinquency directly to the borrower's network, social-contact calls may make disclosure a sunk cost and push borrowers to their families, whose transfers and informal loans provide liquidity (Fafchamps and Lund, 2003) that can both fund repayment and substitute for external credit.

I test this mechanism by examining effect heterogeneity with respect to the share of kin among each borrower's callable social contacts. Based on the relationship labels attached to each contact number in the platform's records, I define kin contacts as the borrower's parents, siblings, children, and other relatives, in contrast to non-kin contacts such as colleagues and friends; spouses form a separate category, as spousal money is intra-household rather than insurance from the wider family network. Since informal credit and insurance flow largely through family and kin networks (La Ferrara, 2003; Angelucci et al., 2018), the mechanism predicts stronger effects—larger recovery gains and larger reborrowing reductions—among borrowers whose contact networks are kin-rich. Appendix Table B9 augments the regressions of both outcomes on $\%social$ with an indicator for the top tercile of the kin contact fraction and its interaction with $\%social$. I find no evidence for the channel: none of the interaction terms is statistically significant at the 5% level and their signs display no consistent pattern. The only marginally significant interaction, a larger recovery effect among kin-rich third-party borrowers, is not matched by a larger reborrowing reduction in the same group.

## 4.6 Interpretation: two effects of harsh collection

The results so far reveal a difference in who is deterred by social-contact calling. When in-house collectors make the calls, borrowers are less likely to return after repaying,

especially among those of higher credit quality. In contrast, calls from third-party collectors leave returning largely intact and, if anything, drive away riskier borrowers. I interpret this difference through blame shifting: the collector's identity changes what borrowers believe about the sources of harshness.

The key distinction is between two components of perceived collection harshness. *Lender-specific* harshness is the part that borrowers attribute to their current lender, whether as resentment toward the lender or as an expectation of the same treatment if they borrow from it again. Switching lenders offers a way to avoid this cost. *Market-wide* (non-lender-specific) harshness is the part that borrowers associate with debt collection in general and expect to encounter after a future default regardless of which lender they choose. Avoiding this cost requires reducing borrowing rather than merely switching lenders.[35] The collector's identity affects the relative importance of these components. Calls made by a lender's own employees are more readily attributed to that lender; outsourcing allows borrowers to attribute more of the experience to the agency or to collection practices in the broader credit market.

This distinction matters because the two components create different incentives to leave. Borrowers who are more likely to default on a future loan face greater exposure to the harsh treatment again, so either component can discourage riskier borrowers from borrowing again from any lender.[36] Lender-specific harshness, however, adds another consideration since borrowers can avoid the offending lender without giving up credit altogether. Safer borrowers tend to have better options at other lenders and can therefore switch more easily. This switching margin makes lender-specific harshness relatively more likely to drive away safer borrowers, worsening the credit quality of those who remain.

These forces yield the central prediction. In-house collection makes lender-specific harshness more salient, encouraging borrowers with attractive outside options to leave the lender. Delegation shifts perceived harshness toward its market-wide component, reducing this selective loss of safer customers. Therefore, even if both forms of collection discourage some future borrowing, delegation should leave the lender with a better-quality returning pool relative to in-house collection, as suggested by the performance results in Section 4.3.

The evidence on borrowers' subsequent credit choices is consistent with this interpretation. Following in-house collection, the decline in returning is concentrated among

[35]The distinction between harshness that borrowers can escape by switching lenders and harshness they cannot is also central to Fedaseyeu and Hunt (2018), in which lenders' delegation to a common collection agency arises in equilibrium and borrowers rationally anticipate it. Here, the split instead depends on whom borrowers hold responsible for the harshness they experience.

[36]Another possibility is that riskier borrowers are less sensitive to this risk of future harsh treatment if they have less to lose from a future default, for example because their access to credit is already impaired or their standing among their social contacts is already damaged by earlier collection.

small applications and is accompanied by more applications to other lenders, consistent with switching away from the offending lender. Following third-party collection, the decline is larger for larger applications, while applications to other lenders, if anything, fall, consistent with a broader reduction in demand for credit. The expected cost of default, examined in Section 4.5.2, can be interpreted as the market-wide part of harsh collection, but because it does not depend on whom borrowers blame, it cannot explain why the damage to the lender arises only in-house.

# 5 Calling Efforts and Communication Strategies

As discussed in Section 3.1.4, if callers with high social-calling propensities also differ systematically in other dimensions of collection activities, the estimand captures the effect of the bundled style. To characterize the bundle, Section 5.1 focuses on generic operational calling efforts, such as calling intensity and sequence, and Section 5.2 explores variations in communication strategies in call transcripts using large language models.

## 5.1 Other calling efforts

I regress a set of observable calling effort metrics, each measured over days 26–30, on instrumented $\%social$, using the same UJIVE specification as in Table 2. The metrics cover calling volume and frequency, case coverage, contact sequence, actual outreach, and payment waivers.

Table 7 reports the results. First, social-contact calling is positively associated with the total calling intensity among third-party collectors. Specifically, a 20-percentage-point (roughly one standard deviation) higher $\%social$ corresponds to 20% more calls per day, so the third-party estimates could capture the impacts of calling intensity. Among in-house collectors, in contrast, the relationship is also positive but insignificant. If calling intensity harms the customer relationship, which is likely since frequent calls can be quite disturbing, the sector difference in overall calling intensity could not reconcile the gap in the sensitivity of return propensity to $\%social$. At the phone-number level, the fraction of days receiving any call attempts and the time intervals between attempts both decline with social-contact calling, suggesting that high-$\%social$ collectors tend to concentrate their attempts, instead of exerting persistent efforts on borrowers.

Higher overall calling intensity translates to higher overall outreach. In both sectors, a 20-percentage-point increase in $\%social$ raises total answered calls by about 14% of the mean, leading to a 15% rise in total call duration in-house and 10% through third parties.

The duration per answered call, however, turns out to be flat in both sectors. Therefore, part of the overall calling intensity reflects the "breadth" of contact rather than the "depth".

Further analysis of other calling patterns supports the "breadth" versus "depth" trade-off. In both sectors, callers with higher %*social* actually make fewer calls per number per day, begin calling social contacts after fewer calls to the borrower's own numbers, and, among in-house collectors, are less likely to follow up with the borrower on the same day after having reached them.

Finally, the style involves no differential leniency in debt forgiveness. As reported in the last row, conditional on borrowers making any payment, the payment reduction granted per unit of overdue balance is unrelated to %*social* in either sector. Therefore, the concession policy does not appear to be part of the strategy bundle.

Taken together, regarding operational calling efforts and strategies, high %*social* corresponds to high overall calling intensity and outreach, but also more dispersed efforts leaning towards borrowers' social contacts. High-%*social* collectors in both sectors act similarly relative to their same-sector colleagues, so the sector differences in calling efforts are unlikely to explain the sector gaps in outcomes documented earlier.

## 5.2 Communication strategies and borrower responses in call transcripts

Apart from generic calling effort, the social-contact calling harshness index, %*social*, could also capture other aspects of harshness in collectors' persuasion strategies and language. This section explores this possibility by using large language models (LLMs) to analyze phone call transcripts, which are transcribed by the company primarily for collector conduct auditing.

Among the 332,665 answered calls between days 26 and 30 in the estimation sample that last at least 6 seconds and are not answered by an automated voice message or voicemail, 263,090 (79%) are matched to a transcript provided by the company according to the encrypted phone number and the time of the call (78% of in-house calls and 91% of third-party calls).[37] Unmatched phone calls are generally due to operational reasons unrelated to the content or outcomes of the calls.[38] Nevertheless, within each sector, I reweight the matched calls by the inverse of the conditional matching probability estimated

[37]The company transcribed few calls shorter than 6 seconds since they usually contain no information.

[38]For example, among third-party calls in 2021–22, calls starting during noon–12:30 P.M. and calls ending after 7:45 P.M. were transcribed at lower rates, and calls still in progress when the clock hour changed were lost probably due to hourly batching. Transcription rates also differ by calendar month and dialing system, reflecting differences in workload and in the difficulty of retrieving the recordings for transcription.

with an extensive set of inputs to an XGBoost model, whose explanatory power turns out to be mostly from operation-related variables.[39] I analyze all matched third-party calls and all matched calls to borrowers' own numbers; for in-house calls to social contacts, I analyze only the calls of a random 25% sample of borrowers and weight them by the inverse of the sampling rate. The final transcript sample includes 79,440 calls with in-house collectors and 31,407 calls with third-party collectors. The agency relationship disclosure clauses in third-party calls are identified and masked using a keyword matching algorithm so the disclosure does not bias the LLM's judgment.[40]

### 5.2.1 Collector communication strategies

I use an LLM to extract the persuasion and communication strategies used by the collector from the phone call transcripts (e.g., Asirvatham et al., 2026; Ludwig et al., 2025; Jha et al., 2024). For each transcript, the LLM is asked to read the whole conversation and choose the communication strategies used each time the collector speaks (a speaking turn) from a set of 25 strategies listed in Appendix Table B10 Panel A. The prompt describes each strategy with a short definition, followed by clarification of situations to which the label does not apply.[41] Each collector turn may carry more than one strategy. I construct the list based on the collection and persuasion literature (e.g., Liao et al., 2021; Laudenbach and Siegel, 2025; Choi et al., 2025), the company's and regulator's guidelines, and strategies nominated by an LLM reading a subsample of conversations, and refine it over several rounds of pilot runs and manual verification (Appendix A.3.2).

I use Qwen3.8-27B in non-reasoning mode with zero temperature, which provides a good balance between accuracy, stability, and speed.[42] The labels are validated by comparing to a human reader's labels on a random sample and by showing several labels'

[39]The operation-related variables include dialing system indicators (e.g., cell phone/landline, manual/auto-dialing), call timing (e.g., hours, minutes, seconds, days of the week, days of the month, calendar months), and call characteristics (e.g., ring seconds, call recipient type, and call order). See Appendix A.3.1 for details.

[40]The algorithm searches for agency self-identifications such as "I'm a delegated representative of <the lender>," and rewrites them into "I'm with <the lender>," the in-house callers' standard self-introduction. In an audit sample, an LLM asked whether the caller is an outside agency or the lender's own staff is correct 84% of the time on the raw transcripts, and 52% of the time, i.e., at chance, on the masked transcripts.

[41]For example, the description of the `legal threat` strategy (translated from the Chinese prompt) is: the collector says or implies that a legal or judicial process will follow (a lawsuit, court, arbitration, a lawyer's letter, the police, criminal liability, enforcement, or listing as a dishonest debtor). It does not apply to a handover to a non-legal department, a visit, or "the process" (`escalation threat`), or to threats to personal safety (`intimidation`); a legal claim that is false or impossible for this debt, such as imprisonment for an unpaid consumer loan, is additionally coded as `deception`.

[42]Enabling reasoning takes about 14 times as long per call as the non-reasoning mode without improving classification accuracy. See Appendix A.3.4 for details.

predictive power on subsequent collection outcomes after the calls.[43] More details on the validation are in Appendix A.3.3.

To characterize how collectors' communication strategies relate to %*social*, I reuse the same UJIVE specifications as in Section 5.1, but, for each of the 25 strategies, I change the dependent variable to an indicator of the presence of the strategy in the first conversation with the borrower's own phone number in days 26–30. I first focus on calls with borrowers' own numbers since they are generally more informative than calls to borrowers' social contacts, which will be briefly discussed later. Restricting to the first conversation of each borrower removes the influence of earlier contacts. The sample comprises 14,251 in-house and 3,454 third-party borrowers with a transcribed first conversation.

Figure 5 shows how the strategies a caller uses in the first conversation with the borrower vary with %*social* in the two sectors. The figure visualizes the UJIVE estimates and their 95% confidence intervals, rescaled to the estimated effect of a 20-percentage-point (roughly one standard deviation) increase in %*social* (numerical results in Appendix Table B11, Panel A). I classify the strategies into four broader categories. The first category, information-related communication strategies, shows similar and relatively small UJIVE estimates for both in-house and third-party sectors.[44]

The most salient correlation between strategies and %*social* concentrates in the "threats and consequences" category. For example, a one-standard-deviation increase in %*social* in-house is associated with a 5-percentage-point higher probability of using a legal threat, which emphasizes the possibility of a future legal or judicial process. The association is much larger (15 percentage points) and more significant among third-party collectors. This pattern suggests that the %*social* measure also captures the effect of legal threats as a dimension of harshness, but it is unlikely to explain the sector difference in return propensity: if legal threats damaged the customer relationship, third-party collection, where the association is stronger, should show the larger reduction in borrower return, the opposite of what the data show. Similar patterns are also observed in escalation threats, credit warnings, and cost warnings in this category, as well as deception and accusations of bad faith in the "misconduct" category, to which similar rationales apply.

Another interesting observation is that in-house %*social* is significantly associated with social-contact warnings, which mention the possibility of informing the borrower's

[43]The F1 score is 0.83 on average across the collector strategies and about 0.9 for the main ones such as legal threats, credit and social-contact warnings, and urgency.

[44]One exception is the verification of the call recipient's identity where the third-party estimate is negative, so harsh third-party collectors as measured by %*social* are less likely to follow the standard procedure of identity verification. That said, this procedural flaw is usually considered minor and unlikely to have a substantial impact on recovery and return outcomes.

social contacts if the delinquency continues. The actual calls to contacts presumably make the warning credible. Among third-party collectors, in contrast, the relationship is insignificant, yet the recovery effect of %*social* is similar across sectors, suggesting that actions may matter more than words.

The other strategies in the second group, namely stating vague consequences of delinquency, stressing urgency and deadlines, and making moral appeals, are also positively associated with %*social*. Although the in-house estimates are more significant than third-party estimates, the magnitudes are similar across sectors, so they are unlikely to explain the sector difference in relationship damage.

%*social* also captures variation in the use of negotiation strategies among collectors. High-%*social* in-house collectors are more likely to provide suggestions to seek temporary liquidity from other sources ("Funds-source suggestion"), typically family members and friends, so part of the association might simply reflect the social-contact calling tendency. These collectors also tend to downplay the amount and respond to borrowers' complaints more often, but the relationships are only marginally significant after the Benjamini and Hochberg (1995) multiple-testing adjustment. In contrast, high-%*social* third-party collectors are more likely to mention fee waivers and concessions during negotiation ("Concession offer"). However, as we have seen in Table 7, the payment reduction actually realized, which is observed only when the borrower pays and takes up the offer, is similar across sectors and flat with respect to %*social*.

In addition, I ask the LLM to rate each conversation from 1 (least) to 5 (most) from five aspects, including the overall imposed pressure, professionalism, empathy, and negotiation effort of the caller and the resistance of the borrower. A grading scheme, briefly reported in Appendix Table B10 Panel B, is provided to the LLM. Figure 6 reports the results in a way similar to Figure 5 (numerical estimates in Appendix Table B11, Panel B). High-%*social* callers generally impose greater pressure on borrowers, consistent with the use of various threats and consequence warnings, especially among third-party collectors. High-%*social* third-party collectors also display less professionalism, meaning that they are more likely to use a harsh tone and show procedural flaws and misconduct. High-%*social* callers also tend to show less empathy and receive greater borrower resistance, although none of the effects is significant after the multiple-testing $p$-value adjustment.

### 5.2.2 Borrower response

Figure 7 turns to the borrower's side of the same first conversations: the borrower's moves and expressed emotions, as defined in Appendix Table B10 Panels C and D, with numerical estimates in Appendix Table B12. The sector contrast shows a clear and consistent

pattern. Borrowers facing high-%*social* in-house collectors push back harder. Specifically, a 20-percentage-point increase in %*social* raises the probability that the borrower insults the collector by 3.0 percentage points, refuses to pay by 2.7 percentage points, asks the collector to stop calling by 2.9 percentage points, and requests relief by 3.1 percentage points, and the borrower's expressed emotion becomes more negative and more intense, with the probability that the borrower expresses anger rising by 4.9 percentage points. Borrowers facing high-%*social* third-party collectors, in comparison, show none of these reactions.[45] The same harshness bundled with social-contact calling therefore provokes anger and withdrawal when it comes from the lender's own staff but not when it comes from a third-party agency.

### 5.2.3 Additional analysis

Appendix Tables B13 and B14 report the same estimates for the calls to the borrower's social contacts. The patterns are qualitatively similar to those on calls to the borrower's own number: high-%*social* third-party collectors rely more on threats and consequence warnings, whereas high-%*social* in-house collectors elicit more hostility, stop-contact requests, and anger from the contacts they reach.

Because the two sectors dial differently, the borrowers who answer may differ between sectors. I therefore re-estimate the in-house regressions with in-house answering borrowers reweighted to the covariate distribution of third-party answering borrowers.[46] Appendix Table B11 reports the results, which are quantitatively similar.

### 5.2.4 Summary of call transcript LLM analysis

Overall, the call transcript analysis suggests that social-contact calling propensity also captures other aspects of harshness in communication, such as a higher tendency to use threats and consequence warnings and to display misconduct. These associations are stronger among third-party collectors. If these communication strategies reduced borrower return regardless of who used them, the reduction would be larger under third-party collection, the opposite of what the data show. The borrower's side of the conversations shows that the harshness bundled with social-contact calling provokes anger, hostility, and requests to stop contact when it comes from the lender's own staff, but not when it comes

[45] The only significant association is that they ask for more information about the debt (4.3 percentage points), and their expressed valence is, if anything, slightly more positive.

[46] Specifically, I first use an XGBoost model to estimate $e_i = \Pr[\text{third-party} \mid W_i]$, the probability of answering borrower $i$ being third-party-treated among all answering borrowers, conditional on pre-treatment borrower and loan characteristics $W_i$. Each in-house-treated borrower $i$ is then additionally re-weighted by $e_i/(1-e_i)$ on top of the transcription probability re-weighting.

from a third-party agency. Delegation therefore appears to attenuate the damage from harsh communication bundled with social-contact calling.

# 6 The Lender's Problem and Regulation

## 6.1 Profitability of social calling to the lender

As we have seen in Section 4, calling social contacts significantly improves debt recovery through day 30 past due but, when performed in-house, also causes relationship damage, as suggested by a reduction in the return of cured borrowers, especially among those with high credit quality. A natural question, then, is the combined effect of social-contact calling on the lender's total profitability.

I define the lender's total profitability (profits and losses, P&L) as the sum of the NPV of collections recovered by day 30 of delinquency, plus the realized lifetime NPV of all payments from the borrower's return loans originated within a stated window *after the day-26 assignment* (cash collected net of principal, discounted to the assignment date), scaled by the overdue balance at the assignment date. The sum includes both the short-run benefits of recovery and long-run customer relationship value of reborrowing.

This definition of P&L is a natural choice given the conventions established in previous sections, but it warrants some further clarification. First, the definition applies to *all* borrowers, including non-cured, non-returning, and always-rejected borrowers, whose long-run relationship value is zero by construction. Second, the choice of the evaluation window for short-run recovery is not obvious. From the perspective of callers, the day-30 window may be reasonable, since their effort and contribution are front-loaded, as discussed in Section 1.3.2. Meanwhile, the day-30 recovery is also central from the lender's perspective, since the day-30 delinquency ratio is a major metric of the lender's loan performance and could affect the lender's regulatory compliance and future funding costs. Finally, it is unclear whether the lender weighs the two components equally at face value. On the one hand, the lender may emphasize short-run recovery for the reasons given above. On the other hand, the long-run damage to the firm's customer relationship could go beyond the current borrower if the adverse collection experience spills over to other borrowers. I therefore view the analysis here as a simple back-of-the-envelope illustration of the lender's decision, rather than a complete profitability analysis.

Table 8 reports the estimates of $\%social$'s effect on total P&L under UJIVE specifications. First, notice that the magnitudes of the coefficients are similar to the baseline recovery regressions in Table 2, suggesting that the total P&L is mainly driven by recovery-

side gains. However, the customer relationship side also shows up in the total P&L. In particular, the third-party coefficients are generally larger than those from recovery alone and are significant across all return-loan origination windows. In contrast, in-house coefficients remain weak, fall slightly below the recovery-only estimate, and do not significantly differ from zero at the 5% level in any window. The results cannot reject the hypothesis that in-house collectors already operate at the level of $\%social$ that maximizes the lender's total P&L (i.e., where the derivative of profitability with respect to social calling intensity is zero), while the significantly positive third-party coefficients imply that third-party calling remains below the P&L-maximizing level. This is consistent with third-party collectors operating at a constrained optimum due to regulatory requirements, a point I return to in Section 6.3.

## 6.2 The outsourcing decision

This analysis speaks to one important consideration for the lender's make-or-buy decision in collections. Specifically, outsourcing mitigates the customer-relationship cost of aggressive collection while maintaining its benefits for recovery. The same social-contact calling that is at best marginally profitable in-house, therefore, raises the lender's total P&L when a third-party agency makes the calls. This blame-shifting gain should be larger where harshness is more valuable for recovery, and smaller the more borrowers hold the lender responsible for its agencies' conduct. The first implication is consistent with the common industry practice of outsourcing later delinquency stages, in which borrowers are harder to collect and harshness matters more for recovery.

The same logic suggests how the lender may organize delegation. Because the lender incurs less relationship cost under delegation, the agency needs to be made to internalize less of it, so organizational separation is compatible with higher-powered incentives, such as a contingency fee that pays the agency a share of recoveries. The lender may also prefer to stay at arm's length and forgo monitoring or control that would reduce conventional agency losses, if closer control makes the agency's conduct more attributable to the lender, as experimental evidence suggests (Hill, 2015).

The full decision, however, involves considerations beyond the scope of the current analysis. The lender's managers raised three points in my conversations with them. First, in-house performance on comparable cases serves as an internal benchmark for the lender to evaluate third-party agencies and to set and update their incentive schemes over time. Second, in-house collection activities generate important data for training the lender's models of collection strategy and default risk. Third, the in-house department carries

organizational value of its own: it retains experienced senior collectors and trains future managers of the collection operation. Other standard considerations may also be relevant. For example, capacity flexibility over the credit cycle and economies of scale favor outsourcing, while tighter control over compliance and data security favors keeping collection in-house. I therefore interpret the lender's persistent dual structure as an interior solution to this richer problem rather than as evidence of misoptimization on the collection margins measured here. The benchmarking motive also helps explain the lender's motivation for assigning comparable cases to in-house and third-party collectors over time, which is the feature my empirical design utilizes.

## 6.3 Regulatory implications

Regulators in many countries explicitly address outsourced debt collection. In the United States, the Fair Debt Collection Practices Act of 1977 (FDCPA), which is the federal statute governing collection conduct, applies to outside debt collectors and debt buyers but generally exempts creditors collecting their own debts in their own names.[47] In China, regulators make lenders directly accountable for the conduct of the collection agencies they hire and restrict whom they may hire.[48] Similar regulatory emphasis on outsourced providers appears in many other service industries.[49]

These findings provide an economic rationale for this regulatory emphasis. Social-contact calling can impose externalities beyond the borrower–lender relationship, such as shocks to the borrower's social networks and employment, as well as additional financial and mental distress. My results show that part of this externality is internalized when collection is performed in-house through the customer-relationship channel: losses of returning borrowers, especially the high-quality ones, offset the recovery gains. When a third party makes the calls, the borrower–lender relationship suffers less, so the feedback

[47]The FDCPA defines a covered "debt collector" as a person "in any business the principal purpose of which is the collection of any debts, or who regularly collects or attempts to collect …debts owed or due or asserted to be owed or due another," and expressly excludes "any officer or employee of a creditor while, in the name of the creditor, collecting debts for such creditor" (15 U.S.C. §1692a(6)).

[48]The national standard on post-loan collection for online consumer credit (GB/T 45251-2025), developed from the 2024 industry guideline of the National Internet Finance Association of China, provides that financial institutions "shall fulfill the primary responsibility for post-loan collection …and implement outsourcing prudently," and "shall strengthen the management, supervision, and inspection of third-party collection agencies."

[49]Regulators often refuse to let firms shed responsibility by outsourcing customer-facing services. In U.S. telemarketing, a seller may be held vicariously liable for unlawful calls placed by third-party telemarketers acting on its behalf (FCC, *In re DISH Network*, 28 FCC Rcd 6574, 2013); in healthcare, HIPAA holds providers and insurers accountable for the outsourced "business associates" that handle patient information (45 C.F.R. §§164.502(e), 164.504(e)); and under the EU's General Data Protection Regulation, firms remain responsible for their outsourced data processors (Article 28).

channel is attenuated. Absent external regulation, a lender therefore has an incentive to tolerate or even encourage social-contact calling by its outsourced agencies while restraining it in-house. Because restricting collection also carries real costs in reduced credit supply (Fedaseyeu, 2020; Romeo and Sandler, 2021), targeting stricter rules at third-party conduct concentrates regulation where private internalization is weakest and the market failure is largest.

The two regulatory approaches described above work through different channels. Conduct rules that apply only to outside collectors, as under the FDCPA, restrict delegated harshness directly. Rules that make the lender accountable for its agencies' conduct, as in China, instead restore part of the relationship cost that delegation removes: by making delegated conduct more attributable to the lender, or by imposing a direct liability cost on the lender for it, they give the lender itself a reason to restrain its agencies through the contracts it writes. Such rules also counteract the lender's incentive to stay at arm's length from its agencies' conduct discussed in Section 6.2.

# 7 Conclusion

This paper examines why a firm would delegate a task to outside agents when their conduct could damage the firm's goodwill, even when incentives are misaligned. Utilizing the exogenous variation in the harshness of collection borrowers receive, the paper shows the benefit of blame shifting by delegation. Specifically, the recovery benefits of harshness cross the firm boundary, while its costs to customer relationships stay within the firm. Differences in generic calling effort and communication strategies are unlikely to explain this gap.

Blame shifting changes how one should view the agency costs of outsourcing. The standard view treats an outside agent's failure to internalize the firm's customer relationships as a cost of delegation. When customers attribute the agent's conduct less to the firm, the same misalignment can be less costly to the firm's customer relationship.

The logic is not specific to debt collection. It applies wherever a firm must perform productive but unwelcome tasks, such as restructuring, layoffs, claim denials, or enforcing contract terms, and it suggests that the more valuable harshness is, the more likely such tasks are placed outside the firm. The blame-shifting motive also provides a rationale for the emphasis on regulating third parties, as blame shifting attenuates the externality of harshness internalized as customer relationship damage.

Of course, the ultimate make-or-buy problem is complex and involves multiple considerations beyond the blame-shifting margin, such as the regulatory environment, relative

cost-effectiveness, and informational advantages. The empirical setting holds the lender's organizational choice fixed and randomizes borrowers across the two arrangements, which allows me to identify the blame-shifting margin that a firm should consider when setting its organizational boundaries, but does not allow me to weigh blame shifting against these other considerations. Doing so would require variation in the make-or-buy decision and in the factors that shape it.

# References


Ackerberg, D. A. and P. J. Devereux (2009). Improved JIVE estimators for overidentified linear models with and without heteroskedasticity. *Review of Economics and Statistics 91*(2), 351–362.

Adams, W., L. Einav, and J. Levin (2009). Liquidity constraints and imperfect information in subprime lending. *American Economic Review 99*(1), 49–84.

Angelucci, M., G. De Giorgi, and I. Rasul (2018). Consumption and investment in resource pooling family networks. *Economic Journal 128*(615), 2613–2651.

Angrist, J. and M. Kolesár (2024). One instrument to rule them all: The bias and coverage of just-id iv. *Journal of Econometrics 240*(2), 105398.

Angrist, J. D., G. W. Imbens, and A. B. Krueger (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics 14*(1), 57–67.

Arnold, D., W. Dobbie, and P. Hull (2022). Measuring racial discrimination in bail decisions. *American Economic Review 112*(9), 2992–3038.

Asirvatham, H., E. Mokski, and A. Shleifer (2026). GPT as a measurement tool. Working Paper 34834, National Bureau of Economic Research.

Aydin, D. (2022). Consumption response to credit expansions: Evidence from experimental assignment of 45,307 credit lines. *American Economic Review 112*(1), 1–40.

Bartling, B. and U. Fischbacher (2012). Shifting the blame: On delegation and responsibility. *Review of Economic Studies 79*(1), 67–87.

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*(1), 289–300.

Besley, T. and S. Coate (1995). Group lending, repayment incentives and social collateral. *Journal of Development Economics 46*(1), 1–18.

Breza, E. and A. G. Chandrasekhar (2019). Social networks, reputation, and commitment: Evidence from a savings monitors experiment. *Econometrica 87*(1), 175–216.

Bursztyn, L., S. Fiorin, D. Gottlieb, and M. Kanz (2019). Moral incentives in credit card debt repayment: Evidence from a field experiment. *Journal of Political Economy 127*(4), 1641–1683.

Chandrasekhar, A. G., B. Golub, and H. Yang (2018). Signaling, shame, and silence in social learning. Working Paper 25169, National Bureau of Economic Research.

Chen, T. and C. Guestrin (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794.

Choi, J. J., D. Huang, Z. Yang, and Q. Zhang (2025). How good is AI at twisting arms? Experiments in debt collection. Working Paper 33669, National Bureau of Economic Research.

Chyn, E., B. Frandsen, and E. Leslie (2025). Examiner and judge designs in economics: A practitioner's guide. *Journal of Economic Literature 63*(2), 401–439.

Coffman, L. C. (2011). Intermediation reduces punishment (and reward). *American Economic Journal: Microeconomics 3*(4), 77–106.

Consumer Financial Protection Bureau (2023, October). The consumer credit card market. Report, Consumer Financial Protection Bureau, Washington, DC.

Consumer Financial Protection Bureau (2025a, December). The consumer credit card market: Report to Congress. Report, Consumer Financial Protection Bureau, Washington, DC.

Consumer Financial Protection Bureau (2025b, August). Defining larger participants of the consumer debt collection market. Federal Register 90(151): 38418–38421. Advance Notice of Proposed Rulemaking; Docket No. CFPB-2025-0030; RIN 3170-AB51.

Consumer Financial Protection Bureau (2026, March). Consumer response annual report, January 1 – December 31, 2025. Report, Consumer Financial Protection Bureau, Washington, DC.

Dobbie, W., A. Liberman, D. Paravisini, and V. Pathania (2021). Measuring bias in consumer lending. *The Review of Economic Studies 88*(6), 2799–2832.

Dobbie, W. and P. M. Skiba (2013). Information asymmetries in consumer credit markets: Evidence from payday lending. *American Economic Journal: Applied Economics 5*(4), 256–282.

Drozd, L. A. and R. Serrano-Padial (2017). Modeling the revolving revolution: The debt collection channel. *American Economic Review 107*(3), 897–930.

Fafchamps, M. and S. Lund (2003). Risk-sharing networks in rural Philippines. *Journal of Development Economics 71*(2), 261–287.

Fedaseyeu, V. (2020). Debt collection agencies and the supply of consumer credit. *Journal of Financial Economics 138*(1), 193–221.

Fedaseyeu, V. and R. M. Hunt (2018). The economics of debt collection: Enforcement of consumer credit contracts. Working Paper 18-04, Federal Reserve Bank of Philadelphia.

Fonseca, J. (2023). Less mainstream credit, more payday borrowing? Evidence from debt collection restrictions. *Journal of Finance 78*(1), 63–103.

Frandsen, B., L. Lefgren, and E. Leslie (2023). Judging judge fixed effects. *American Economic Review 113*(1), 253–277.

Frandsen, B. R., L. J. Lefgren, E. C. Leslie, and S. P. McIntyre (2026). The incredible shrinking instruments: Using empirical Bayes to increase efficiency in IV designs with many instruments. Working Paper 34634, National Bureau of Economic Research.

Goldsmith-Pinkham, P., P. Hull, and M. Kolesár (2026). Leniency designs: An operator's manual. *Journal of Economic Perspectives 40*(3), 213–240.

Hart, O., A. Shleifer, and R. W. Vishny (1997). The proper scope of government: Theory and an application to prisons. *Quarterly Journal of Economics 112*(4), 1127–1161.

Hill, A. (2015). Does delegation undermine accountability? experimental evidence on the relationship between blame shifting and control. *Journal of Empirical Legal Studies 12*(2), 311–339.

Holmström, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization 7*(Special Issue), 24–52.

Holmström, B. and P. Milgrom (1994). The firm as an incentive system. *American Economic Review 84*(4), 972–991.

Hunt, R. M. (2007). Collecting consumer debt in America. *Federal Reserve Bank of Philadelphia Business Review* (Q2), 11–24.

Imbens, G. W. and C. F. Manski (2004). Confidence intervals for partially identified parameters. *Econometrica 72*(6), 1845–1857.

Jha, M., J. Qian, M. Weber, and B. Yang (2024). ChatGPT and corporate policies. Working Paper 32161, National Bureau of Economic Research.

Karlan, D., M. Mobius, T. Rosenblat, and A. Szeidl (2009). Trust and social collateral. *Quarterly Journal of Economics 124*(3), 1307–1361.

Kling, J. R. (2006). Incarceration length, employment, and earnings. *American Economic Review 96*(3), 863–876.

Kolesár, M. (2013, November). Estimation in an instrumental variables model with treatment effect heterogeneity. Working paper, version 1.3.1, Cowles Foundation, Yale University.

La Ferrara, E. (2003). Kin groups and reciprocity: A model of credit transactions in Ghana. *American Economic Review 93*(5), 1730–1751.

Lafontaine, F. and M. Slade (2007). Vertical integration and firm boundaries: The evidence. *Journal of Economic Literature 45*(3), 629–685.

Laudenbach, C. and S. Siegel (2025). Personal communication in an automated world: Evidence from loan repayments. *Journal of Finance 80*(1), 515–559.

Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies 76*(3), 1071–1102.

Lee, S. and P. Persson (2016). Financing from family and friends. *Review of Financial Studies 29*(9), 2341–2386.

Liao, C., P. Du, Y. Yang, and Z. Huang (2021). Carrots or sticks in debt collection services? A voice metrics and text analysis of debt collection calls. *Journal of Service Theory and Practice 31*(6), 950–973.

Liao, L., Z. Wang, H. Yan, J. Yang, and C. Zhou (2026). Angry borrowers: Ex post effects of social shaming on debt repayment. *Management Science 72*(4), 3061–3080.

Ludwig, J., S. Mullainathan, and A. Rambachan (2025). Large language models: An applied econometric framework. Working Paper 33344, National Bureau of Economic Research. Revised December 2025.

Oexl, R. and Z. J. Grossman (2013). Shifting the blame to a powerless intermediary. *Experimental Economics 16*(3), 306–312.

Pei, H. D. (2015). Machiavellian delegation. Working paper, Massachusetts Institute of Technology.

Perez-Truglia, R. and U. Troiano (2018). Shaming tax delinquents. *Journal of Public Economics 167*, 120–137.

Phillips, G. D. A. and C. Hale (1977). The bias of instrumental variable estimators of simultaneous equation systems. *International Economic Review 18*(1), 219–228.

Romeo, C. and R. Sandler (2021). The effect of debt collection laws on access to credit. *Journal of Public Economics 195*, 104320.

Wilf-Townsend, D. (2022). Assembly-line plaintiffs. *Harvard Law Review 135*(7), 1704–1778.

Zhou, Y. (2024). Artificial intelligence and debt collection: Evidence from a field experiment. SSRN Working Paper 4905228, revised January 2025.

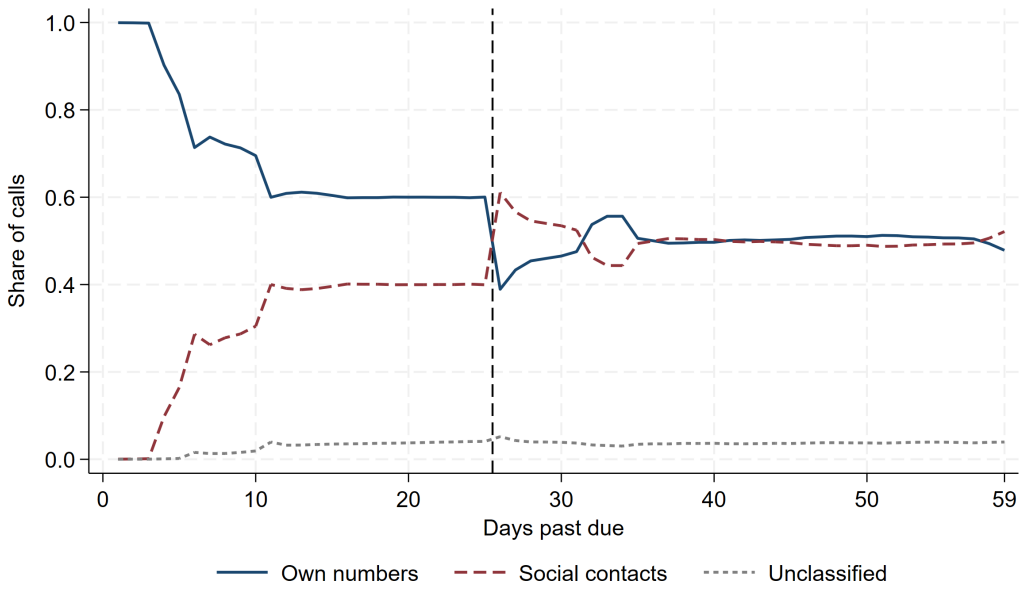



Figure 1: Call composition over days past due

*Notes*: The figure plots, for each day past due, the share of collection calls dialed to the borrower's own phone numbers and to social contact numbers (all numbers other than the borrower's own), pooled across all case-days of borrowers whose debt remains open on day 26 and weighted by the number of calls. The small unclassified share (numbers whose relationship with the borrower cannot be determined, included in the social-contact line) is also shown separately. The dashed line marks day 26, when debts are assigned to in-house and third-party collectors.

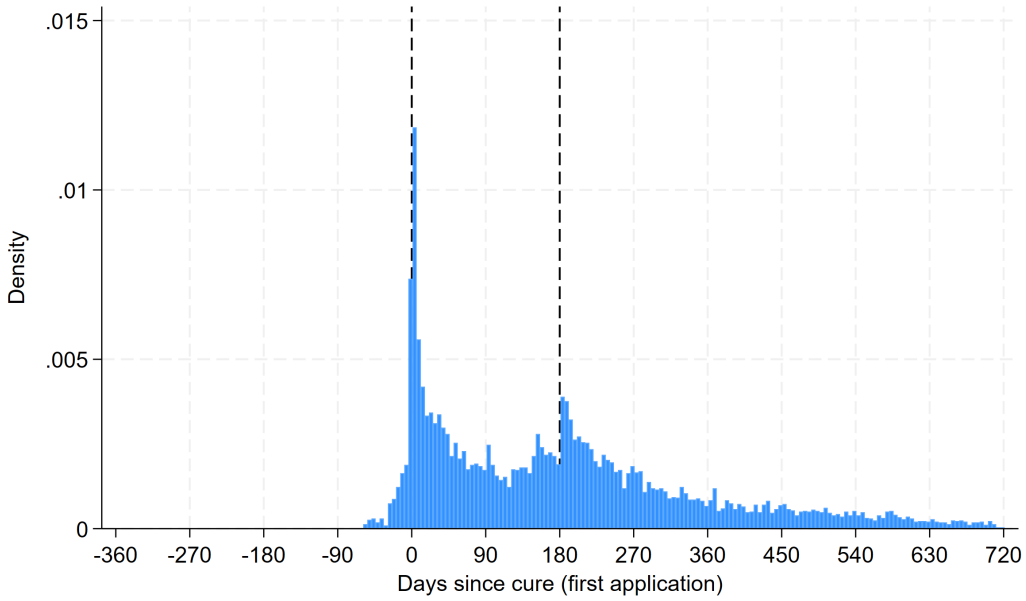



Figure 2: Timing of the first return application among cured borrowers

*Notes*: The histogram shows, for each borrower cured within 60 days of the day-26 assignment, the timing of their first return loan application, measured in days since the overdue balance closing date (5-day bins). The dashed lines mark the cure day (0) and the 180-day relaxation of credit access restrictions.

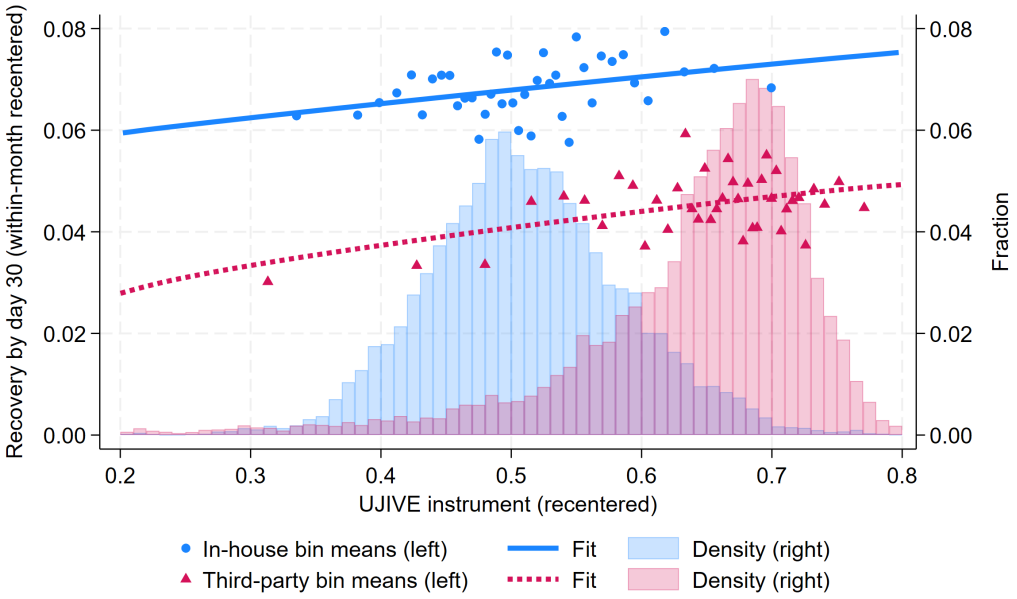
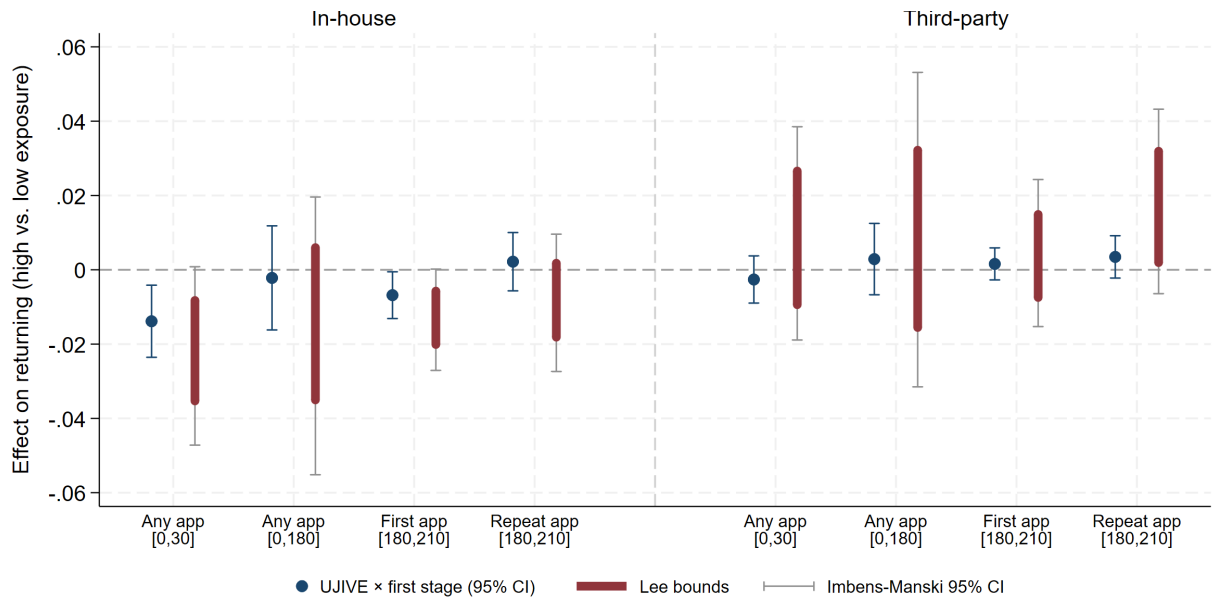



Figure 3: Reduced-form effect of social-contact calling on recovery

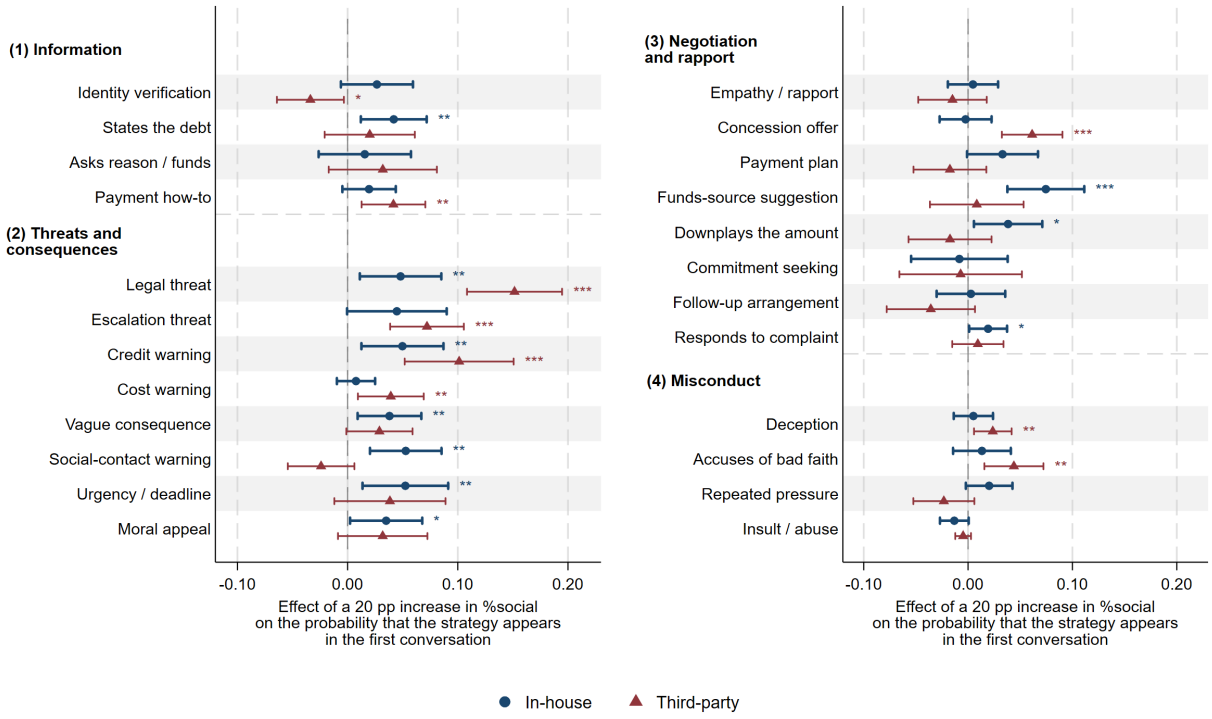
*Notes*: The figure plots the reduced-form relationship between the cumulative recovered NPV by day 30, scaled by the overdue balance at the day-26 assignment, and the social-contact calling propensity measured by the UJIVE instrument in Section 3.1, separately for the in-house and third-party samples. Recovery is recentered within sector by calendar month with each sector's sample mean added back; the instrument is recentered within sector by calendar month and shifted by the sector's sample mean of $\%social$. Markers show mean recovery within 40 quantile bins of the instrument, and lines are local-linear fits. The recentered instrument is trimmed to $[0.2, 0.8]$ for illustrative purposes; its distributions in the two samples are shown as histograms (right axis, fraction).

Figure 4: Lee bounds on the always-cured borrowers' returning effects



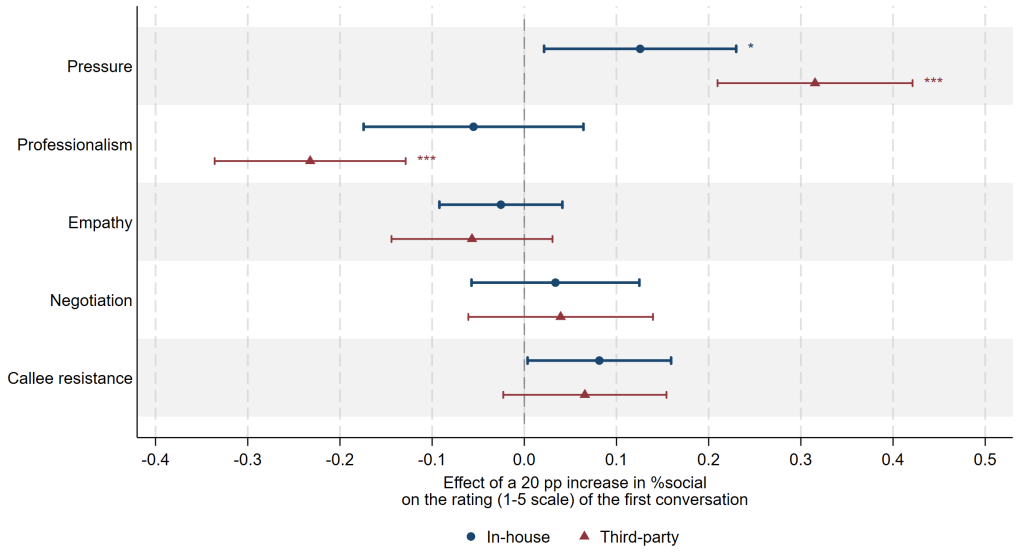

*Notes*: The figure reports Lee (2009) bounds on the effect of collection style on the returning of *always-cured* borrowers, who cure within 60 days regardless of the style they experience. Style exposure is binarized as $Z = 1$ if the assigned caller-month leave-one-out instrument of Section 3.1 exceeds its calendar-month median; the outcomes are the return indicators of Table 3. The thick bars show the Lee bounds on the effect of high vs. low $Z$ on the outcome; the thin capped spikes show the Imbens–Manski 95% confidence intervals based on caller-clustered bootstrap standard errors. The dots show the corresponding UJIVE estimates of Table 3 multiplied by the first stage of $Z$ on $\%social$, with their 95% confidence intervals based on the standard errors robust to heteroskedasticity and treatment effect heterogeneity, as in Table 3.

Figure 5: Social-contact calling and collection strategies in the first conversation with the borrower



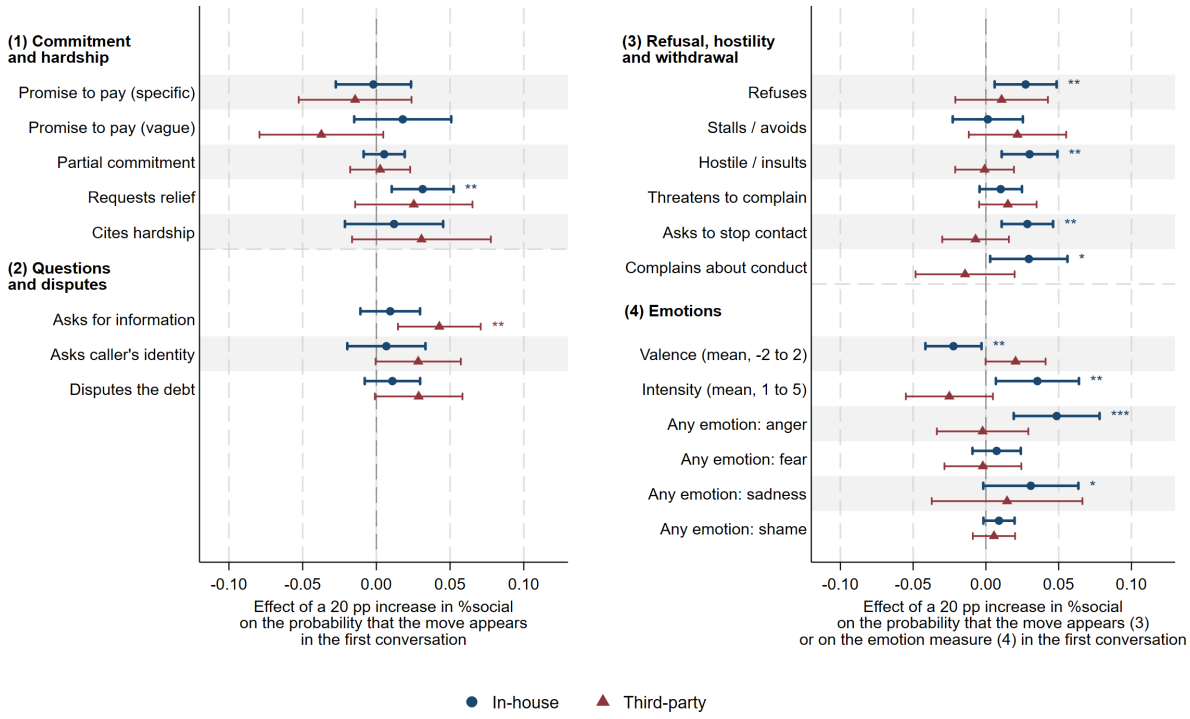

*Notes*: Each row reports, separately for in-house and third-party callers, the UJIVE coefficient on %*social* from a separate regression of a collection-strategy indicator on the instrumented treatment as in the headline specifications, rescaled to a 20-percentage-point (roughly one standard deviation in both sectors) increase in %*social*. The outcome is an indicator for whether the strategy appears in the borrower's first answered conversation with the caller on the borrower's own phone number during days 26–30, as labeled by the LLM annotation of the call transcripts. Capped segments show 95% confidence intervals based on standard errors clustered at the collector level. The $p$-values are Benjamini and Hochberg (1995) false-discovery-rate adjusted $p$-values, computed within sector and strategy family. Intimidation is omitted because it does not appear in the third-party sample and, therefore, is not estimable.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 6: Social-contact calling and the ratings of the first conversation with the borrower




*Notes*: Each row reports, separately for in-house and third-party callers, the UJIVE coefficient on %*social* from a separate regression of a call rating on the instrumented treatment, with the same specification, sample definition, and rescaling to a 20-percentage-point increase in %*social* as in Figure 5. The outcomes are the LLM-assigned ratings of the borrower's first answered conversation with the caller on the borrower's own phone number during days 26–30, on a 1–5 scale: the caller's pressure, professionalism, empathy, and negotiation effort, and the borrower's resistance. Capped segments show 95% confidence intervals based on standard errors clustered at the collector level. The $p$-values are Benjamini and Hochberg (1995) false-discovery-rate adjusted $p$-values, computed within sector across the five ratings.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 7: Social-contact calling and the borrower's reactions in the first conversation




*Notes*: Each row reports, separately for in-house and third-party callers, the UJIVE coefficient on %*social* from a separate regression of the row's outcome on the instrumented treatment, with the same specification, sample definition, and rescaling to a 20-percentage-point increase in %*social* as in Figure 5. The outcomes are measured on the borrower's first answered conversation with the caller on the borrower's own phone number during days 26–30, as labeled by the LLM annotation of the call transcripts: in groups (1)–(3), indicators for whether the borrower move appears in the conversation; in group (4), the mean valence ($-2$ to 2) and intensity (1 to 5) of the borrower's turns and indicators for whether any of the borrower's turns expresses anger, fear, sadness, or shame, defined for the conversations in which the borrower speaks. Capped segments show 95% confidence intervals based on standard errors clustered at the collector level. The $p$-values are Benjamini and Hochberg (1995) false-discovery-rate adjusted $p$-values, computed within sector separately for the borrower moves and the emotion measures.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1: Summary statistics

| | In-house | | | | | | Third-party | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | P10 | Median | P90 | *N* | Mean | SD | P10 | Median | P90 | *N* |
| *Panel A: Borrower and loan characteristics* | | | | | | | | | | | | |
| Overdue balance at assignment (yuan) | 2,589 | 2,592 | 815 | 1,933 | 4,779 | 66,664 | 2,565 | 2,608 | 809 | 1,924 | 4,711 | 60,271 |
| Remaining principal (yuan) | 19,248 | 11,673 | 9,134 | 15,578 | 33,500 | 66,664 | 19,179 | 11,615 | 9,122 | 15,501 | 33,383 | 60,271 |
| Repaid in days 1–25 (share of balance) | 0.017 | 0.064 | 0.000 | 0.000 | 0.033 | 66,664 | 0.017 | 0.064 | 0.000 | 0.000 | 0.035 | 60,271 |
| Credit-score decile (1 = highest risk) | 5.0 | 2.7 | 1.0 | 5.0 | 9.0 | 66,664 | 5.0 | 2.7 | 1.0 | 5.0 | 9.0 | 60,271 |
| Age | 31.1 | 6.9 | 23.0 | 30.0 | 41.0 | 66,664 | 31.1 | 6.9 | 23.0 | 30.0 | 41.0 | 60,271 |
| Male | 0.696 | 0.460 | 0.000 | 1.000 | 1.000 | 66,664 | 0.694 | 0.461 | 0.000 | 1.000 | 1.000 | 60,271 |
| Bachelor's degree or above | 0.095 | 0.293 | 0.000 | 0.000 | 0.000 | 66,664 | 0.095 | 0.294 | 0.000 | 0.000 | 0.000 | 60,271 |
| Currently a college student | 0.051 | 0.219 | 0.000 | 0.000 | 0.000 | 66,664 | 0.047 | 0.212 | 0.000 | 0.000 | 0.000 | 60,271 |
| Emergency contacts on file (#) | 2.7 | 1.4 | 2.0 | 2.0 | 4.0 | 66,664 | 2.7 | 1.4 | 2.0 | 2.0 | 4.0 | 60,271 |
| Phone book authorized | 0.844 | 0.363 | 0.000 | 1.000 | 1.000 | 66,664 | 0.844 | 0.363 | 0.000 | 1.000 | 1.000 | 60,271 |
| *Panel B: Collector characteristics* | | | | | | | | | | | | |
| Collector level: | | | | | | | | | | | | |
| Age | 26.7 | 3.8 | 22.2 | 26.0 | 32.0 | 482 | 26.4 | 5.3 | 20.0 | 25.0 | 34.2 | 340 |
| Male | 0.649 | 0.478 | 0.000 | 1.000 | 1.000 | 482 | 0.497 | 0.501 | 0.000 | 0.000 | 1.000 | 340 |
| Tenure (months) | 16.9 | 14.9 | 3.5 | 11.6 | 37.0 | 482 | 13.8 | 7.7 | 4.5 | 14.0 | 24.0 | 340 |
| Cases handled (total, #) | 138 | 133 | 21 | 91 | 334 | 482 | 177 | 176 | 20 | 118 | 505 | 340 |
| Collector-month level: | | | | | | | | | | | | |
| Cases handled per month (#) | 23.5 | 6.2 | 16.0 | 23.0 | 32.0 | 2,838 | 24.4 | 7.0 | 16.0 | 24.0 | 34.0 | 2,466 |
| Calls placed per month (days 26–30, #) | 860 | 319 | 483 | 825 | 1,300 | 2,838 | 173 | 73 | 86 | 161 | 271 | 2,466 |
| Working days per month (#) | 29.9 | 2.2 | 28.0 | 31.0 | 31.0 | 2,838 | 29.8 | 2.1 | 28.0 | 30.0 | 31.0 | 2,466 |
| Month level: | | | | | | | | | | | | |
| Cases per calendar month (#) | 2,469 | 1,206 | 889 | 2,793 | 4,193 | 27 | 2,232 | 1,121 | 1,165 | 1,805 | 4,165 | 27 |
| Active collectors per calendar month (#) | 105.1 | 37.5 | 48.0 | 118.0 | 143.0 | 27 | 91.3 | 31.2 | 62.0 | 79.0 | 139.0 | 27 |

Table 1: Summary statistics (continued)

| | In-house | | | | | | Third-party | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | P10 | Median | P90 | *N* | Mean | SD | P10 | Median | P90 | *N* |
| *Panel C: Collection intensity and outcomes* | | | | | | | | | | | | |
| % social-contact calls (days 26–30) | 0.511 | 0.191 | 0.250 | 0.542 | 0.725 | 66,664 | 0.638 | 0.212 | 0.375 | 0.667 | 0.833 | 60,271 |
| Calls placed (days 26–30, #) | 36.6 | 20.5 | 12.0 | 34.0 | 64.0 | 66,664 | 7.1 | 4.0 | 3.0 | 6.0 | 12.0 | 60,271 |
| Answer rate (days 26–30) | 0.159 | 0.147 | 0.000 | 0.125 | 0.333 | 66,664 | 0.106 | 0.166 | 0.000 | 0.000 | 0.333 | 60,271 |
| Talk time (days 26–30, minutes) | 5.6 | 7.3 | 0.0 | 3.4 | 13.7 | 66,664 | 1.1 | 2.9 | 0.0 | 0.0 | 3.3 | 60,271 |
| Recovery rate by day 30 | 0.068 | 0.240 | 0.000 | 0.000 | 0.000 | 66,664 | 0.045 | 0.197 | 0.000 | 0.000 | 0.000 | 60,271 |
| Recovery rate by day 59 | 0.137 | 0.325 | 0.000 | 0.000 | 0.930 | 66,664 | 0.130 | 0.316 | 0.000 | 0.000 | 0.908 | 60,271 |
| Cured within 60 days of assignment | 0.154 | 0.361 | 0.000 | 0.000 | 1.000 | 66,664 | 0.140 | 0.347 | 0.000 | 0.000 | 1.000 | 60,271 |
| Returned within [0,30] of cure | 0.069 | 0.253 | 0.000 | 0.000 | 0.000 | 10,257 | 0.066 | 0.248 | 0.000 | 0.000 | 0.000 | 8,439 |
| Returned within [0,180] of cure | 0.175 | 0.380 | 0.000 | 0.000 | 1.000 | 10,257 | 0.187 | 0.390 | 0.000 | 0.000 | 1.000 | 8,439 |
| First-time return in [180,210] | 0.034 | 0.181 | 0.000 | 0.000 | 0.000 | 10,257 | 0.040 | 0.196 | 0.000 | 0.000 | 0.000 | 8,439 |
| Repeat return in [180,210] | 0.049 | 0.217 | 0.000 | 0.000 | 0.000 | 10,257 | 0.057 | 0.232 | 0.000 | 0.000 | 0.000 | 8,439 |

*Notes*: Panel B: collector-level rows report one observation per collector serving the working sample, with age and tenure averaged over the collector's active months; tenure is months since the collector first started working for the lender. Panel C: the recovery rates are cumulative collections through day 30 and day 59, respectively, discounted at 24% per year, as shares of the assignment balance; the four returning rows are computed on cases cured within 60 days of assignment, with windows in days since the cure date and first-time vs. repeat referring to whether the borrower's first-ever return application falls in $[180, 210]$.

Table 2: Social-contact calling and recovery

| | (1) OLS | (2) UJIVE | (3) OLS | (4) UJIVE | (5) UJIVE Emergency contacts 0–2 | (6) UJIVE Emergency contacts 3+ |
|---|---|---|---|---|---|---|
| Sample | Full | | | | | |
| *Panel A: in-house* | | | | | | |
| %*social* | 0.063*** | 0.034** | 0.060*** | 0.034** | 0.021 | 0.060** |
| | (12.96) | (2.04) | (12.20) | (2.08) | (1.10) | (1.97) |
| Mean dep. var. | 0.068 | 0.068 | 0.068 | 0.068 | 0.056 | 0.097 |
| Covariates | No | No | Yes | Yes | Yes | Yes |
| Observations | 66,664 | 66,664 | 66,664 | 66,664 | 47,387 | 19,277 |
| *Panel B: third-party* | | | | | | |
| %*social* | 0.020*** | 0.033*** | 0.016*** | 0.033*** | 0.020** | 0.060*** |
| | (5.90) | (3.74) | (4.87) | (3.75) | (2.04) | (3.54) |
| Mean dep. var. | 0.045 | 0.045 | 0.045 | 0.045 | 0.035 | 0.069 |
| Covariates | No | No | Yes | Yes | Yes | Yes |
| Observations | 60,271 | 60,271 | 60,271 | 60,271 | 42,837 | 17,434 |
| *Panel C: third-party − in-house* | | | | | | |
| %*social* | -0.043*** | -0.001 | -0.043*** | -0.002 | -0.001 | 0.000 |
| | (-7.26) | (-0.06) | (-7.26) | (-0.09) | (-0.04) | (0.01) |
| Covariates | No | No | Yes | Yes | Yes | Yes |
| Observations | 126,935 | 126,935 | 126,935 | 126,935 | 90,224 | 36,711 |

*Notes*: The table reports regressions of the cumulative recovered NPV by day 30, scaled by the overdue balance at the day-26 assignment, on %*social* (days 26–30). Calendar-month fixed effects are included in all columns. Columns (3)–(6) additionally control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status. UJIVE columns instrument %*social* with the UJIVE instrument of Section 3.1. Columns (5)–(6) split the sample by the number of borrowers' emergency contacts. Mean dep. var. reports the mean of the dependent variable on the column's estimation sample. Panel C reports the third-party minus in-house coefficient difference, with the $z$-statistic of an independent-samples test based on the two panels' standard errors in parentheses. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Social-contact calling and borrower return propensity

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | Any app [0,30] | | Any app [0,180] | | First app [180,210] | | Repeat app [180,210] | |
| | OLS | UJIVE | OLS | UJIVE | OLS | UJIVE | OLS | UJIVE |
| *Panel A: in-house* | | | | | | | | |
| %*social* | 0.037*** | -0.155*** | 0.093*** | -0.024 | 0.011 | -0.076** | 0.025** | 0.024 |
| | (2.71) | (-2.79) | (4.58) | (-0.30) | (1.12) | (-2.13) | (2.18) | (0.55) |
| Mean dep. var. | 0.069 | 0.069 | 0.175 | 0.175 | 0.034 | 0.034 | 0.049 | 0.049 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 |
| *Panel B: third-party* | | | | | | | | |
| %*social* | 0.004 | -0.026 | 0.023 | 0.028 | 0.024** | 0.016 | 0.023* | 0.034 |
| | (0.29) | (-0.80) | (1.07) | (0.59) | (2.45) | (0.72) | (1.96) | (1.19) |
| Mean dep. var. | 0.066 | 0.066 | 0.187 | 0.187 | 0.040 | 0.040 | 0.057 | 0.057 |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 |
| *Panel C: third-party − in-house* | | | | | | | | |
| %*social* | -0.033* | 0.129** | -0.069** | 0.053 | 0.014 | 0.092** | -0.002 | 0.010 |
| | (-1.68) | (2.02) | (-2.33) | (0.57) | (1.01) | (2.19) | (-0.11) | (0.19) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 18,696 | 18,696 | 18,696 | 18,696 | 18,696 | 18,696 | 18,696 | 18,696 |

*Notes*: The table reports regressions of return-borrowing indicators on %*social* (days 26–30), on the sample of borrowers cured within 60 days of the day-26 assignment. Outcomes in columns (1)–(4) indicate any return loan application within the stated window of days since the overdue balance closing date (the cure day). Columns (5)–(6) indicate that the borrower's *first-ever* return application falls in days [180,210] since cure, while columns (7)–(8) indicate an application in [180,210] by a borrower who had already applied before day 180. All regressions control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status, with calendar-month fixed effects. UJIVE columns instrument %*social* with the UJIVE instrument of Section 3.1. Mean dep. var. reports the mean of the dependent variable on the column's estimation sample. Panel C reports the third-party minus in-house coefficient difference, with the $z$-statistic of an independent-samples test based on the two panels' standard errors in parentheses. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Social-contact calling and the repayment performance of returning loans

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | In-house | | | Third-party | | | Third-party − in-house | | |
| Installment horizon: | 1m | 3m | All | 1m | 3m | All | 1m | 3m | All |
| *Panel A: loans originated in [0,180] since cure* | | | | | | | | | |
| Bad-loan rate (30+ DPD) | 0.937** | 0.803 | 0.225 | -0.659** | -0.523* | -0.734** | -1.595*** | -1.326** | -0.959 |
| | (2.01) | (1.54) | (0.40) | (-2.44) | (-1.82) | (-2.28) | (-2.97) | (-2.23) | (-1.48) |
| Mean dep. var. | 0.244 | 0.349 | 0.461 | 0.203 | 0.288 | 0.415 | | | |
| Payment fulfillment | -1.285** | -0.936** | -0.727 | 0.347 | 0.333 | 0.401 | 1.633** | 1.269** | 1.128** |
| | (-2.27) | (-2.04) | (-1.57) | (1.08) | (1.29) | (1.44) | (2.51) | (2.41) | (2.09) |
| Mean dep. var. | 0.603 | 0.667 | 0.652 | 0.606 | 0.703 | 0.690 | | | |
| ROA | -0.510 | -0.522 | -0.524 | 0.225 | 0.214 | 0.227 | 0.736* | 0.736* | 0.751* |
| | (-1.26) | (-1.34) | (-1.34) | (1.19) | (1.12) | (1.15) | (1.65) | (1.70) | (1.72) |
| Mean dep. var. | -0.175 | -0.155 | -0.168 | -0.167 | -0.150 | -0.159 | | | |
| Observations | 295 | 295 | 295 | 248 | 248 | 248 | 543 | 543 | 543 |
| Return loans | 639 | 639 | 639 | 691 | 691 | 691 | 1,330 | 1,330 | 1,330 |
| *Panel B: first application in [180,210]* | | | | | | | | | |
| Bad-loan rate (30+ DPD) | 0.314 | 0.038 | 0.044 | -0.076 | -0.030 | -0.063 | -0.390 | -0.068 | -0.107 |
| | (1.05) | (0.08) | (0.05) | (-0.33) | (-0.10) | (-0.07) | (-1.04) | (-0.12) | (-0.09) |
| Mean dep. var. | 0.029 | 0.082 | 0.232 | 0.023 | 0.040 | 0.213 | | | |
| Payment fulfillment | -0.970 | -0.336 | -0.253 | 0.540 | 0.158 | -0.227 | 1.510 | 0.494 | 0.026 |
| | (-1.21) | (-0.70) | (-0.43) | (0.86) | (0.48) | (-0.58) | (1.49) | (0.86) | (0.04) |
| Mean dep. var. | 0.824 | 0.886 | 0.849 | 0.867 | 0.935 | 0.879 | | | |
| ROA | -0.720 | -0.685 | -0.697 | 0.138 | 0.068 | 0.027 | 0.858 | 0.753 | 0.724 |
| | (-1.30) | (-1.24) | (-1.25) | (0.50) | (0.58) | (0.23) | (1.39) | (1.34) | (1.27) |
| Mean dep. var. | -0.033 | -0.025 | -0.029 | -0.003 | 0.010 | 0.006 | | | |
| Observations | 115 | 115 | 115 | 132 | 132 | 132 | 247 | 247 | 247 |
| Return loans | 310 | 310 | 310 | 424 | 424 | 424 | 734 | 734 | 734 |

*Notes*: The table reports UJIVE regressions of the performance of return loans on %*social* (days 26–30), on the sample of borrowers cured within 60 days of the day-26 assignment. Each row is a performance measure, aggregated over loans originated to returning borrowers in the panel's window and evaluated at horizons of one and three months of scheduled installments or over the full loan term ("All"). *Bad-loan rate*: the share of originated principal in loans whose any installment scheduled within the horizon ever becomes delinquent for at least 30 days. *Payment fulfillment*: the share of scheduled installment payments actually collected by the horizon-end due date. *ROA*: the payments actually collected within the horizon, minus the scheduled principal in the same horizon, per yuan-month of book outstanding principal balance cumulated over the horizon, where the book balance is the originated principal net of scheduled principal repayments. All regressions control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status, with calendar-month fixed effects, and instrument %*social* with the UJIVE instrument of Section 3.1. Mean dep. var. reports the mean of the row's dependent variable on the column group's estimation sample. Return loans reports the number of originated return loans aggregated into the panel's borrower-level outcomes. Columns (7)–(9) report the third-party minus in-house coefficient difference, with the $z$-statistic of an independent-samples test based on the two sectors' standard errors in parentheses. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Social-contact calling and cured borrowers' applications to other lenders

| | (1) Any inquiry | (2) Inquiry day | (3) | (4) Number of applications within 7 days before inquiry | (5) | (6) |
|---|---|---|---|---|---|---|
| Sample: | All cured borrowers | | | | Non-returners | |
| Inquiries in: (days since cure) | 30d | 30d | 30d | 180d | 30d | 180d |
| *Panel A: in-house* | | | | | | |
| %*social* | -0.034 | 1.002 | 0.718** | 0.159 | 0.792** | 0.310 |
| | (-0.50) | (0.20) | (2.21) | (0.61) | (2.32) | (1.03) |
| Mean dep. var. | 0.130 | 9.0 | 1.195 | 1.203 | 1.199 | 1.192 |
| Timing bins | No | No | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 10,257 | 1,329 | 1,329 | 2,389 | 1,216 | 1,760 |
| *Panel B: third-party* | | | | | | |
| %*social* | -0.017 | -0.146 | -0.533 | -0.303 | -0.663 | -0.158 |
| | (-0.46) | (-0.04) | (-1.19) | (-1.30) | (-1.33) | (-0.53) |
| Mean dep. var. | 0.130 | 9.1 | 1.173 | 1.218 | 1.171 | 1.192 |
| Timing bins | No | No | Yes | Yes | Yes | Yes |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 8,439 | 1,096 | 1,096 | 1,968 | 1,010 | 1,422 |
| *Panel C: third-party − in-house* | | | | | | |
| %*social* | 0.017 | -1.147 | -1.251** | -0.462 | -1.454** | -0.468 |
| | (0.22) | (-0.18) | (-2.27) | (-1.32) | (-2.41) | (-1.11) |
| Covariates | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 18,696 | 2,425 | 2,425 | 4,357 | 2,226 | 3,182 |

*Notes*: The table reports UJIVE regressions on %*social* (days 26–30), on the sample of borrowers cured within 60 days of the day-26 assignment. The outcomes are built from the lender's inquiries to a credit aggregator about its borrowers. Only inquiries dated within the stated window of days since cure are used. Column (1) uses an indicator of at least one inquiry; column (2) is the mean inquiry day since cure among borrowers with at least one inquiry; columns (3)–(6) are the application count averaged over the borrower's inquiries in the window, among borrowers with at least one inquiry. Columns (1)–(4) use all cured borrowers; columns (5)–(6) use borrowers with no return loan application from cure through the end of the window. Timing bins control for when the inquiries occur in the window, as the shares of the borrower's inquiries in 5-day bins (0–30 days) or 30-day bins (0–180 days). All regressions control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status, with calendar-month fixed effects, and instrument %*social* with the UJIVE instrument of Section 3.1. Mean dep. var. reports the mean of the dependent variable on the column's estimation sample. Panel C reports the third-party minus in-house coefficient difference, with the $z$-statistic of an independent-samples test based on the two panels' standard errors in parentheses. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Social-contact calling and returning, by loan-size threshold

| Amount cutoff (yuan): | (1) >0 | (2) >200 | (3) >500 | (4) >1,000 | (5) >2,000 | (6) >5,000 | (7) >8,000 |
|---|---|---|---|---|---|---|---|
| *Panel A: in-house, any application in [0,30]* | | | | | | | |
| %*social* | -0.155*** | -0.148*** | -0.132*** | -0.126*** | -0.113*** | -0.055 | -0.035 |
| | (-2.79) | (-2.88) | (-2.64) | (-2.63) | (-2.61) | (-1.57) | (-1.19) |
| Base rate | 0.069 | 0.058 | 0.052 | 0.047 | 0.040 | 0.025 | 0.018 |
| Coef. / base rate | -2.2 | -2.6 | -2.5 | -2.7 | -2.8 | -2.2 | -2.0 |
| Observations | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 |
| *Panel B: in-house, first application in [180,210]* | | | | | | | |
| %*social* | -0.076** | -0.073** | -0.068** | -0.044 | -0.032 | -0.021 | -0.016 |
| | (-2.13) | (-2.15) | (-2.01) | (-1.37) | (-1.07) | (-0.80) | (-0.67) |
| Base rate | 0.034 | 0.031 | 0.029 | 0.028 | 0.025 | 0.018 | 0.014 |
| Coef. / base rate | -2.3 | -2.4 | -2.3 | -1.6 | -1.3 | -1.2 | -1.1 |
| Observations | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 | 10,257 |
| *Panel C: third-party, any application in [0,30]* | | | | | | | |
| %*social* | -0.026 | -0.027 | -0.023 | -0.031 | -0.023 | -0.038* | -0.039** |
| | (-0.80) | (-0.90) | (-0.80) | (-1.12) | (-0.92) | (-1.81) | (-2.23) |
| Base rate | 0.066 | 0.055 | 0.050 | 0.046 | 0.038 | 0.025 | 0.017 |
| Coef. / base rate | -0.4 | -0.5 | -0.5 | -0.7 | -0.6 | -1.5 | -2.3 |
| Observations | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 |
| *Panel D: third-party, first application in [180,210]* | | | | | | | |
| %*social* | 0.016 | 0.021 | 0.027 | 0.024 | 0.023 | 0.015 | 0.008 |
| | (0.72) | (1.04) | (1.37) | (1.27) | (1.27) | (0.89) | (0.50) |
| Base rate | 0.040 | 0.038 | 0.036 | 0.035 | 0.031 | 0.026 | 0.021 |
| Coef. / base rate | 0.4 | 0.6 | 0.7 | 0.7 | 0.7 | 0.6 | 0.4 |
| Observations | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 | 8,439 |

*Notes*: The table reports UJIVE regressions of loan-size threshold indicators on %*social* (days 26–30), on the sample of borrowers cured within 60 days of the day-26 assignment. The outcome in each column is the unconditional indicator that the borrower's cumulative applied return-loan amount in the stated window since the overdue balance closing date exceeds the column's cutoff. In Panels A and C the window is days [0,30] since cure; in Panels B and D the indicator additionally requires the borrower's *first-ever* return application to fall in days [180,210] since cure, with the amount cumulated over that window. All regressions control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status, with calendar-month fixed effects, and instrument %*social* with the UJIVE instrument of Section 3.1. Base rate reports the mean of the dependent variable on the estimation sample, and Coef. / base rate the proportional effect. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Social-contact calling and other collection efforts

| | In-house | | | | Third-party | | | |
|---|---|---|---|---|---|---|---|---|
| | Coef. | Mean dep. var. | % of mean | Obs. | Coef. | Mean dep. var. | % of mean | Obs. |
| (1) Calls per day | 0.060 | 7.670 | 0.8 | 66,664 | 0.296*** | 1.485 | 19.9 | 60,271 |
| | (0.30) | | | | (13.50) | | | |
| (2) Day coverage | -0.052*** | 0.817 | -6.3 | 66,664 | -0.052*** | 0.353 | -14.9 | 60,271 |
| | (-7.35) | | | | (-8.85) | | | |
| (3) Repeat dial gap (min) | -17.481*** | 118.098 | -14.8 | 65,556 | -22.720*** | 78.905 | -28.8 | 16,941 |
| | (-4.85) | | | | (-2.79) | | | |
| (4) # answered | 0.747*** | 5.058 | 14.8 | 66,664 | 0.098*** | 0.708 | 13.9 | 60,271 |
| | (6.52) | | | | (4.08) | | | |
| (5) Duration (min) | 0.842*** | 5.638 | 14.9 | 66,664 | 0.115** | 1.116 | 10.3 | 60,271 |
| | (4.99) | | | | (2.26) | | | |
| (6) Duration per answer | 0.028 | 1.186 | 2.3 | 58,965 | -0.113 | 1.642 | -6.9 | 26,731 |
| | (0.83) | | | | (-1.48) | | | |
| (7) Calls per number-day | -0.381*** | 2.465 | -15.4 | 66,445 | -0.082*** | 1.102 | -7.4 | 60,081 |
| | (-9.39) | | | | (-5.16) | | | |
| (8) # self before social | -0.723*** | 1.807 | -40.0 | 64,088 | -0.219*** | 1.253 | -17.4 | 56,401 |
| | (-8.99) | | | | (-6.13) | | | |
| (9) Self follow-up | -0.111*** | 0.715 | -15.5 | 21,413 | 0.011 | 0.192 | 5.5 | 5,319 |
| | (-7.66) | | | | (0.65) | | | |
| (10) Reduction / balance | -0.002 | 0.074 | -2.3 | 6,274 | 0.007 | 0.074 | 9.3 | 3,977 |
| | (-0.24) | | | | (1.25) | | | |

*Notes*: Each row reports, separately for in-house and third-party collectors, the UJIVE coefficient on %*social* from a separate regression of the row's collection-strategy metric, measured over days 26–30, on the instrumented treatment, rescaled to a 20-percentage-point (roughly one standard deviation in both sectors) increase in %*social*, together with the mean of the dependent variable, the rescaled coefficient as a percentage of that mean, and the number of observations, with the same samples, borrower controls, and month fixed effects as the headline recovery specifications; the number of observations varies across rows with the availability of the metric. Outcome definitions: (1) Calls per day: number of calls dialed per active day (calls divided by the number of days the case remains open within the five-day window); (2) Day coverage: number of days with at least one dialed call divided by active days; (3) Repeat dial gap: mean minutes between consecutive calls to the same number within a day, for borrowers with at least one same-number multi-call day; (4) # answered: number of answered calls; (5) Duration: total call duration in minutes; (6) Duration per answer: call duration per answered call, for borrowers with at least one answered call; (7) Calls per number-day: mean number of calls dialed per phone number per dialing day; (8) # self before social: number of calls to the borrower's own number before the first social-contact attempt, for borrowers with at least one social-contact attempt; (9) Self follow-up: share of days with a borrower-answered call on which the caller dials the borrower again later that day, for borrowers with at least one such day; (10) Reduction / balance: total payment reduction granted as a fraction of the day-26 overdue balance, restricted to borrowers with a positive payment in days 26–30. $t$-statistics based on standard errors robust to heteroskedasticity, treatment effect heterogeneity, and clustering at the collector level in parentheses beneath the coefficients. Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Social-contact calling and the lender's total P&L

| Return-loan window (days): | (1) [0,30] | (2) [0,90] | (3) [0,180] | (4) [0,360] | (5) [0,720] |
|---|---|---|---|---|---|
| *Panel A: in-house* | | | | | |
| %*social* | 0.029* | 0.025 | 0.026 | 0.031 | 0.029 |
| | (1.68) | (1.33) | (1.32) | (1.32) | (0.82) |
| Mean dep. var. | 0.068 | 0.067 | 0.066 | 0.067 | 0.064 |
| Observations | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 |
| *Panel B: third-party* | | | | | |
| %*social* | 0.033*** | 0.034*** | 0.037*** | 0.040*** | 0.101*** |
| | (3.47) | (3.60) | (3.29) | (2.91) | (3.29) |
| Mean dep. var. | 0.044 | 0.044 | 0.042 | 0.043 | 0.042 |
| Observations | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 |

*Notes*: The table reports UJIVE regressions of the lender's total P&L per yuan of assigned balance on %*social* (days 26–30), on the full working samples with no conditioning on cure. The outcome is the sum of two components on a common scale: the NPV of collections recovered by day 30 of delinquency, as a share of the assigned overdue balance, plus the realized lifetime net value of the borrower's return loans originated within the stated window of days since the day-26 assignment (cash collected net of principal, discounted to the assignment date at 5% per year), scaled by the same balance; borrowers who take no return loan in the window carry a zero second term. All regressions control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status, with calendar-month fixed effects, and instrument %*social* with the UJIVE instrument of Section 3.1. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

# Online Appendix

Delegating Dirty Work:
Blame Shifting and the Boundary of the Firm

Dong Huang


| | | |
|---|---|---|
| Appendix A | Technical Details | p. 63 |
| A.1 | Construction of the jackknife instrumental variables | p. 63 |
| A.2 | Lee bounds | p. 67 |
| A.3 | LLM annotation | p. 70 |
| Appendix B | Additional Tables and Figures | p. 78 |

# Appendix A Technical Details

## A.1 Construction of the jackknife instrumental variables

This appendix derives the instruments behind the estimators compared in Appendix Tables B5 and B8. Throughout, $i$ indexes borrowers, $X_i$ is the treatment, and $c(i)$ and $t(i)$ are the assigned caller and the assignment month. Let $Z$ be the $n \times K$ matrix of caller-by-month indicators, $W$ the $n \times L$ matrix of month indicators, $Q = [Z, W]$, and $x$ and $y$ the $n$-vectors of treatments and outcomes. For a design matrix $A$, $H_A = A(A^\top A)^{-1}A^\top$ denotes the projection matrix, $M_A = I - H_A$ the residual maker, $D_A = \mathrm{diag}(H_A)$, and $h_{A,i} = (H_A)_{ii}$ the leverage of observation $i$. The structure that yields closed forms is that every caller-month cell lies within a month: the columns of $W$ are spanned by those of $Z$, so $H_Q = H_Z$. Let $n_1$ and $n_2$ denote the numbers of borrowers in borrower $i$'s caller-month cell and in borrower $i$'s month, *including* $i$ (so $n_1 = N_1 + 1$ and $n_2 = N_2 + 1$ in the notation of Equation (1)), $\overline{X}_{ct}$ and $\overline{X}_t$ the corresponding full averages of the treatment, and $\overline{X}_{ct,-i}$ and $\overline{X}_{t,-i}$ the averages that leave borrower $i$ out.

### A.1.1 UJIVE

Goldsmith-Pinkham et al. (2026) write the UJIVE instrument of Kolesár (2013) as $\hat{x} = G_U\, x$ with

$$G_U = H_{\tilde{Z}} - \mathrm{diag}\Big(\frac{h_{\tilde{Z},i}}{1 - h_{W,i} - h_{\tilde{Z},i}}\Big)\,(M_W - H_{\tilde{Z}}), \qquad \tilde{Z} = M_W Z, \tag{A.1}$$

where $\tilde{Z}$ collects the instruments residualized on the controls, and the estimator is $\hat{\beta} = \hat{x}^\top y \big/ \hat{x}^\top x$. Because the caller-month cells are nested in the months, $\hat{x}$ is exactly orthogonal to the month indicators.

The formula presented in the main text (Equation (1)) follows in two steps. First, by the Frisch–Waugh–Lovell theorem, the projection on the residualized instruments is the difference of two projections, $H_{\tilde{Z}} = H_Q - H_W$, so that $M_W - H_{\tilde{Z}} = M_Q$, $h_{\tilde{Z},i} = h_{Q,i} - h_{W,i}$, and $1 - h_{W,i} - h_{\tilde{Z},i} = 1 - h_{Q,i}$. Row $i$ of $G_U x$ is then

$$\hat{X}_i = \big[(H_Q - H_W)\,x\big]_i - \frac{h_{Q,i} - h_{W,i}}{1 - h_{Q,i}}\,\big(M_Q x\big)_i, \tag{A.2}$$

the difference between the fitted values of the full and the controls-only first stages, minus a leverage-weighted share of borrower $i$'s own first-stage residual. Second, with the nested

indicator structure, $H_Q = H_Z$ and every projection is an average: $(H_Z x)_i = \overline{X}_{ct}$ with $h_{Q,i} = 1/n_1$, $(H_W x)_i = \overline{X}_t$ with $h_{W,i} = 1/n_2$, and $(M_Q x)_i = X_i - \overline{X}_{ct}$. Substituting,

$$\hat{X}_i = (\overline{X}_{ct} - \overline{X}_t) - (X_i - \overline{X}_{ct}) \frac{1/n_1 - 1/n_2}{1 - 1/n_1}. \tag{A.3}$$

Equation (A.3) is the instrument used in every UJIVE estimate in the paper: the caller's calling propensity in the month relative to the month average, corrected by borrower $i$'s own deviation from the caller average scaled by the cell leverage net of the month leverage.

Equation (A.3) is the leave-one-out form of Equation (1). Writing the full averages in terms of the leave-one-out averages, $\overline{X}_{ct} = \overline{X}_{ct,-i} + (X_i - \overline{X}_{ct})/(n_1 - 1)$ and $\overline{X}_t = \overline{X}_{t,-i} + (X_i - \overline{X}_t)/(n_2 - 1)$, and collecting terms gives

$$\hat{X}_i = \left(1 - \frac{1}{n_2}\right)[\overline{X}_{ct,-i} - \overline{X}_{t,-i}], \tag{A.4}$$

which is Equation (1) with $n_2 = N_2 + 1$.

### A.1.2 JIVE1 and IJIVE

*JIVE1.* The jackknife instrumental variable estimator of Phillips and Hale (1977) and Angrist et al. (1999) instruments with the leave-one-out fitted value of the full first stage,

$$\hat{x}^{J1} = (I - D_Q)^{-1}(H_Q - D_Q)\, x, \qquad \hat{X}_i^{J1} = \frac{(H_Q x)_i - h_{Q,i} X_i}{1 - h_{Q,i}} = Q_i^\top \hat{\pi}_{-i}, \tag{A.5}$$

where $\hat{\pi}_{-i}$ is the first-stage coefficient vector estimated without observation $i$ (the last equality is the standard leave-one-out identity), and the controls $W$ are included as exogenous regressors in the second stage, so that the effective instrument is $M_W \hat{x}^{J1}$. With nested indicators, $(H_Q x)_i = \overline{X}_{ct}$ and $h_{Q,i} = 1/n_1$, and the jackknifed fitted value collapses to the leave-one-out cell average,

$$\hat{X}_i^{J1} = \frac{\overline{X}_{ct} - X_i/n_1}{1 - 1/n_1} = \overline{X}_{ct,-i} = \frac{1}{N_1} \sum_{j \neq i} \mathbb{1}\{c(j) = c(i),\, t(j) = t(i)\}\, X_j, \tag{A.6}$$

which is the instrument of the JIVE1 rows of Appendix Tables B5 and B8.

*IJIVE.* Ackerberg and Devereux (2009) reverse the order of the two operations. The instruments, the treatment, and the outcome are first residualized on the controls on the

full sample, $\tilde{Z} = M_W Z$, $\tilde{x} = M_W x$, and $\tilde{y} = M_W y$, and JIVE1 is then applied to the residualized system, in which no controls remain:

$$\hat{x}^{IJ} = (I - D_{\tilde{Z}})^{-1}(H_{\tilde{Z}} - D_{\tilde{Z}})\,\tilde{x}, \qquad \hat{\beta}^{IJ} = \frac{\hat{x}^{IJ\top}\tilde{y}}{\hat{x}^{IJ\top}\tilde{x}}. \tag{A.7}$$

Because $H_{\tilde{Z}} = H_Q - H_W = H_Z - H_W$ here, $(H_{\tilde{Z}}\tilde{x})_i = \overline{X}_{ct} - \overline{X}_t$ and the leverage of the residualized design is $\tilde{h}_i \equiv h_{\tilde{Z},i} = 1/n_1 - 1/n_2$, so that

$$\hat{X}_i^{IJ} = \frac{(\overline{X}_{ct} - \overline{X}_t) - \tilde{h}_i\,(X_i - \overline{X}_t)}{1 - \tilde{h}_i}, \tag{A.8}$$

the instrument of the IJIVE rows.

### A.1.3 Why the three coincide in this design

The three instruments look different but, when the caller-month cells are nested in the month fixed effects, they differ only by a within-month constant, which the fixed effects absorb, and by own-observation terms of order $1/n_2$. Let $\hat{X}_i^{\mathrm{LOO}} \equiv \overline{X}_{ct,-i} - \overline{X}_{t,-i}$ denote the bracket in (A.4), so that the UJIVE instrument is $(1 - 1/n_2)\,\hat{X}_i^{\mathrm{LOO}}$.

*JIVE1 versus UJIVE.* Within a month, the leave-one-out cell averages add up to $n_2\overline{X}_t$ (each cell contributes $n_1\overline{X}_{ct}$), so partialling the month fixed effects out of $\hat{X}_i^{J1}$ subtracts the full month average and leaves $\overline{X}_{ct,-i} - \overline{X}_t$. Using $\overline{X}_t = \overline{X}_{t,-i} + (X_i - \overline{X}_t)/(n_2 - 1)$,

$$\left(M_W \hat{x}^{J1}\right)_i = \hat{X}_i^{\mathrm{LOO}} - \frac{X_i - \overline{X}_t}{n_2 - 1}. \tag{A.9}$$

JIVE1 leaves borrower $i$ out of the cell average but not out of the month average through which the controls are partialled, so the borrower's own treatment re-enters the instrument with a factor of $-1/(n_2 - 1)$, whose magnitude is only around 0.0005 since $n_2 \approx 2000$. This is the own-observation bias through the controls discussed by Goldsmith-Pinkham et al. (2026): under homoskedasticity, the JIVE1 bias equals the 2SLS bias times $-E[F]\,L/((E[F]-1)K - L)$, where $L$ is the number of controls, $K$ the number of instruments, and $E[F]$ the expected first-stage $F$ statistic. With $L$ equal to the number of months and $K$ to the number of caller-months, the term is of the order of one over the number of borrowers per month, the statement made in Section 3.1.

*IJIVE versus UJIVE.* Since $1/n_1 - 1/n_2 = \tilde{h}_i$, Equations (A.3) and (A.8) can both be written as $\overline{X}_{ct} - \overline{X}_t$ minus $X_i - \overline{X}_{ct}$ times a coefficient, $\tilde{h}_i/(1 - 1/n_1)$ for UJIVE and

$\tilde{h}_i/(1-\tilde{h}_i)$ for IJIVE. The two coefficients differ only in whether the month leverage $1/n_2$ is netted out of the denominator:

$$\hat{X}_i^{IJ} - \hat{X}_i = (X_i - \overline{X}_{ct})\,\tilde{h}_i\,\frac{1/n_2}{(1-1/n_1)(1-\tilde{h}_i)}, \tag{A.10}$$

an own-deviation term of order $1/(n_1 n_2)$. Residualizing on the controls before rather than after the leave-one-out step does not remove the own observation from the residualization itself, which is carried out on the full sample, but the remainder is of second order: Goldsmith-Pinkham et al. (2026) show that the trace of the IJIVE operator is approximately $LK/(n-K)$ under a balanced design, compared with $-L$ for JIVE1 and zero for UJIVE.

*Consequences.* Because $n_2$ is in the thousands, the three instruments produce identical estimates to three decimals in Appendix Tables B5 and B8; the $t$-statistics differ across the rows only because the UJIVE rows use the standard errors robust to treatment effect heterogeneity. The 2SLS rows, which instrument with the in-sample fitted value $H_Q x$ and hence give borrower $i$'s own treatment the weight $1/n_1$, show the bias that all three jackknife constructions remove, the many-instrument bias toward OLS.

### A.1.4 SJIVE

The leave-one-out cell propensity is a noisy estimate of a caller's true propensity when the cell is small (its sampling variance is proportional to $1/(n_1-1)$), yet the jackknife estimators weight all cells alike. The shrunken jackknife instrumental variable estimator (SJIVE) of Frandsen et al. (2026) replaces the leave-one-out propensity by an empirical-Bayes combination of it and a lower-dimensional target,

$$\tilde{P}_i = \lambda_i\,\hat{P}_i + (1-\lambda_i)\,S_i, \qquad \hat{\beta}^{SJ} = \frac{\sum_i \tilde{P}_i\,y_i}{\sum_i \tilde{P}_i\,x_i}, \tag{A.11}$$

with the month fixed effects partialled out as before. In this implementation, which follows the authors' `sjive` command in Stata, $\hat{P}_i = \hat{X}_i^{\mathrm{LOO}}$ is the leave-one-out instrument of (A.4) and the target $S_i$ is the caller's propensity pooled across months: the leave-one-out fitted value of $x$ on caller indicators and month indicators, minus $\overline{X}_{t,-i}$. Shrinking toward $S_i$ borrows a caller's calling style in the other months to stabilize the estimate of the style in a thin month. The shrinkage weights are chosen to minimize the asymptotic variance of the estimator.

## A.2 Lee bounds

This appendix describes how the bounds and confidence intervals in Figure 4 are constructed from the data, following Lee (2009).

*Setup and the selection problem.* For each sector separately, I binarize collection-style exposure as $Z_i = 1$ if the collector–month UJIVE instrument of Section 3.1 exceeds its median within borrower $i$'s calendar month of assignment, and $Z_i = 0$ otherwise. Because assignment to callers is random within month, $Z$ acts like a randomly assigned binary treatment: being assigned a high-$Z$ caller raises the $\%social$ a borrower experiences by 9.0 percentage points in-house and 10.1 percentage points third-party (the "first stage" of $Z$, i.e., the average shift in actual exposure that the assignment induces). The return outcomes $Y_i$ of Table 3, however, are defined only for borrowers who cure within 60 days of assignment ($S_i = 1$), since uncured borrowers generally cannot apply for new loans. This creates a sample-selection problem: exposure itself moves who cures, so the cured borrowers observed under high exposure are a differently composed group than the cured borrowers observed under low exposure, and a simple comparison of their return rates mixes the causal effect of exposure with this composition difference. The clean question is therefore asked about a fixed population: the *always-cured* borrowers, who would cure under either exposure. The parameter of interest is the effect of high versus low exposure on the returning of these always-cured borrowers.

*The idea of the bounds.* Suppose that high exposure never prevents anyone from curing — one of the key assumptions, monotonicity, discussed below. Then every borrower who cures under *low* exposure would also have cured under high exposure, so the low-exposure cured pool consists of always-cured borrowers only, and its return rate directly measures the always-cured return rate without treatment. The high-exposure cured pool is the contaminated one: it contains the same always-cured borrowers *plus* the "compliers," borrowers who cure only because of the extra pressure. Although we cannot tell which individual cured borrowers are compliers, we do know how many there are: the gap between the two cure rates reveals the compliers' share, $p_0$, of the high-exposure cured pool. To compare like with like, we would want to remove exactly this share of compliers from the high-exposure pool and average the returns of the remainder. Since the compliers cannot be identified individually, we instead remove, or "trim", the most extreme candidates and ask how far the answer can move. If the removed $p_0$ share happens to be the cured borrowers with the *highest* return outcomes, the remaining pool has the lowest return rate consistent with the data; if it is those with the *lowest* outcomes, the remainder has the highest possible rate. The always-cured return rate under high exposure must lie between

these two extremes, whatever the compliers' true identity. Subtracting the low-exposure (always-cured) return rate from each extreme turns the bracket into a lower and an upper bound on the effect itself; the interval between them is the *identified set*, the range of effects consistent with the data without any assumption on who the compliers are.

*Which pool is trimmed: direction from the data.* The assumption above — exposure moves cure in one direction for every borrower ("monotonicity") — determines which cured pool is the contaminated one, and I take the direction from the data. In-house, high-$Z$ borrowers cure more often (15.5% vs. 15.2%), so the high-$Z$ cured pool is trimmed, as in the example. Third-party, the cure margin runs marginally the other way (13.8% vs. 14.2%): there the *low-$Z$* cured pool is the larger, contaminated one, and the roles of the two pools are exchanged throughout.

*Month-by-month construction.* Because assignment is random only within the assignment month, every object above is computed within calendar month $m$ (27 months per sector) and aggregated afterwards. Let $q_1(m)$ and $q_0(m)$ denote the cure rates of the high- and low-$Z$ groups in month $m$, and $r_1(m)$ and $r_0(m)$ their return rates among cured borrowers. In-house, the compliers' share of the high-$Z$ cured pool is $p_0(m) = \max\{(q_1(m) - q_0(m))/q_1(m),\, 0\}$ (mirrored third-party). In some months the raw share is negative — the sampling noise of small monthly samples can flip the cure gap against the assumed direction — and those months, carrying 47% (in-house) and 33% (third-party) of the weight, are floored at zero, i.e., treated as months with nothing to trim. The in-house cure margin is small (0.3 percentage points), so in nearly half of the months the sampling noise flips its sign, and there is nothing to trim; the bounds in those months collapse to the raw high-minus-low contrast. Because the outcome is a 0/1 indicator, the trimmed averages have closed forms: removing the highest-outcome share removes borrowers with $Y = 1$ first, so the smallest average the remainder can have is $(r_1(m) - p_0(m))/(1 - p_0(m))$; removing the lowest-outcome share removes $Y = 0$ borrowers first, giving the largest possible average $r_1(m)/(1 - p_0(m))$. The month-$m$ bounds on the always-cured effect are therefore

$$\Delta^{UB}(m) = \min\Big(\frac{r_1(m)}{1 - p_0(m)},\, 1\Big) - r_0(m), \qquad \Delta^{LB}(m) = \max\Big(\frac{r_1(m) - p_0(m)}{1 - p_0(m)},\, 0\Big) - r_0(m)$$

in-house, with the roles of the groups exchanged third-party. The month-level bounds are then averaged across months, weighting each month by the number of cured borrowers in its uncontaminated group, which approximates how the always-cured population is distributed across months. The weighted average trimmed share is only 3.6% in-house and 4.4% third-party, so the identified set is about $p_0/(1 - p_0) \approx 3.8$ and 4.6 percentage points

wide before sampling error: the same small cure margin that creates the compositional concern in the first place also keeps the bounds tight. Because the median split gives the high- and low-$Z$ groups equal shares in every month, Ferman, Siqueira, and Possebom (2026) argue that the bounds could equally be computed on the pooled sample with a single trimmed share; doing so trims only 2.0% (in-house) and 2.9% (third-party) and yields narrower identified sets with the same sign readings as Figure 4.

*Confidence intervals.* Standard errors for the two endpoints come from a nonparametric bootstrap of the entire pipeline — the within-month median cut, cure rates, trimmed shares, and weighted aggregation — with 500 replications resampling collectors, matching the clustering of the main tables; a borrower-level bootstrap stratified by calendar month yields similar results. The confidence interval must cover the *parameter*, which may sit anywhere inside the identified set, rather than the set itself: when the set is wide, each endpoint only needs to guard against being crossed from its own side (a one-sided requirement), while if the set were a single point the usual two-sided coverage would apply. The Imbens and Manski (2004) construction interpolates between these two cases. The thin spikes in Figure 4 are its 95% confidence intervals, $[\hat{\Delta}^{LB} - c\,\widehat{se}_{LB},\ \hat{\Delta}^{UB} + c\,\widehat{se}_{UB}]$, where the critical value $c$ solves

$$\Phi\Big(c + \frac{\hat{\Delta}^{UB} - \hat{\Delta}^{LB}}{\max(\widehat{se}_{LB}, \widehat{se}_{UB})}\Big) - \Phi(-c) = 0.95$$

and thus moves from the one-sided value (1.645) for a wide set toward the two-sided value (1.96) as the set narrows. A percentile variant that takes the same per-end one-sided levels from the bootstrap replicate distributions instead of the normal approximation yields nearly identical intervals.

*Reference points.* The dots in Figure 4 overlay the UJIVE estimates of Table 3. Those estimates are per unit of $\%social$, while the bounds compare high versus low $Z$; to put both on the same scale, each UJIVE estimate is multiplied by the sector's first stage of $Z$ (0.090 in-house, 0.101 third-party), with its confidence interval scaled accordingly.

*Additional references:*

Ferman, B., D. Siqueira, and V. Possebom (2026). Partial Identification under Stratified Randomization. arXiv:2601.12566.

## A.3 LLM annotation

### A.3.1 Transcript matching and reweighting

I estimate the probability that an eligible call is matched to a transcript separately within each sector with an XGBoost model (Chen and Guestrin, 2016). The inputs are the pre-treatment borrower and loan characteristics used as controls in the main regressions and a set of operation-related variables: dialing-system indicators (e.g., cell phone or landline, manual or automatic dialing), call timing (e.g., the hour, minute, and second of the call, the day of the week and of the month, and the calendar month), and call characteristics (e.g., ring seconds, the type of call recipient, and the order of the call within the day). The model achieves out-of-fold areas under the receiver operating characteristic (ROC) curve of 0.96 for in-house calls and 0.91 for third-party calls. Adding the variables sequentially shows that the operation-related variables account for all of the explanatory power, while the borrower and loan characteristics add none. Each matched call is weighted by the inverse of its estimated matching probability. The weights are capped at 30, which binds for 199 of the 228,978 matched in-house calls and for one of the 34,112 matched third-party calls.

### A.3.2 List construction and refinement

The annotation uses four lists, reported in Appendix Table B10: collector strategies (Panel A), call ratings (Panel B), borrower moves (Panel C), and emotion measures (Panel D). The strategy and move lists are built from the collection literature, collection guidelines, and an open-ended discovery pass, and are refined against human reading. The rating and the emotion lists are based on common collector evaluation metrics and the major emotion classifications, respectively. Only the wording of the grading schemes is refined.

**Candidate strategies and moves.** The initial list is built on three sources. The first is the literature of debt-collection calls (Liao et al., 2021; Laudenbach and Siegel, 2025; Choi et al., 2025). The second is the conduct rules for collectors: the company's collector guideline, the industry guideline of the National Internet Finance Association of China (2024), and the national standard for collection agencies (GB/T 45251-2025), which together describe the required procedures and the prohibited practices (e.g., threats, insults, deception, and disclosure to social contacts). The third is an open-ended discovery pass on a pilot sample of 500 randomly drawn calls. Without showing it any list, I ask an LLM to name and briefly describe each strategy the collector uses and, separately, each move the borrower makes. The 2,591 collector and 2,392 borrower descriptions are converted to

text embedding vectors and grouped by $k$-means clustering (100 and 120 clusters, chosen by the silhouette criterion). The clusters serve as a reading aid: I read each cluster and assign it to an existing code or, when none fits, to a new code. The discovery pass adds codes that the initial sources miss, such as vague consequences, accusations of bad faith, downplaying the amount, responding to complaints, and arranging follow-up calls on the collector side, and asking for the caller's identity on the borrower side.

**Ratings and emotions.** The five ratings—the caller's pressure, professionalism, empathy, and negotiation effort, and the borrower's resistance—are specified ex ante based on commonly used collector evaluation metrics, each with a written anchor for every point of the 1–5 scale. Professionalism is anchored on procedural flaws and misconduct, with 5 for a call with neither. The emotion measures follow standard emotion theories, including James A. Russell's circumplex model of emotion. Each borrower turn is scored on valence ($-2$ to 2) and intensity (or arousal, 1 to 5), and tagged with discrete emotions, including anger, fear, sadness, and shame. For each of these four emotions, a conversation-level indicator equals one if any of the borrower's turns carries it.

**Refinement.** I refine the strategy and move lists over several rounds on the pilot sample. In each round, the LLM annotates the pilot calls, I read samples of flagged and unflagged turns for each code and revise the list: merging codes that the model cannot tell apart, dropping codes that are rare or redundant, narrowing codes that the model over-applies (e.g., deception is restricted to a closed list of verifiably false claims), and adding exclusion rules at the boundaries between neighboring codes (e.g., a threat to cancel the installment plan and to pay the whole loan at once is moved from cost warnings to credit warnings, since it is the result of losing the eligibility to pay in installments). The rating anchors are revised in the same reads. The final prompt gives, for each code, a definition, the situations to which it does and does not apply, and Chinese examples. The LLM may assign several codes to one turn and must quote the text span that supports each label. Codes that only apply to calls with social contacts, such as asking the contact to relay a message, are offered only on those calls. The final lists comprise 25 collector strategies and 14 borrower moves on calls to the borrower's own number (29 and 19 including the contact-only codes).

### A.3.3 Validation

I validate the labels in three ways: agreement with a human reader on a pilot sample, the stability of the labels across repeated runs, and two internal checks on the full transcript

sample of 110,847 calls. The last way verifies whether the labels predict the behavior their names imply, and whether the three annotation families describe the same call consistently.

**Agreement with a human reader.** An author labeled turns from the 500 pilot calls, one turn and one code at a time, without seeing the model's labels. For each code, the sample contains about 25 turns the model flagged and about 25 turns it did not flag. The unflagged turns are drawn mostly from turns where a miss is likely. They include turns flagged in other annotation runs, turns given a neighboring code, and turns adjacent to a flagged turn, as well as the remaining turns at random. The benchmark comprises 2,339 judgments. Each judgment is weighted by the inverse of its sampling rate, so that precision and recall estimate the model's performance on the pilot sample rather than on the sampled turns. Recall counts misses among the aforementioned unflagged sources, where misses are likely, so the reported recall is conservative by design.

Table A1 reports the results. Across the 29 collector codes, the macro-averaged precision is 0.87, recall 0.80, and F1 0.83; across the 19 borrower codes, they are 0.82, 0.72, and 0.76. Agreement is highest for the codes that carry the main results, such as legal threats (F1 0.99), social-contact warnings and urgency (0.93), escalation threats (0.90), credit warnings (0.85), and concession offers (0.88), and for most borrower moves, such as relief requests (0.88), hardship (0.87), and requests to stop contact (0.85). A few codes agree poorly with the reader, mostly because of low recall, such as stalls or avoids (F1 0.19), empathy or rapport (0.50), asking a contact to pay (0.50), refuses (0.51), insult or abuse (0.56), and deception (0.60). Some of the codes depend on intent and context across several turns, which the LLM may find difficult to understand. Estimates based on these codes should be read with caution.

Table A1: LLM labels versus the human reader: strategy and move codes

| | Precision | Recall | F1 | | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|
| *Panel A: Collector strategies* | | | | *Panel B: Borrower moves* | | | |
| (1) Information | | | | Own-number and contact calls | | | |
| Identity verification | 1.00 | 0.91 | 0.96 | Promise to pay (specific) | 0.83 | 0.68 | 0.74 |
| States the debt | 0.97 | 0.91 | 0.94 | Promise to pay (vague) | 0.95 | 0.72 | 0.82 |
| Asks reason / funds | 0.94 | 1.00 | 0.97 | Partial commitment | 0.88 | 0.92 | 0.90 |
| Payment how-to | 0.78 | 0.87 | 0.82 | Requests relief | 0.94 | 0.82 | 0.88 |
| (2) Threats and consequences | | | | Cites hardship | 0.85 | 0.90 | 0.87 |
| Legal threat | 1.00 | 0.99 | 0.99 | Disputes the debt | 1.00 | 0.52 | 0.68 |
| Escalation threat | 0.93 | 0.88 | 0.90 | Asks for information | 0.89 | 0.62 | 0.73 |
| Credit warning | 0.93 | 0.78 | 0.85 | Asks caller's identity | 0.71 | 0.90 | 0.79 |
| Cost warning | 0.88 | 0.74 | 0.81 | Refuses | 0.62 | 0.43 | 0.51 |
| Vague consequence | 0.85 | 0.91 | 0.88 | Stalls / avoids | 0.23 | 0.17 | 0.19 |
| Social-contact warning | 0.91 | 0.96 | 0.93 | Hostile / insults | 0.80 | 0.86 | 0.83 |
| Urgency / deadline | 0.96 | 0.91 | 0.93 | Threatens to complain | 0.90 | 0.93 | 0.92 |
| Moral appeal | 0.90 | 0.95 | 0.93 | Asks to stop contact | 0.87 | 0.83 | 0.85 |
| (3) Negotiation and rapport | | | | Complains about conduct | 0.70 | 0.69 | 0.70 |
| Empathy / rapport | 0.80 | 0.36 | 0.50 | Contact calls only | | | |
| Concession offer | 0.84 | 0.91 | 0.88 | Confirms knowing the borrower | 0.80 | 0.78 | 0.79 |
| Payment plan | 0.97 | 0.95 | 0.96 | Cannot reach the borrower | 0.80 | 0.80 | 0.80 |
| Funds-source suggestion | 0.89 | 0.62 | 0.73 | Denies knowing the borrower | 1.00 | 0.88 | 0.94 |
| Downplays the amount | 0.88 | 0.94 | 0.91 | Agrees to relay | 0.92 | 0.74 | 0.82 |
| Commitment seeking | 0.84 | 0.89 | 0.87 | Refuses to relay | 0.89 | 0.57 | 0.70 |
| Follow-up arrangement | 0.97 | 0.75 | 0.85 | | | | |
| Responds to complaint | 0.82 | 0.58 | 0.68 | | | | |
| (4) Misconduct | | | | | | | |
| Deception / misrepresentation | 0.79 | 0.49 | 0.60 | | | | |
| Accuses of bad faith | 0.88 | 0.92 | 0.90 | | | | |
| Repeated pressure | 0.74 | 0.58 | 0.65 | | | | |
| Insult / abuse | 0.67 | 0.49 | 0.56 | | | | |
| Intimidation | 0.67 | 1.00 | 0.80 | | | | |
| (5) Contact calls only | | | | | | | |
| Asks contact to relay | 0.94 | 0.90 | 0.92 | | | | |
| Asks about the borrower | 0.86 | 0.86 | 0.86 | | | | |
| Explains why contact is called | 1.00 | 0.76 | 0.86 | | | | |
| Asks contact to pay | 0.67 | 0.40 | 0.50 | | | | |
| Macro average (29 codes) | 0.87 | 0.80 | 0.83 | Macro average (19 codes) | 0.82 | 0.72 | 0.76 |
| 95% interval | [0.84, 0.91] | [0.77, 0.84] | [0.80, 0.86] | 95% interval | [0.79, 0.87] | [0.69, 0.77] | [0.73, 0.80] |

*Notes*: F1 is the harmonic mean of precision and recall. Macro averages are simple means over codes, with 95% intervals from a stratified bootstrap with 2,000 replications.

The same author rated 100 pilot calls on the five rating scales. Table A2 reports the agreement. The intraclass correlation between the LLM and the human ratings ranges from 0.72 to 0.81 across the scales, and the two ratings are within one point of each other on 93–100% of the calls. The interval for empathy is wide because the scale has little variation in routine calls: the reader rated 92% of the calls at 2, normal politeness in an interpersonal relationship.

Table A2: LLM ratings versus the human reader

| | ICC(2,1) | | Agreement | | |
|---|---|---|---|---|---|
| | Estimate | 95% interval | Exact | Within ±1 | Mean difference (LLM − human) |
| Pressure | 0.77 | [0.65, 0.85] | 0.62 | 0.95 | 0.21 |
| Professionalism | 0.81 | [0.71, 0.88] | 0.75 | 0.95 | −0.06 |
| Empathy | 0.80 | [0.39, 0.91] | 0.93 | 1.00 | −0.01 |
| Negotiation | 0.72 | [0.56, 0.83] | 0.77 | 0.93 | −0.06 |
| Borrower resistance | 0.74 | [0.58, 0.85] | 0.78 | 0.95 | 0.01 |

*Notes*: ICC(2,1) is the two-way random-effects intraclass correlation for absolute agreement, with a bootstrap 95% interval. Exact and within-±1 agreement are shares of calls.

**Stability.** Repeating the annotation of 100 random calls with the same configuration reproduces 93% of the turn-level labels (Cohen's $\kappa = 0.92$) and 97% of the call-level indicators, and the ratings of a repeated run have intraclass correlations of 0.97–1.00 with the original. Presenting the rating options in a permuted order changes the ratings little (intraclass correlations of 0.82–0.94).

**Predictive validity.** If a label captures what its name says, it should predict the behavior that the name implies after the call, but not before it. For each label $x$ and outcome $y$, I regress an indicator of $y$ in a window *after* the call on $x$, and, separately, an indicator of $y$ in an equally long window *before* the call on $x$, controlling for the borrower and loan characteristics of the headline specifications, an indicator of any payment in days 1–25, the call's rank among the borrower's transcribed calls, and an indicator of the last transcribed call (to account for the fact that calls stop after repayment), with month fixed effects, transcription-probability weights, and standard errors clustered at the collector level. The borrower's unobserved heterogeneity affects both windows, so the difference between the forward and backward coefficients removes the part of the association that reflects who the borrower is rather than what the call contains. Calls on days 26–27, whose backward window falls mostly in days 1–25, before the case reached the current collector, provide a cleaner comparison.

Table A3: Predictive validity of the LLM labels: forward versus backward windows

| | | | | Forward − backward | | By sector (all days) | |
|---|---|---|---|---|---|---|---|
| Label | Outcome | Forward | Backward | All days | Days 26–27 | In-house | Third-party |
| *Panel A: Calls to the borrower's own number* | | | | | | | |
| Promise to pay (any) | Payment, 3 days | 0.030 | 0.004 | 0.026 | 0.032 | 0.026 | 0.028 |
| | | | | (0.004) | (0.005) | (0.004) | (0.010) |
| Promise to pay (specific) | Payment, 3 days | 0.074 | 0.010 | 0.064 | 0.074 | 0.065 | 0.059 |
| | | | | (0.006) | (0.008) | (0.006) | (0.014) |
| Promise to pay (vague) | Payment, 3 days | 0.011 | 0.001 | 0.010 | 0.014 | 0.010 | 0.006 |
| | | | | (0.004) | (0.005) | (0.004) | (0.010) |
| Partial commitment | Payment, 3 days | 0.085 | 0.015 | 0.071 | 0.084 | 0.073 | 0.061 |
| | | | | (0.017) | (0.026) | (0.020) | (0.024) |
| Concession offer (collector) | Waiver payment, 7 days | 0.047 | 0.003 | 0.044 | 0.032 | 0.042 | 0.047 |
| | | | | (0.007) | (0.008) | (0.008) | (0.015) |
| Requests relief | Waiver payment, 7 days | 0.033 | −0.000 | 0.033 | 0.022 | 0.033 | 0.032 |
| | | | | (0.006) | (0.008) | (0.007) | (0.013) |
| Valence (mean) | Payment, 3 days | 0.018 | 0.008 | 0.010 | 0.023 | 0.009 | 0.028 |
| | | | | (0.006) | (0.006) | (0.006) | (0.016) |
| Valence trend | Payment, 3 days | 0.015 | 0.004 | 0.010 | 0.016 | 0.010 | 0.013 |
| | | | | (0.004) | (0.005) | (0.004) | (0.008) |
| Intensity (mean) | Payment, 3 days | −0.011 | −0.005 | −0.006 | −0.015 | −0.005 | −0.023 |
| | | | | (0.004) | (0.004) | (0.004) | (0.010) |
| *Panel B: All calls* | | | | | | | |
| Threatens to complain | Complaint, 30 days | 0.003 | 0.001 | 0.002 | 0.001 | 0.002 | −0.001 |
| | | | | (0.002) | (0.003) | (0.003) | (0.005) |
| *Panel C: Calls to social contacts* | | | | | | | |
| Denies knowing the borrower | Later dials to the number | −0.880 | −0.276 | −0.605 | | −0.624 | −0.078 |
| | | | | (0.061) | | (0.070) | (0.012) |
| Agrees to relay | Borrower answers, 1 day | 0.031 | 0.017 | 0.014 | | 0.017 | 0.006 |
| | | | | (0.006) | | (0.006) | (0.003) |

*Notes*: Standard errors of the forward-minus-backward differences in parentheses. Each outcome is measured over the same length of window before and after the call. The valence trend is the within-call slope of the borrower's valence.

Table A3 reports the results. A promise to pay is associated with a 3.0-percentage-point higher probability of a payment in the next three days, against 0.4 in the three days before. The gradient follows the codebook: a specific promise (6.4 percentage points net of the backward window) predicts payment far better than a vague one (1.0), and the slopes are the same for in-house and third-party borrowers. A concession offer by the collector and a relief request by the borrower each predict a payment with a fee reduction in the next seven days by about 3–4 percentage points net of the backward window. On calls to social contacts, a contact who denies knowing the borrower receives far fewer later calls on that number, and a contact who agrees to relay a message is followed by a higher probability that the borrower answers an own-number call within a day. The borrower's expressed emotion carries a weaker signal: on days 26–27, a more positive valence and a more intense expression predict, respectively, a higher and a lower probability of a later

payment net of the backward window, and an improving valence over the call predicts payment on all days, but on all days the mean valence and intensity are not significantly different between the two windows. Complaint threats do not predict filed complaints, which are rare in the sample.

**Coherence across label families.** The strategy and move codes, the ratings, and the emotion labels are produced by separate prompts. If they measure what they claim, each rating should move most closely with the codes and emotions that describe the same aspect of the call. For each rating, I build a composite of the strategy or move codes that its rubric refers to (e.g., the number of threat and consequence codes for pressure, the number of misconduct codes for professionalism, and the number of resistance moves minus the number of commitments for borrower resistance) and a composite of the corresponding emotion measures (see Table A4 note). Table A4 reports the correlations. Each rating correlates most strongly with its own strategy composite—0.74 for pressure, 0.59 for negotiation, 0.60 for borrower resistance, −0.57 for professionalism with misconduct, and 0.34 for empathy—and in regressions of each rating on all composites of a family, the own composite has the largest coefficient in every case. At the turn level, collector turns carrying a threat or consequence code are negative in tone 22% of the time against 9% of other turns, turns coded as insults or intimidation 97% against 10%, and turns coded as empathy positive 74% against 1%.

Table A4: Coherence across label families: ratings and composites of codes and emotions

| | Strategy and move composites | | | | | Emotion composites | | | |
|---|---|---|---|---|---|---|---|---|---|
| Rating | Pressure | Empathy | Misconduct | Negotiation | Resistance | Pressure | Empathy | Misconduct | Resistance |
| Pressure | **0.74** | 0.11 | 0.50 | 0.41 | 0.20 | **0.47** | −0.27 | 0.41 | 0.31 |
| Empathy | −0.00 | **0.34** | −0.17 | 0.23 | −0.18 | −0.20 | **0.38** | −0.40 | −0.12 |
| Professionalism | −0.50 | −0.01 | **−0.57** | −0.14 | −0.20 | −0.41 | 0.30 | **−0.40** | −0.23 |
| Negotiation | 0.46 | 0.23 | 0.11 | **0.59** | −0.08 | 0.10 | 0.10 | 0.00 | 0.06 |
| Borrower resistance | −0.06 | −0.01 | 0.06 | −0.17 | **0.60** | 0.29 | −0.27 | 0.33 | **0.58** |

*Notes*: Weighted Spearman correlations on the 100,938 transcribed calls with ratings, codes, and emotion labels. Bold entries mark the composite targeted by each rating. Emotion composites average standardized emotion measures: the collector's intensity and negative and impatient turn shares (pressure), the collector's warm, courteous, and mean-valence measures (empathy), the collector's rude, contemptuous, and angry turn shares (misconduct), and the borrower's angry, negative, and rude turn shares and minus mean valence (resistance).

### A.3.4 Model configuration

I run the annotation with Qwen3.8-27B in non-reasoning mode at zero temperature. Enabling reasoning, with medium thinking effort and the officially recommended tempera-

ture of 1.0, takes about 14 times as long per call without improving classification accuracy: on the pilot sample, the average F1 score across the collector strategies against the human benchmark is 0.80 in reasoning mode, compared with 0.83 in non-reasoning mode. Reasoning also makes the annotation less stable. In a repeated run on 100 random calls, the non-reasoning mode at zero temperature reproduces 93% of the turn-level labels, while the reasoning mode reproduces only 86%.

# Appendix B Additional Tables and Figures

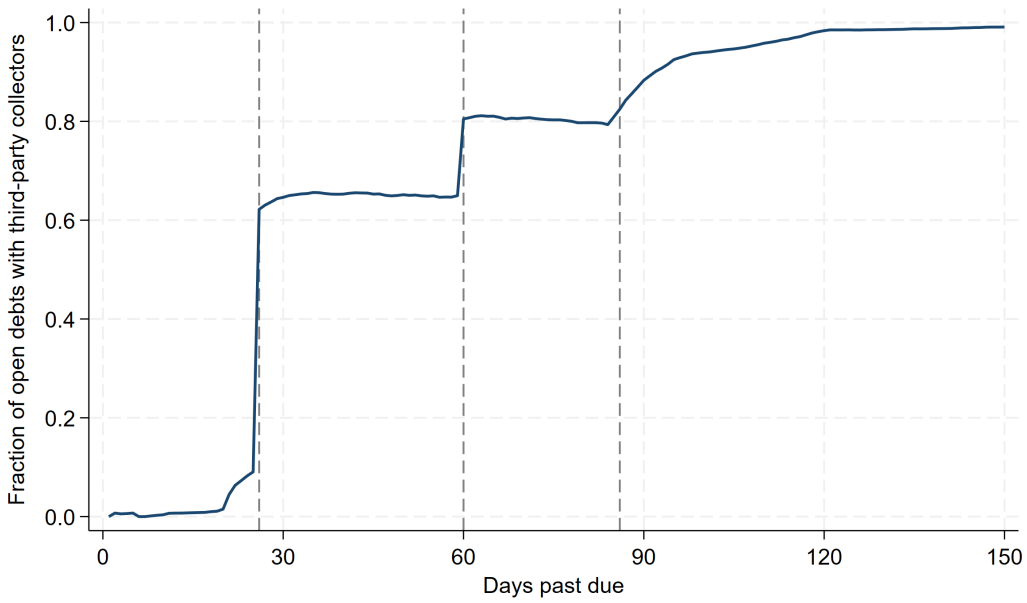



Figure B1: Fraction of delinquent debts with third-party collectors, by days past due

*Notes*: The figure plots the fraction of active delinquent debts assigned to third-party collection queues on each day past due, using a representative subsample of the whole delinquent debt inventory. The dashed lines mark days 26, 60, and 86, the stage transitions at which batches of debts are allocated to third-party agencies.

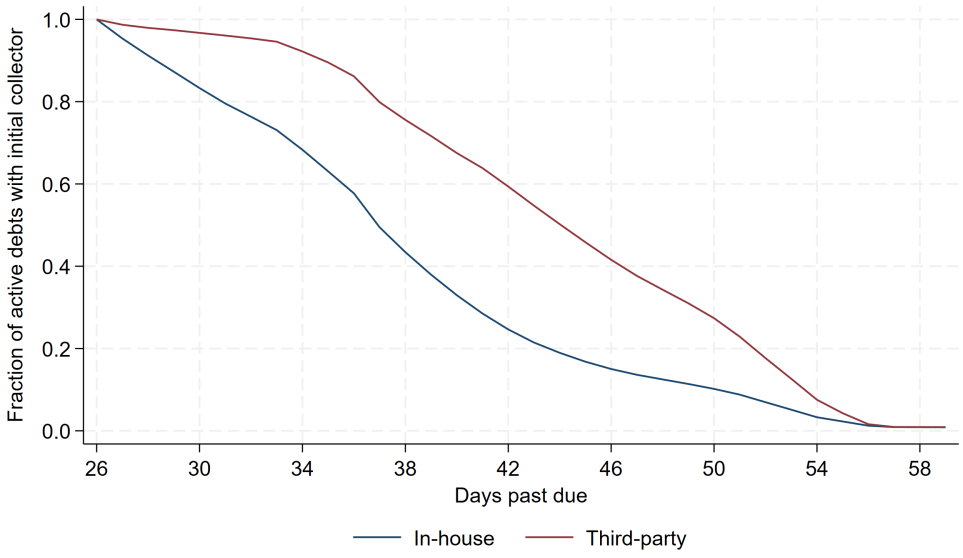



Figure B2: Debt retention with the initial collector over days 26–59

*Notes*: The figure plots, for each day past due, the fraction of active debts served by the same collector who served them on day 26, separately for in-house and third-party debts.

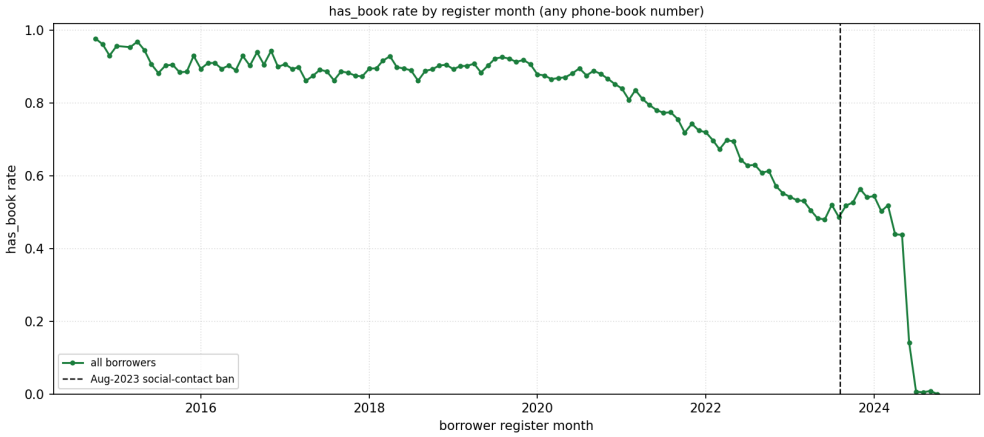



Figure B3: Phone-book availability by borrower registration month

*Notes*: The figure plots, by borrower registration month, the fraction of borrowers for whom the lender holds at least one phone number accessed from the borrower's cell-phone contact book and phone logs. The decline starting with late-2020 registration cohorts follows the tightening of personal-information regulation on lending apps' access to contact lists; the collection of phone-book contacts stops entirely for cohorts registering after mid-2024. The dashed line marks the August 2023 regulatory ban on collection calls to phone-book contacts.

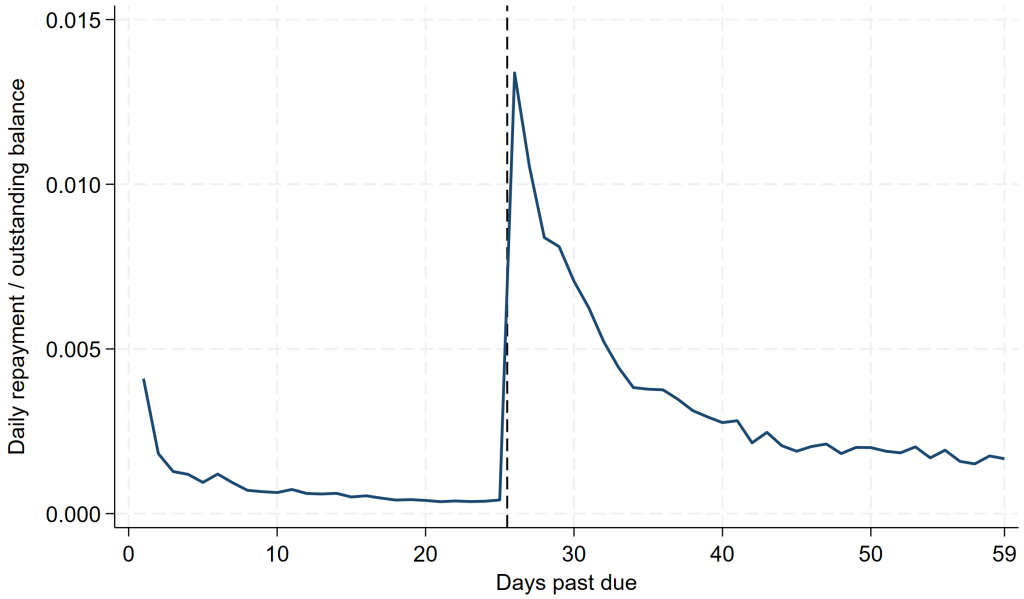



Figure B4: Daily repayment rate over days past due

*Notes*: The figure plots aggregate repayment on each day past due as a fraction of the aggregate outstanding balance, computed on the working sample, which consists of first-time delinquent borrowers whose overdue balance remains open on day 26. The outstanding balance is the day-1 overdue balance, less cumulative repayments to date. The dashed line marks day 26.

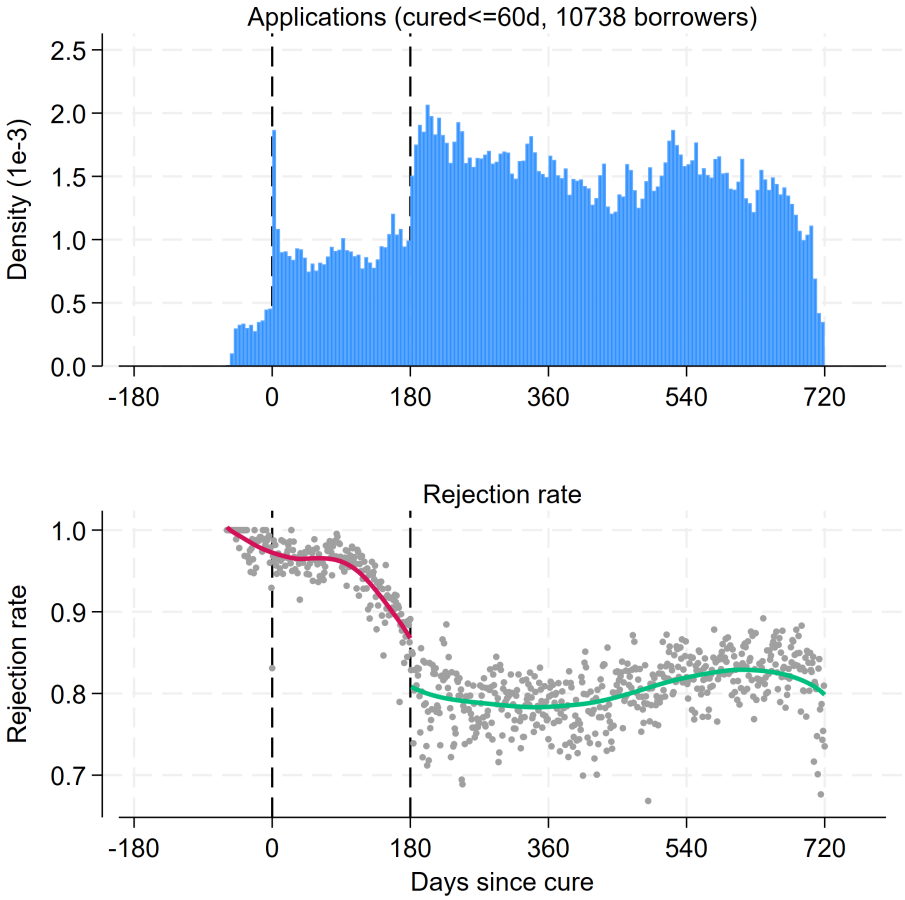
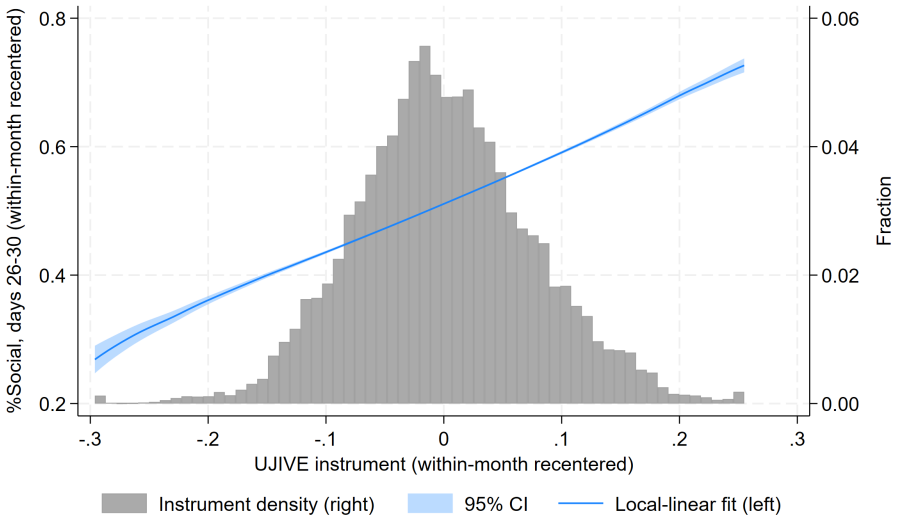



Figure B5: Return applications and rejection rates by days since cure

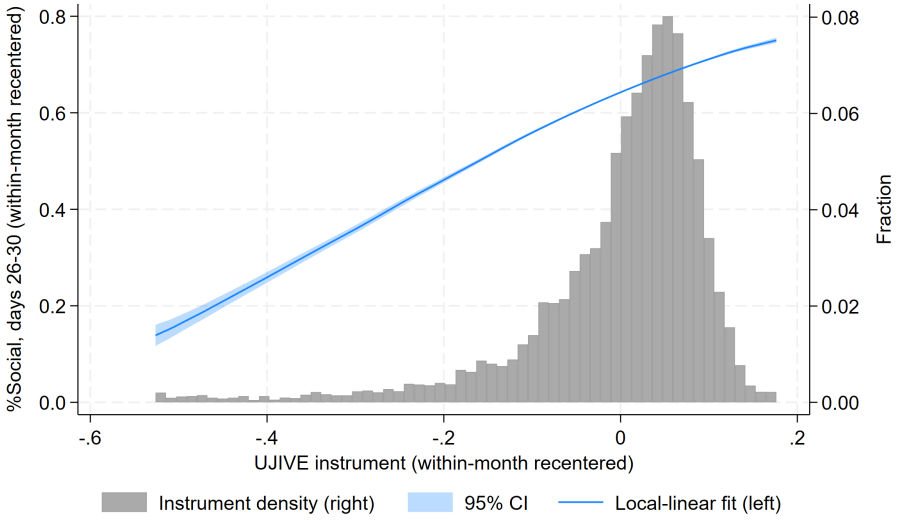
*Notes*: Both panels condition on borrowers cured within 60 days of the day-26 assignment and cover all their return loan applications within 720 days. The upper panel shows the density of applications by days since the overdue balance closing date (5-day bins). The lower panel shows the daily rejection rate (gray dots) with locally weighted scatterplot smoothing (LOWESS) fits estimated separately before and after day 180. The dashed lines mark the cure day (0) and the 180-day relaxation of credit access restrictions.

Figure B6: First-stage relationships

*Panel A. In-house*




*Panel B. Third-party*




*Notes*: Each panel plots, for the stated estimation sample, the distribution of the leave-one-out UJIVE instrument of Section 3.1, recentered within calendar month and winsorized at its 0.1/99.9 percentiles for illustrative purposes (gray histogram, right axis, fraction), together with a local-linear fit of %*social* (days 26–30; within-month recentered, sample mean added back) on the instrument, with the shaded 95% confidence band.

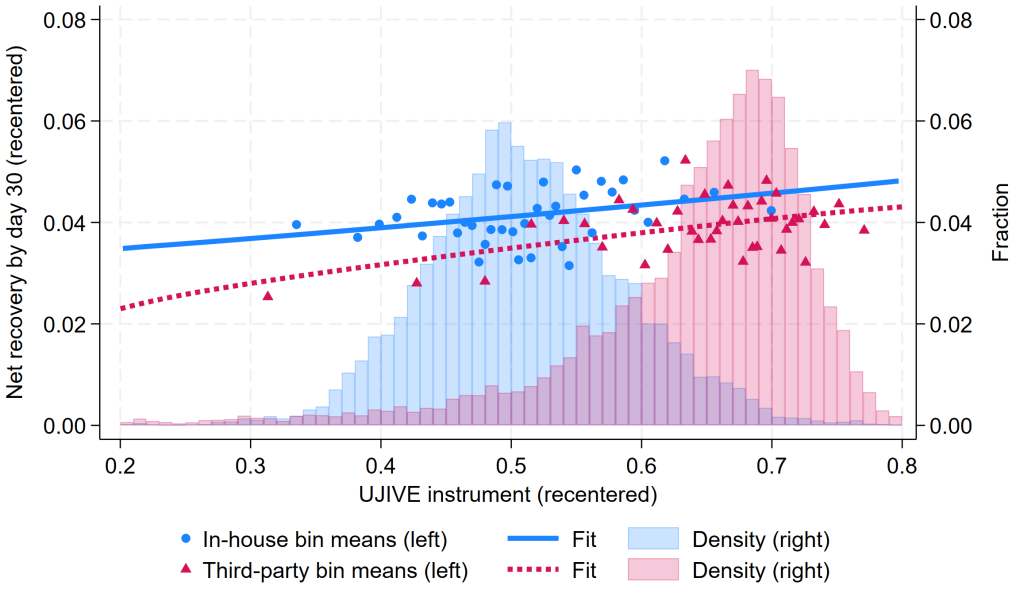



Figure B7: Reduced-form effect of social-contact calling on net recovery

*Notes*: The figure repeats Figure 3 with the outcome replaced by net recovery by day 30: the recovered NPV less a variable cost of 3% of collections and 1 yuan (0.14 USD) per minute of talk and ring time over all call attempts during days 26–30, both divided by 0.45 to obtain total costs, scaled by the overdue balance at the day-26 assignment. Recentering, binning, local-linear fits, trimming, and histograms are as in Figure 3.

Table B1: Front-loading of collection activity over days 26–59

| | Days past due window | | | | | Caller order | | First × | |
|---|---|---|---|---|---|---|---|---|---|
| | [26,30] | [31,35] | [36,40] | [41,50] | [51,59] | First | Later | [26,30] | N |
| *Panel A: in-house* | | | | | | | | | |
| No. of calls | 0.328 | 0.189 | 0.133 | 0.204 | 0.146 | 0.567 | 0.433 | 0.306 | 75,722 |
| No. of answered calls | 0.404 | 0.175 | 0.117 | 0.176 | 0.128 | 0.609 | 0.391 | 0.379 | 72,055 |
| Call duration | 0.459 | 0.166 | 0.106 | 0.157 | 0.112 | 0.638 | 0.362 | 0.432 | 72,019 |
| Money collected | 0.424 | 0.176 | 0.111 | 0.168 | 0.121 | 0.616 | 0.384 | 0.400 | 15,320 |
| *Panel B: third-party* | | | | | | | | | |
| No. of calls | 0.207 | 0.178 | 0.108 | 0.227 | 0.280 | 0.540 | 0.460 | 0.204 | 78,716 |
| No. of answered calls | 0.212 | 0.174 | 0.110 | 0.222 | 0.283 | 0.537 | 0.463 | 0.208 | 63,485 |
| Call duration | 0.229 | 0.169 | 0.113 | 0.221 | 0.268 | 0.550 | 0.450 | 0.225 | 63,387 |
| Money collected | 0.295 | 0.210 | 0.145 | 0.198 | 0.152 | 0.703 | 0.297 | 0.290 | 15,711 |

*Notes*: The table reports the average within-borrower distribution of collection activity over days 26–59. For each borrower, the share of each activity measure falling in each day window, and the share delivered by the initial (day-26) collector versus later collectors, are computed over days 26–59; days after the case closes count as zeros, and case-days after a transfer to the collection queues for debts 60 or more days past due are excluded. The table reports the average of these within-borrower shares, separately for in-house-collected (Panel A) and third-party-collected (Panel B) debts. Rows are the number of calls dialed, the number of answered calls, total call duration, and money collected (cash repayments, excluding waived amounts), each attributed to the collector serving the day on which the activity occurred. Borrowers with none of a given activity are excluded from that row; N reports the remaining borrower count. The sample is all cases assigned to collectors on day 26.

Table B2: Balance test: pre-treatment characteristics and the UJIVE instrument

| *Panel A: in-house* | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Overdue amount | Remaining principal | Score decile | Male | Age | Bachelor or above | Days 1–25 repaid share |
| %*social* (instrumented) | -303.341* | -219.893 | -0.255 | 0.003 | 0.012 | -0.004 | 0.003 |
| | (-1.69) | (-0.25) | (-1.31) | (0.09) | (0.02) | (-0.20) | (0.68) |
| Mean dep. var. | 2590.976 | 19247.726 | 4.997 | 0.696 | 31.118 | 0.095 | 0.017 |
| Observations | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 |
| *Panel B: third-party* | | | | | | | |
| | Overdue amount | Remaining principal | Score decile | Male | Age | Bachelor or above | Days 1–25 repaid share |
| %*social* (instrumented) | -47.108 | -154.835 | 0.044 | -0.002 | -0.220 | 0.013 | 0.001 |
| | (-0.39) | (-0.27) | (0.33) | (-0.07) | (-0.64) | (0.87) | (0.25) |
| Mean dep. var. | 2566.421 | 19178.868 | 5.007 | 0.694 | 31.116 | 0.095 | 0.017 |
| Observations | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 |

*Notes*: Each column reports a separate UJIVE regression of a pre-treatment characteristic on %*social* (days 26–30), instrumented by the leave-one-out collector–month instrument of Section 3.1, with calendar-month fixed effects and no other controls. Columns: overdue balance at day 26 (yuan), remaining loan principal (yuan), internal credit-score decile (1–10), male indicator, age, bachelor-degree-or-above indicator, and the fraction of the overdue balance repaid during days 1–25. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B3: Average monotonicity test

| | In-house | | | Third-party | | |
|---|---|---|---|---|---|---|
| | Share | Estimate | t | Share | Estimate | t |
| *Overdue amount (terciles)* | | | | | | |
| Low | 0.330 | 0.360 | 20.09 | 0.333 | 0.352 | 28.67 |
| Middle | 0.331 | 0.322 | 17.75 | 0.336 | 0.345 | 27.51 |
| High | 0.338 | 0.319 | 17.50 | 0.331 | 0.303 | 24.56 |
| *Remaining principal (terciles)* | | | | | | |
| Low | 0.332 | 0.366 | 19.98 | 0.334 | 0.347 | 27.81 |
| Middle | 0.333 | 0.295 | 16.09 | 0.333 | 0.344 | 27.42 |
| High | 0.335 | 0.339 | 18.23 | 0.333 | 0.310 | 24.75 |
| *Age (terciles)* | | | | | | |
| Low | 0.365 | 0.349 | 18.41 | 0.363 | 0.369 | 28.59 |
| Middle | 0.303 | 0.314 | 17.42 | 0.306 | 0.307 | 25.01 |
| High | 0.333 | 0.337 | 18.50 | 0.331 | 0.324 | 26.10 |
| *Credit-score decile* | | | | | | |
| Deciles 1–3 | 0.348 | 0.349 | 18.75 | 0.349 | 0.336 | 26.41 |
| Deciles 4–7 | 0.425 | 0.426 | 21.98 | 0.421 | 0.431 | 32.69 |
| Deciles 8–10 | 0.227 | 0.225 | 13.84 | 0.230 | 0.233 | 21.05 |
| *Gender* | | | | | | |
| Female | 0.304 | 0.311 | 17.10 | 0.306 | 0.303 | 24.83 |
| Male | 0.696 | 0.689 | 37.85 | 0.694 | 0.697 | 57.08 |
| *Education* | | | | | | |
| Below bachelor | 0.905 | 0.902 | 78.90 | 0.905 | 0.898 | 118.15 |
| Bachelor or above | 0.095 | 0.098 | 8.62 | 0.095 | 0.102 | 13.43 |
| *Repayment before day 26* | | | | | | |
| No repayment in days 1–25 | 0.820 | 0.843 | 56.37 | 0.814 | 0.836 | 79.19 |
| Positive repayment in days 1–25 | 0.180 | 0.157 | 10.53 | 0.186 | 0.164 | 15.58 |

*Notes*: Each cell reports a UJIVE regression of $V_i \times \%social$ on $\%social$, where $V_i$ indicates membership in the stated borrower group, with the same specification and samples as Appendix Table B2; Share is the group's sample share; tercile cutoffs are set on the pooled in-house and third-party sample. Under non-negative IV weights the estimate is a weighted average of the 0/1 indicator and must lie between zero and one; within each partition the estimates sum to one by construction. $t$-statistics are based on standard errors robust to heteroskedasticity and treatment effect heterogeneity. $N = 66{,}664$ (in-house) and 60,271 (third-party) in every cell.

Table B4: Complaints and subsequent return borrowing

| *Panel A: in-house, any reborrow application within horizon since dpd-26 (no cure conditioning)* | | | | | | |
|---|---|---|---|---|---|---|
| | 1 month | 3 months | 6 months | 1 year | 2 years | |
| Any complaint (dpd 26–59) | -0.0083*** | -0.0163*** | -0.0265*** | -0.0436*** | -0.0507*** | |
| | (-2.84) | (-3.73) | (-5.03) | (-6.67) | (-6.60) | |
| Mean dep. var. | 0.0116 | 0.0243 | 0.0380 | 0.0641 | 0.0812 | |
| Mean complaint | 0.0104 | 0.0104 | 0.0104 | 0.0104 | 0.0104 | |
| Observations | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 | |
| *Panel B: in-house, any application within days since cure + gate decomposition (cured≤60d)* | | | | | | |
| | [0,30] | [0,60] | [0,120] | [0,180] | First [180,210] | Repeat [180,210] |
| Any complaint (dpd 26–59) | -0.0205 | -0.0529* | -0.0902*** | -0.0953** | -0.0297*** | -0.0485*** |
| | (-0.68) | (-1.67) | (-2.74) | (-2.42) | (-5.21) | (-6.09) |
| Mean dep. var. | 0.0689 | 0.0940 | 0.1310 | 0.1745 | 0.0337 | 0.0496 |
| Mean complaint | 0.0059 | 0.0059 | 0.0059 | 0.0059 | 0.0059 | 0.0059 |
| Observations | 10,230 | 10,230 | 10,230 | 10,230 | 10,230 | 10,230 |
| *Panel C: third-party, any reborrow application within horizon since dpd-26 (no cure conditioning)* | | | | | | |
| | 1 month | 3 months | 6 months | 1 year | 2 years | |
| Any complaint (dpd 26–59) | 0.0004 | -0.0045 | -0.0078 | -0.0232** | -0.0339*** | |
| | (0.08) | (-0.67) | (-0.96) | (-2.52) | (-3.45) | |
| Mean dep. var. | 0.0099 | 0.0244 | 0.0384 | 0.0641 | 0.0816 | |
| Mean complaint | 0.0084 | 0.0084 | 0.0084 | 0.0084 | 0.0084 | |
| Observations | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 | |
| *Panel D: third-party, any application within days since cure + gate decomposition (cured≤60d)* | | | | | | |
| | [0,30] | [0,60] | [0,120] | [0,180] | First [180,210] | Repeat [180,210] |
| Any complaint (dpd 26–59) | 0.0212 | 0.0168 | -0.0000 | -0.0260 | -0.0114 | -0.0479* |
| | (0.43) | (0.31) | (-0.00) | (-0.42) | (-0.46) | (-1.81) |
| Mean dep. var. | 0.0661 | 0.0987 | 0.1426 | 0.1876 | 0.0398 | 0.0569 |
| Mean complaint | 0.0049 | 0.0049 | 0.0049 | 0.0049 | 0.0049 | 0.0049 |
| Observations | 8,432 | 8,432 | 8,432 | 8,432 | 8,432 | 8,432 |

*Notes*: OLS regressions of return-borrowing indicators on an indicator for the borrower filing any complaint during days 26–59, estimated on the in-house (Panels A and B) and third-party (Panels C and D) working samples of Table 1. Panels A and C outcomes are indicators for any return application within the stated horizon since the day-26 date (no cure conditioning); Panels B and D outcomes condition on cure within 60 days of assignment and are indicators for any application within the stated window of days since cure, plus the first-time vs. repeat decomposition of applications in days [180,210] since cure. All regressions control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status, and include collector and calendar-month fixed effects. $t$-statistics based on heteroskedasticity-robust standard errors in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B5: Recovery effects across alternative jackknife IV estimators

| | In-house | | | | Third-party | | | |
|---|---|---|---|---|---|---|---|---|
| | No controls | | With controls | | No controls | | With controls | |
| | Estimate | $t$ | Estimate | $t$ | Estimate | $t$ | Estimate | $t$ |
| *Panel A: alternative jackknife estimators, caseload $\geq$ 15 (N = 66,664 / 60,271)* | | | | | | | | |
| OLS | 0.063 | 12.96 | 0.060 | 12.20 | 0.020 | 5.90 | 0.016 | 4.87 |
| UJIVE | 0.034 | 2.04 | 0.034 | 2.08 | 0.033 | 3.74 | 0.033 | 3.75 |
| JIVE1 | 0.034 | 2.28 | 0.034 | 2.32 | 0.033 | 4.03 | 0.033 | 4.04 |
| IJIVE | 0.034 | 2.28 | 0.034 | 2.32 | 0.033 | 4.03 | 0.033 | 4.04 |
| SJIVE | 0.037 | 2.50 | 0.037 | 2.54 | 0.037 | 4.60 | 0.037 | 4.60 |
| 2SLS, collector-by-month dummies | 0.040 | 3.38 | 0.040 | 3.42 | 0.030 | 4.30 | 0.030 | 4.26 |
| *Panel B: caseload filter* | | | | | | | | |
| UJIVE, caseload $\geq$ 10 (N = 72,761 / 63,453) | 0.033 | 2.15 | 0.034 | 2.21 | 0.033 | 4.05 | 0.033 | 4.00 |
| UJIVE, caseload $\geq$ 20 (N = 52,238 / 50,397) | 0.046 | 2.38 | 0.045 | 2.37 | 0.032 | 3.13 | 0.032 | 3.20 |
| UJIVE, no caseload filter (N = 73,945 / 63,654) | 0.033 | 2.12 | 0.034 | 2.21 | 0.035 | 4.31 | 0.035 | 4.25 |
| SJIVE, no caseload filter (N = 73,945 / 63,654) | 0.038 | 2.84 | 0.039 | 2.94 | 0.039 | 5.19 | 0.038 | 5.16 |

*Notes*: The table re-estimates the recovery regressions of Table 2 columns 1–4 (outcome: cumulative recovered NPV by day 30, scaled by the overdue balance at assignment; treatment: %*social* over days 26–30; calendar-month fixed effects partialled out) under alternative estimators and sample filters. Panel A: the UJIVE row is the headline estimator with the leave-one-out instrument of Section 3.1; JIVE1 uses the leave-one-out cell mean without UJIVE's recentering by the leave-out month mean; IJIVE residualizes on the month fixed effects before the leave-out construction; SJIVE shrinks the caller–month cell propensities toward caller-level propensities with empirical-Bayes weights; 2SLS uses caller-by-month dummies directly as instruments. Panel B varies the collector-month caseload filter of the working sample (at least 15 cases in Panel A) while keeping the other sample filters. "With controls" adds the borrower and loan characteristics of Table 2 column 4. $t$-statistics use heteroskedasticity-robust standard errors; for the UJIVE rows, they are based on the standard errors robust to treatment effect heterogeneity, the convention of Table 2. $N$s are the sizes of estimation samples (in-house / third-party).

Table B6: Characteristics of borrowers driving the identifying variation

| | In-house | | | Third-party | | |
|---|---|---|---|---|---|---|
| | UJIVE-weighted | Sample mean | Norm. diff. | UJIVE-weighted | Sample mean | Norm. diff. |
| Overdue amount (yuan) | 2,335 | 2,591 | -0.099 | 2,407 | 2,566 | -0.061 |
| | (96.8) | | | (66.7) | | |
| Log overdue amount | 7.547 | 7.593 | -0.066 | 7.535 | 7.586 | -0.074 |
| | (0.0263) | | | (0.0179) | | |
| Credit-score decile (1–10) | 4.96 | 5.00 | -0.013 | 5.03 | 5.01 | 0.007 |
| | (0.1062) | | | (0.0724) | | |
| Repaid share, dpd 1–25 | 0.0167 | 0.0171 | -0.007 | 0.0179 | 0.0174 | 0.008 |
| | (0.0024) | | | (0.0017) | | |
| Age | 31.1 | 31.1 | 0.001 | 31.0 | 31.1 | -0.018 |
| | (0.2668) | | | (0.1828) | | |
| Male | 0.689 | 0.696 | -0.015 | 0.697 | 0.694 | 0.006 |
| | (0.0182) | | | (0.0122) | | |
| Bachelor's degree or above | 0.098 | 0.095 | 0.013 | 0.102 | 0.095 | 0.022 |
| | (0.0114) | | | (0.0076) | | |
| Observations | 66,664 | | | 60,271 | | |

*Notes*: Each cell in the UJIVE-weighted columns reports a UJIVE regression of $W_i \times \%social$ on $\%social$, where $W_i$ is the stated pre-treatment characteristic, using the no-covariate specification and samples of Table 2 column 2 (calendar-month fixed effects only). The estimate equals a weighted average of $W_i$ whose weights sum to one and are the same weights with which the UJIVE aggregates individual treatment effects (Goldsmith-Pinkham et al., 2026); the Sample mean columns report the unweighted mean of $W_i$ on the same estimation sample, and the Norm. diff. columns the difference between the two, divided by the sample standard deviation of $W_i$. Standard errors robust to heteroskedasticity and treatment effect heterogeneity in parentheses.

Table B7: Social-contact calling and recovery at longer horizons

| | (1) Day 30 | (2) Day 35 | (3) Day 40 | (4) Day 45 | (5) Day 50 | (6) Day 55 | (7) Day 59 |
|---|---|---|---|---|---|---|---|
| *Panel A: in-house* | | | | | | | |
| %*social* | 0.034** | 0.026 | 0.029 | 0.030 | 0.033 | 0.023 | 0.027 |
| | (2.08) | (1.39) | (1.47) | (1.44) | (1.57) | (1.08) | (1.21) |
| Mean dep. var. | 0.068 | 0.091 | 0.104 | 0.114 | 0.123 | 0.131 | 0.137 |
| Observations | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 | 66,664 |
| *Panel B: third-party* | | | | | | | |
| %*social* | 0.033*** | 0.024** | 0.009 | 0.016 | 0.009 | -0.004 | -0.003 |
| | (3.75) | (2.17) | (0.71) | (1.23) | (0.66) | (-0.24) | (-0.21) |
| Mean dep. var. | 0.045 | 0.072 | 0.090 | 0.102 | 0.113 | 0.123 | 0.130 |
| Observations | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 | 60,271 |
| *Panel C: third-party − in-house* | | | | | | | |
| %*social* | -0.002 | -0.002 | -0.020 | -0.013 | -0.024 | -0.027 | -0.030 |
| | (-0.09) | (-0.09) | (-0.86) | (-0.54) | (-0.95) | (-1.03) | (-1.12) |
| Observations | 126,935 | 126,935 | 126,935 | 126,935 | 126,935 | 126,935 | 126,935 |

*Notes*: The table re-estimates the covariate-adjusted UJIVE specification of Table 2, column (4), with the recovery horizon extended in five-day steps through day 59: the dependent variable in each column is the cumulative recovered NPV by the stated day since delinquency, scaled by the overdue balance at the day-26 assignment. Each cell is a separate regression of the outcome on %*social* (days 26–30), instrumented by the UJIVE instrument of Section 3.1, with calendar-month fixed effects and controls for log overdue amount, credit-score decile dummies, age, age squared, gender, education, and student status. Column (1) reproduces the day-30 estimates of Table 2, column (4), for reference. Mean dep. var. reports the mean of the dependent variable on the column's estimation sample. Panel C reports the third-party minus in-house coefficient difference, with the $z$-statistic of an independent-samples test based on the two panels' standard errors in parentheses. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B8: Returning effects across alternative jackknife IV estimators

| | Any application in [0,30] | | | | First application in [180,210] | | | |
|---|---|---|---|---|---|---|---|---|
| | In-house | | Third-party | | In-house | | Third-party | |
| | Estimate | $t$ | Estimate | $t$ | Estimate | $t$ | Estimate | $t$ |
| *Panel A: alternative jackknife estimators, caseload ≥ 15 (N = 10,257 / 8,439)* | | | | | | | | |
| OLS | 0.037 | 2.71 | 0.004 | 0.29 | 0.011 | 1.12 | 0.024 | 2.45 |
| UJIVE | −0.155 | −2.79 | −0.026 | −0.80 | −0.076 | −2.13 | 0.016 | 0.72 |
| JIVE1 | −0.155 | −3.07 | −0.026 | −0.86 | −0.076 | −2.39 | 0.016 | 0.78 |
| IJIVE | −0.155 | −3.07 | −0.026 | −0.86 | −0.076 | −2.39 | 0.016 | 0.78 |
| SJIVE | −0.145 | −3.05 | −0.021 | −0.71 | −0.071 | −2.27 | 0.022 | 1.13 |
| 2SLS, collector-by-month dummies | −0.017 | −0.66 | 0.002 | 0.11 | −0.013 | −0.82 | 0.019 | 1.30 |
| *Panel B: caseload filter* | | | | | | | | |
| UJIVE, caseload ≥ 10 (N = 10,999 / 8,872) | −0.127 | −2.36 | −0.037 | −1.21 | −0.072 | −2.12 | 0.015 | 0.72 |
| UJIVE, caseload ≥ 20 (N = 8,315 / 7,179) | −0.149 | −2.44 | −0.036 | −0.95 | −0.082 | −2.04 | 0.011 | 0.43 |
| UJIVE, no caseload filter (N = 11,137 / 8,897) | −0.122 | −2.29 | −0.034 | −1.14 | −0.071 | −2.12 | 0.016 | 0.78 |
| SJIVE, no caseload filter (N = 11,137 / 8,897) | −0.124 | −2.71 | −0.027 | −0.97 | −0.066 | −2.27 | 0.022 | 1.15 |

*Notes*: The table re-estimates the returning regressions of Table 3 columns 2 and 6 (left column group: any application within 30 days of cure; right column group: first-ever application in the $[180, 210]$-day window; treatment: $\%social$ over days 26–30; calendar-month fixed effects partialled out; borrower and loan controls included) under the alternative estimators and caseload filters of Appendix Table B5, on the same cured-within-60-days samples as Table 3 (Panel A) and on the corresponding samples under the alternative collector-month caseload filters (Panel B). $t$-statistics use heteroskedasticity-robust standard errors; for the UJIVE rows, they are based on the standard errors robust to treatment effect heterogeneity. $N$s are the sizes of estimation samples (in-house / third-party).

Table B9: Social-contact calling, recovery, and returning, by kin share of the contact network

| Outcome: | (1) Recovery | (2) Any [0,30] | (3) Any [0,180] | (4) First [180,210] | (5) Repeat [180,210] |
|---|---|---|---|---|---|
| *Panel A: in-house* | | | | | |
| %*social* | 0.038** | -0.172*** | 0.005 | -0.077** | -0.007 |
| | (2.20) | (-2.82) | (0.05) | (-2.07) | (-0.15) |
| %*social* × High kin | -0.018 | 0.092 | -0.141 | -0.029 | 0.201 |
| | (-0.36) | (0.65) | (-0.62) | (-0.25) | (1.51) |
| High kin | 0.002 | -0.068 | 0.044 | 0.017 | -0.125** |
| | (0.08) | (-0.99) | (0.40) | (0.31) | (-1.97) |
| Mean dep. var. | 0.068 | 0.069 | 0.175 | 0.034 | 0.049 |
| Observations | 66,370 | 10,217 | 10,217 | 10,217 | 10,217 |
| *Panel B: third-party* | | | | | |
| %*social* | 0.024** | -0.042 | -0.009 | 0.011 | 0.028 |
| | (2.46) | (-1.08) | (-0.16) | (0.45) | (0.85) |
| %*social* × High kin | 0.039* | 0.044 | 0.101 | 0.005 | 0.002 |
| | (1.85) | (0.66) | (0.87) | (0.08) | (0.03) |
| High kin | -0.023* | -0.048 | -0.092 | 0.002 | -0.013 |
| | (-1.67) | (-1.10) | (-1.20) | (0.07) | (-0.27) |
| Mean dep. var. | 0.045 | 0.066 | 0.187 | 0.040 | 0.057 |
| Observations | 59,861 | 8,398 | 8,398 | 8,398 | 8,398 |

*Notes*: UJIVE regressions interacting %*social* (days 26–30) with an indicator for a kin-rich contact network. For each borrower, all phone numbers ever called by collectors are classified by their platform-attached relationship labels into kin, spouse, and non-kin; the sample is restricted to borrowers with at least one labeled number, and High kin indicates the top tercile of the kin share of the labeled pool, with the tercile cut set within each panel's cured sample. Kin contacts include the borrower's parents, siblings, children, and other relatives, in contrast to non-kin contacts such as colleagues and friends. Both %*social* and its interaction with High kin are instrumented by the UJIVE instrument of Section 3.1 and its interaction. The column (1) outcome is the recovery rate by day 30; columns (2)–(5) condition on cure within 60 days of assignment, with outcomes an indicator for any return application in the stated window of days since cure and, in columns (4)–(5), the first-time vs. repeat decomposition of applications in [180,210]. All regressions control for log overdue amount, credit-score decile dummies, age, age squared, gender, education, student status, and the number of labeled contacts, with calendar-month fixed effects, and are estimated separately for in-house (Panel A) and third-party (Panel B) cases. $t$-statistics based on standard errors robust to heteroskedasticity and treatment effect heterogeneity are in parentheses.
Significance level: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B10: Transcript annotation: definitions of the codes, ratings, and emotion measures

| Code | Brief Definition |
|---|---|
| *Panel A: Collector strategies (turn-level codes; a call is coded 1 if any collector turn carries the code)* | |
| (1) Information | |
| Identity verification | Asks or confirms who is on the line, the callee's name, the callee's relation to the borrower, or whether the callee is the borrower. |
| States the debt | States factual information about the debt: the amount owed, days overdue, due date, product or instalment, or what has already been paid. |
| Asks reason / funds | Asks why the borrower has not paid, when money will be available, or about income, work, and pay dates. |
| Payment how-to | Explains the mechanics of paying: the app or page, the steps, the channel, or contacting customer service. |
| (2) Threats and consequences | |
| Legal threat | Says or implies that a legal or judicial process will follow: a lawsuit, court, arbitration, a lawyer's letter, the police, criminal liability, enforcement, or listing as a dishonest debtor; a statement that legal liability will be pursued, or a handover to a law firm or legal department. |
| Escalation threat | Says the case will be escalated, handed to another department or team, or visited in person at home or at work, or that collection will intensify. |
| Credit warning | States that non-payment worsens the borrower's standing: the credit record, blacklisting, future borrowing or consumption, or the loss of the terms of this loan (instalments cancelled, the whole loan falling due), or consequences for the borrower's children (schooling, jobs) if they have a bad credit. |
| Cost warning | Warns that the amount owed grows: interest, late fees, penalties, or the loss of a waiver or discount. |
| Vague consequence | States that something bad will follow without saying what ("you will bear the consequences," "it will affect you"). |
| Social-contact warning | Says or implies that family, employer, colleagues, friends, or emergency contacts will be contacted or informed. |
| Urgency / deadline | Sets or stresses a time limit (today, within an hour, the last chance) or demands immediate payment. |
| Moral appeal | Appeals to honesty, keeping one's word, responsibility, fairness, character, or the duty created by the contract, or reproaches the borrower for a promise not kept. |
| (3) Negotiation and rapport | |
| Empathy / rapport | Acknowledges the borrower's difficulty, expresses understanding or concern, offers help or patience, or apologises for disturbing. |
| Concession offer | Offers to reduce, waive, or refund fees, interest, or principal, or a discount conditional on paying, including a warning that an existing waiver is about to expire (also coded as a cost warning). |
| Payment plan | Proposes paying in parts, an extension, a later date, a minimum amount, or one instalment first, as an arrangement. |
| Funds-source suggestion | Suggests where to get the money: family or friends, an advance from the employer, selling something, savings, or another credit line. |

*Continued on next page*

Table B10 (continued)

| Code | Brief Definition |
|---|---|
| Downplays the amount | Presents the amount or the effort as small or easy ("just a few hundred," "a meal's worth"). |
| Commitment seeking | Asks the borrower to name a date or an amount, asks for a definite yes, or repeats a commitment just made to confirm it. |
| Follow-up arrangement | Sets up the next contact: asks the borrower to stay reachable or to call back, or announces a callback or a check at a given time. |
| Responds to complaint | After the borrower objects to the collection, denies harassment or illegality, asserts that the call is lawful or recorded. |
| (4) Misconduct | |
| Deception / misrepresentation | Makes a specific claim that is false or impossible at this stage: impersonating the police, a court, or a government office; framing the unpaid loan as a crime; stating as done what has not happened (already sued, already blacklisted); or directing payment to a private account. |
| Accuses of bad faith | Explicitly charges the borrower with intent or dishonesty: malicious default, loan fraud, lying, or false information. |
| Repeated pressure | Repeats the demand after the borrower has refused or explained, dismisses the explanation without new content, refuses to end the call, or says that calls will continue until payment. |
| Insult / abuse | Insults, mocks, humiliates, or swears at the borrower: name-calling, profanity, or contempt for the person or the family. |
| Intimidation | Threatens harm or trouble outside any legal or contractual process: coming to "find" the person, bringing people, making life difficult, or exposing the person publicly. |
| *Panel B: Ratings (1–5, one rating per call; the collector is rated except for resistance)* | |
| Pressure | Intensity of the means used to make the borrower act, at its peak over the call: 1 = informing and asking only; 2 = a deadline or one stated consequence; 3 = several consequences, or demands repeated after an explanation; 4 = insistent threats (legal, escalation, social contacts) that accept no explanation; 5 = intimidation, insults, or deceptive pressure, or refusing to end the call or saying calls will continue. |
| Professionalism | A misconduct ladder fixed by the worst conduct in the call: 5 = no misconduct (the default); 4 = rough (a harsh tone, vague or exaggerated consequences, no identity check); 3 = clear impropriety (bad-faith accusations, dominant repeated pressure, threats to inform family or the employer about the debt or to visit in person, pressing a contact to pay, continuing after the callee says they are not the borrower, or disclosing the debt to a contact); 2 = false statements or impersonation; 1 = insults or intimidation. |
| Empathy | The degree to which the collector acknowledges the borrower's situation, over the whole call: 1 = cold or hostile, dismissing a stated difficulty; 2 = normal politeness only; 3 = limited acknowledgement of the difficulties, then back to the demand; 4 = repeated understanding, asks about the situation, patient; 5 = adjusts the demand to the stated hardship (e.g., an extension, a partial payment). |

*Continued on next page*

Table B10 (continued)

| Code | Brief Definition |
|---|---|
| Negotiation | How far the collector moves toward a concrete, feasible arrangement: 1 = none (informs, demands payment in full, or pressures only); 2 = a one-way demand for a date that ignores the reply; 3 = interaction without closure; 4 = a specific arrangement closed with a confirmed date or amount; 5 = an arrangement fitted to the borrower's stated constraints, with the steps and the next contact set. |
| Borrower resistance | The borrower's stance toward the request, at its peak over the call: 1 = co-operative (a specific commitment, asks how to pay); 2 = cooperative with reservations (a vague promise, hardship with a relief request); 3 = evasive or passive; 4 = explicit refusal or dispute, a complaint, or a request to stop calling; 5 = hostile (insults, threats to complain or report, abrupt termination). |
| *Panel C: Borrower moves (turn-level codes; a call is coded 1 if any borrower turn carries the code)* | |
| Promise to pay (specific) | Commits to paying and names a time (today, tomorrow, a date, "when the salary comes") or an amount. |
| Promise to pay (vague) | Says they will pay without a time or an amount ("as soon as I can," "when I have money"). |
| Partial commitment | Offers or commits to paying part of the amount, a minimum, or in instalments, as their own proposal. |
| Requests relief | Asks for more time, a lower amount, a fee waiver, an instalment arrangement, or a halt to interest. |
| Cites hardship | Explains why they cannot pay: no money, unemployment, a delayed salary, illness, a family emergency, other debts, or a business loss. |
| Disputes the debt | Makes a factual claim against the debt: denies owing, says it was already paid, says automatic payment is set, disputes the amount, fees, or interest, claims fraud or identity theft, or says the loan is someone else's. |
| Asks for information | Asks about the debt, the amount, how to pay, the consequences, or the terms. |
| Asks caller's identity | Asks who the caller is, which platform or company, whether the caller is a real employee, how they got the number, or why they are calling. |
| Refuses | Explicitly refuses to pay or to discuss the debt ("I won't pay," "do whatever you want"). |
| Stalls / avoids | Visibly puts off or evades a demand or a question about payment without committing, refusing, disputing, or giving a reason ("let's see," "I'm busy," changing the subject). |
| Hostile / insults | Insults, swears at, mocks, or shouts at the collector. |
| Threatens to complain | Threatens to complain to the platform, a regulator, the police, or the media, or to sue, accuses the collector of illegal collection, or announces that the call is being recorded as evidence. |
| Asks to stop contact | Tells the collector not to call again or not to contact others, says they are hanging up, or cuts the conversation off in refusal. |
| Complains about conduct | Objects to how the collection is done (too many calls, calling family or colleagues, tone, timing, privacy) or to the platform's lending itself. |
| *Panel D: Borrower emotions (per-turn labels on the borrower's turns)* | |

*Continued on next page*

Table B10 (continued)

| Code | Brief Definition |
|---|---|
| Valence (mean) | The expressed affective tone of the turn, judged from the wording: $-2$ = clearly negative (hostility, anger, or abuse, or evident distress, despair, or pleading), $-1$ = mildly negative (impatience, complaint, worry, frustration or resignation, a curt refusal), 0 = neutral (statements, questions, routine courtesy), $+1$ = mildly positive (friendliness, cooperation, warmth beyond formula), $+2$ = clearly positive (gratitude, relief, explicit reassurance); the content of a refusal or a dispute does not lower valence unless the wording is hostile, impatient, or distressed. Averaged over the borrower's turns. |
| Intensity (mean) | The strength with which affect is expressed in the words of the turn, whatever its sign: 1 = none (flat, transactional), 2 = slight (one mild marker), 3 = clear emotional wording, 4 = strong (intensifiers, repetition, insults, pleading), 5 = a sustained outburst. Averaged over the borrower's turns. |
| Any emotion | Each borrower turn may carry one or more discrete emotion labels; the indicator for a class equals one if any of the borrower's turns in the call carries it. The four classes: *anger* = anger or hostility toward the collector or the platform (insults, aggressive challenges, heated threats); *fear* = worry or fear about the consequences or the situation (anxious questions, pleading); *sadness* = sadness, helplessness, despair, resignation, or grievance about the situation; *shame* = shame, guilt, or embarrassment about the debt or the delay (an apology, an admission of fault). |

*Notes*: Condensed definitions of the labels assigned by the LLM annotation of the Chinese call transcripts, as used in Figure 5, Figure 6, and Appendix Tables B11 and B12. Collector strategy codes (Panel A) and borrower move codes (Panel C) are assigned turn by turn from a fixed codebook with Chinese anchors and inclusion and exclusion rules; a code may co-occur with others in the same turn, and the call-level indicator equals one if any turn of the respective speaker carries the code. Intimidation is listed because it belongs to the misconduct family used in the multiple-testing adjustment but is omitted from the results table as not estimable in the third-party sample. Ratings (Panel B) are assigned once per call from the whole transcript on a five-point scale with behavioral anchors for each level; the anchor descriptions are abbreviated here. Calls with no substantive exchange are treated as unrated. Emotion labels (Panel D) are assigned to every turn of both speakers by a separate annotation pass from a codebook with Chinese anchors: a valence score, an expressed-intensity score, and zero or more discrete emotion labels; only the borrower's turns enter the measures reported here. Valence and intensity are averaged over the borrower's turns within the call; the emotion indicators are defined for calls in which the borrower speaks.

Table B11: Social-contact calling and the caller's conduct in the first conversation with the borrower

| | In-house | | | | | Third-party | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Coef. | $p$ | Adj. $p$ | Coef. (rewgt.) | Mean | Coef. | $p$ | Adj. $p$ |
| *Panel A: Collector strategies* | | | | | | | | | |
| (1) Information | | | | | | | | | |
| Identity verification | 0.821 | 0.027 | 0.111 | 0.180 | 0.027 | 0.899 | -0.034 | 0.029 | 0.078 |
| States the debt | 0.744 | 0.042 | 0.006 | 0.039 | 0.048 | 0.790 | 0.020 | 0.336 | 0.425 |
| Asks reason / funds | 0.382 | 0.016 | 0.465 | 0.612 | 0.025 | 0.571 | 0.032 | 0.202 | 0.303 |
| Payment how-to | 0.096 | 0.019 | 0.115 | 0.180 | 0.025 | 0.127 | 0.042 | 0.005 | 0.019 |
| (2) Threats and consequences | | | | | | | | | |
| Legal threat | 0.145 | 0.048 | 0.011 | 0.039 | 0.063 | 0.267 | **0.152** | <0.001 | <0.001 |
| Escalation threat | 0.229 | 0.045 | 0.053 | 0.116 | 0.051 | 0.170 | **0.072** | <0.001 | <0.001 |
| Credit warning | 0.196 | 0.050 | 0.009 | 0.039 | 0.053 | 0.432 | **0.101** | <0.001 | <0.001 |
| Cost warning | 0.053 | 0.008 | 0.389 | 0.541 | 0.006 | 0.149 | 0.039 | 0.010 | 0.030 |
| Vague consequence | 0.148 | 0.038 | 0.010 | 0.039 | 0.045 | 0.150 | 0.029 | 0.060 | 0.143 |
| Social-contact warning | 0.181 | 0.053 | 0.001 | 0.018 | 0.051 | 0.112 | -0.024 | 0.119 | 0.213 |
| Urgency / deadline | 0.366 | 0.052 | 0.008 | 0.039 | 0.054 | 0.389 | 0.038 | 0.136 | 0.217 |
| Moral appeal | 0.213 | 0.035 | 0.036 | 0.096 | 0.026 | 0.284 | 0.032 | 0.124 | 0.213 |
| (3) Negotiation and rapport | | | | | | | | | |
| Empathy / rapport | 0.091 | 0.005 | 0.702 | 0.787 | 0.003 | 0.139 | -0.015 | 0.372 | 0.446 |
| Concession offer | 0.091 | -0.002 | 0.852 | 0.870 | -0.001 | 0.137 | **0.061** | <0.001 | <0.001 |
| Payment plan | 0.167 | 0.033 | 0.057 | 0.116 | 0.034 | 0.146 | -0.017 | 0.328 | 0.425 |
| Funds-source suggestion | 0.284 | **0.075** | <0.001 | 0.002 | 0.072 | 0.279 | 0.008 | 0.716 | 0.747 |
| Downplays the amount | 0.195 | 0.038 | 0.022 | 0.068 | 0.042 | 0.204 | -0.017 | 0.394 | 0.451 |
| Commitment seeking | 0.371 | -0.008 | 0.724 | 0.787 | -0.010 | 0.448 | -0.007 | 0.812 | 0.812 |
| Follow-up arrangement | 0.194 | 0.003 | 0.870 | 0.870 | 0.003 | 0.224 | -0.036 | 0.098 | 0.213 |
| Responds to complaint | 0.082 | 0.019 | 0.038 | 0.096 | 0.025 | 0.097 | 0.009 | 0.455 | 0.496 |
| (4) Misconduct | | | | | | | | | |
| Deception / misrepresentation | 0.041 | 0.005 | 0.595 | 0.710 | 0.009 | 0.052 | 0.024 | 0.010 | 0.030 |
| Accuses of bad faith | 0.140 | 0.013 | 0.346 | 0.509 | 0.018 | 0.128 | 0.044 | 0.002 | 0.011 |
| Repeated pressure | 0.101 | 0.020 | 0.076 | 0.136 | 0.017 | 0.060 | -0.023 | 0.120 | 0.213 |
| Insult / abuse | 0.039 | -0.013 | 0.061 | 0.116 | -0.012 | 0.008 | -0.005 | 0.217 | 0.307 |
| *Panel B: Ratings (1–5)* | | | | | | | | | |
| Pressure | 2.547 | 0.126 | 0.018 | 0.090 | 0.150 | 2.531 | **0.315** | <0.001 | <0.001 |
| Professionalism | 4.032 | -0.055 | 0.365 | 0.467 | -0.089 | 4.298 | **-0.232** | <0.001 | <0.001 |
| Empathy | 2.129 | -0.025 | 0.455 | 0.467 | -0.021 | 2.403 | -0.057 | 0.203 | 0.253 |
| Negotiation | 2.022 | 0.034 | 0.467 | 0.467 | 0.040 | 2.311 | 0.039 | 0.440 | 0.440 |
| Borrower resistance | 2.982 | 0.081 | 0.040 | 0.101 | 0.102 | 2.709 | 0.066 | 0.146 | 0.243 |

*Notes*: For each sector, each row reports the mean of the row's outcome in the unweighted estimation sample ("Mean") and the UJIVE coefficient on %*social* ("Coef.") from a separate regression of the outcome on the instrumented treatment, with the headline borrower controls, calendar-month fixed effects, and equal-weighted instrument, rescaled to a 20-percentage-point (roughly one standard deviation in both sectors) increase in %*social*, with its $p$-value and the Benjamini and Hochberg (1995) false-discovery-rate adjusted $p$-value ("Adj. $p$"). Outcomes are measured on the borrower's first answered conversation with the caller on the borrower's own phone number during days 26–30, as labeled by the LLM annotation of the call transcripts and averaged over the borrower's transcribed first-conversation calls with the transcription inverse-probability weight: in Panel A, indicators for whether the caller strategy appears in the conversation, in the grouping of Figure 5; in Panel B, the 1–5 ratings of the caller's pressure, professionalism, empathy, and negotiation effort and of the borrower's resistance, defined in Panels A and B of Appendix Table B10. The sample comprises the borrowers in the headline recovery sample with a transcribed first conversation (14,251 in-house and 3,454 third-party); Panel B uses the rated calls (12,831 and 3,194). "Coef. (rewgt.)" re-estimates the in-house coefficient with the in-house borrowers reweighted to the third-party covariate distribution by XGBoost propensity odds (borrower and loan characteristics, calendar month, contact availability, the day-11 credit score, and days 1–25 pick-up behavior; trimmed at the in-house 99th percentile); its bold and underline marks use its own adjusted $p$-value, which is not reported. Standard errors are clustered at the collector level. The adjustment is applied within column and panel (Panel A's family also includes intimidation, omitted because it is not estimable in the third-party sample). Coefficients are in bold when the adjusted $p$-value is below 0.01 and underlined when it is below 0.05.

Table B12: Social-contact calling and the borrower's responses in the first conversation

| | In-house | | | | | Third-party | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Coef. | $p$ | Adj. $p$ | Coef. (rewgt.) | Mean | Coef. | $p$ | Adj. $p$ |
| *Panel A: Borrower moves* | | | | | | | | | |
| Promise to pay (specific) | 0.159 | -0.002 | 0.880 | 0.919 | 0.004 | 0.209 | -0.014 | 0.461 | 0.630 |
| Promise to pay (vague) | 0.354 | 0.018 | 0.285 | 0.499 | 0.017 | 0.460 | -0.037 | 0.082 | 0.287 |
| Partial commitment | 0.041 | 0.005 | 0.460 | 0.613 | 0.003 | 0.070 | 0.003 | 0.805 | 0.866 |
| Requests relief | 0.119 | 0.031 | 0.003 | 0.016 | 0.034 | 0.187 | 0.025 | 0.212 | 0.371 |
| Cites hardship | 0.490 | 0.012 | 0.482 | 0.613 | 0.013 | 0.610 | 0.031 | 0.202 | 0.371 |
| Disputes the debt | 0.085 | 0.011 | 0.254 | 0.499 | 0.016 | 0.107 | 0.029 | 0.058 | 0.272 |
| Asks for information | 0.095 | 0.009 | 0.364 | 0.567 | 0.015 | 0.139 | 0.043 | 0.003 | 0.040 |
| Asks caller's identity | 0.159 | 0.007 | 0.617 | 0.720 | 0.013 | 0.154 | 0.028 | 0.054 | 0.272 |
| Refuses | 0.159 | 0.027 | 0.012 | 0.042 | 0.027 | 0.135 | 0.011 | 0.507 | 0.630 |
| Stalls / avoids | 0.172 | 0.001 | 0.919 | 0.919 | -0.004 | 0.203 | 0.022 | 0.204 | 0.371 |
| Hostile / insults | 0.084 | 0.030 | 0.002 | 0.015 | 0.035 | 0.053 | -0.001 | 0.934 | 0.934 |
| Threatens to complain | 0.063 | 0.010 | 0.169 | 0.394 | 0.013 | 0.057 | 0.015 | 0.134 | 0.371 |
| Asks to stop contact | 0.078 | 0.028 | 0.002 | 0.015 | 0.027 | 0.086 | -0.007 | 0.540 | 0.630 |
| Complains about conduct | 0.196 | 0.030 | 0.029 | 0.081 | 0.038 | 0.210 | -0.014 | 0.410 | 0.630 |
| *Panel B: Borrower emotions* | | | | | | | | | |
| Valence (mean) | -0.185 | -0.022 | 0.023 | 0.047 | -0.025 | -0.153 | 0.020 | 0.053 | 0.301 |
| Intensity (mean) | 1.252 | 0.035 | 0.015 | 0.045 | 0.043 | 1.199 | -0.025 | 0.100 | 0.301 |
| Any emotion: anger | 0.223 | **0.049** | 0.001 | 0.007 | **0.059** | 0.175 | -0.002 | 0.886 | 0.886 |
| Any emotion: fear | 0.066 | 0.007 | 0.382 | 0.382 | 0.007 | 0.091 | -0.002 | 0.877 | 0.886 |
| Any emotion: sadness | 0.361 | 0.031 | 0.063 | 0.095 | 0.035 | 0.433 | 0.015 | 0.579 | 0.868 |
| Any emotion: shame | 0.021 | 0.009 | 0.098 | 0.117 | 0.007 | 0.029 | 0.006 | 0.450 | 0.868 |

*Notes*: For each sector, each row reports the mean of the row's outcome in the estimation sample and the UJIVE coefficient on %*social* from a separate regression of the outcome on the instrumented treatment, with the same specification, sample, column layout, rescaling to a 20-percentage-point increase in %*social*, reweighted in-house column, and $p$-value conventions as Appendix Table B11. Outcomes are the borrower's side of the first answered conversation on the borrower's own phone number during days 26–30: in Panel A, indicators for whether the borrower move appears in the conversation; in Panel B, the mean valence ($-2$ to 2) and intensity (1 to 5) of the borrower's turns and indicators for whether any of the borrower's turns expresses each emotion, defined for the conversations in which the borrower speaks (about 94% of the sample). The codes and measures are defined in Panels C and D of Appendix Table B10. The Benjamini–Hochberg adjustment is applied within column and panel. Coefficients are in bold when the adjusted $p$-value is below 0.01 and underlined when it is below 0.05.

Table B13: Social-contact calling and the caller's conduct on calls to the borrower's social contacts

| | In-house | | | | | Third-party | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Coef. | $p$ | Adj. $p$ | Coef. (rewgt.) | Mean | Coef. | $p$ | Adj. $p$ |
| *Panel A: Collector strategies* | | | | | | | | | |
| (1) Information | | | | | | | | | |
| Identity verification | 0.541 | 0.033 | 0.162 | 0.424 | 0.038 | 0.872 | -0.001 | 0.969 | 0.969 |
| States the debt | 0.458 | 0.026 | 0.212 | 0.424 | 0.029 | 0.487 | -0.045 | 0.012 | 0.044 |
| Asks reason / funds | 0.027 | 0.006 | 0.184 | 0.424 | 0.006 | 0.060 | 0.002 | 0.758 | 0.842 |
| Payment how-to | 0.031 | 0.003 | 0.525 | 0.656 | 0.004 | 0.052 | 0.001 | 0.854 | 0.884 |
| (2) Threats and consequences | | | | | | | | | |
| Legal threat | 0.111 | 0.030 | 0.030 | 0.186 | 0.038 | 0.169 | **0.085** | <0.001 | <0.001 |
| Escalation threat | 0.035 | -0.007 | 0.188 | 0.424 | -0.006 | 0.051 | 0.022 | 0.005 | 0.023 |
| Credit warning | 0.107 | 0.027 | 0.031 | 0.186 | 0.030 | 0.186 | **0.057** | <0.001 | 0.003 |
| Cost warning | 0.008 | -0.000 | 0.952 | 0.952 | -0.000 | 0.023 | 0.002 | 0.674 | 0.842 |
| Vague consequence | 0.047 | 0.008 | 0.329 | 0.580 | 0.009 | 0.052 | 0.011 | 0.127 | 0.293 |
| Social-contact warning | 0.036 | 0.004 | 0.489 | 0.656 | 0.004 | 0.028 | 0.002 | 0.755 | 0.842 |
| Urgency / deadline | 0.223 | 0.055 | 0.001 | 0.037 | 0.058 | 0.198 | 0.035 | 0.060 | 0.163 |
| Moral appeal | 0.046 | 0.006 | 0.362 | 0.603 | 0.007 | 0.054 | 0.005 | 0.495 | 0.782 |
| (3) Negotiation and rapport | | | | | | | | | |
| Empathy / rapport | 0.021 | 0.002 | 0.676 | 0.811 | 0.001 | 0.028 | -0.005 | 0.258 | 0.517 |
| Concession offer | 0.012 | -0.005 | 0.083 | 0.327 | -0.004 | 0.035 | 0.008 | 0.199 | 0.425 |
| Payment plan | 0.010 | 0.002 | 0.442 | 0.656 | 0.002 | 0.011 | -0.001 | 0.709 | 0.842 |
| Funds-source suggestion | 0.025 | 0.012 | 0.010 | 0.100 | 0.012 | 0.017 | -0.004 | 0.300 | 0.563 |
| Downplays the amount | 0.080 | -0.001 | 0.943 | 0.952 | 0.002 | 0.072 | **-0.048** | <0.001 | <0.001 |
| Commitment seeking | 0.010 | -0.002 | 0.407 | 0.643 | -0.002 | 0.010 | -0.001 | 0.624 | 0.842 |
| Follow-up arrangement | 0.148 | 0.011 | 0.463 | 0.656 | 0.013 | 0.065 | **-0.032** | 0.001 | 0.004 |
| Responds to complaint | 0.061 | 0.016 | 0.048 | 0.239 | 0.012 | 0.045 | 0.011 | 0.047 | 0.141 |
| (4) Misconduct | | | | | | | | | |
| Deception / misrepresentation | 0.038 | 0.010 | 0.197 | 0.424 | 0.013 | 0.041 | **0.022** | <0.001 | 0.001 |
| Accuses of bad faith | 0.082 | 0.003 | 0.739 | 0.821 | 0.006 | 0.114 | **0.042** | 0.001 | 0.003 |
| Repeated pressure | 0.029 | 0.009 | 0.087 | 0.327 | 0.009 | 0.006 | 0.001 | 0.734 | 0.842 |
| Insult / abuse | 0.024 | 0.001 | 0.840 | 0.901 | 0.001 | 0.004 | 0.001 | 0.631 | 0.842 |
| Intimidation | 0.002 | -0.000 | 0.724 | 0.821 | -0.001 | 0.000 | -0.000 | 0.490 | 0.782 |
| (5) Contact-specific | | | | | | | | | |
| Asks contact to relay | 0.546 | 0.024 | 0.290 | 0.544 | 0.024 | 0.463 | -0.013 | 0.487 | 0.782 |
| Asks about the borrower | 0.386 | 0.016 | 0.519 | 0.656 | 0.004 | 0.410 | 0.007 | 0.805 | 0.862 |
| Explains why contact is called | 0.330 | 0.030 | 0.173 | 0.424 | 0.022 | 0.381 | 0.012 | 0.577 | 0.842 |
| Asks contact to pay | 0.018 | 0.012 | 0.002 | 0.037 | 0.013 | 0.008 | -0.005 | 0.113 | 0.283 |
| Discloses the debt to the contact | 0.419 | 0.024 | 0.212 | 0.424 | 0.026 | 0.435 | -0.042 | 0.014 | 0.047 |
| *Panel B: Ratings (1–5)* | | | | | | | | | |
| Pressure | 2.023 | 0.132 | 0.011 | 0.027 | 0.151 | 1.915 | **0.158** | <0.001 | 0.002 |
| Professionalism | 3.886 | -0.057 | 0.242 | 0.403 | -0.077 | 3.952 | -0.030 | 0.468 | 0.585 |
| Empathy | 1.999 | -0.005 | 0.616 | 0.769 | -0.005 | 2.040 | -0.019 | 0.011 | 0.027 |
| Negotiation | 1.351 | 0.007 | 0.782 | 0.782 | 0.013 | 1.322 | -0.021 | 0.365 | 0.585 |
| Contact resistance | 3.104 | **0.107** | <0.001 | <0.001 | **0.097** | 2.919 | 0.015 | 0.588 | 0.588 |

*Notes*: Same specification, column layout, rescaling to a 20-percentage-point increase in %*social*, reweighted in-house column, and $p$-value conventions as Appendix Table B11, with the outcomes now measured on the caller's transcribed conversations with the borrower's *social contacts* during days 26–30 (the stage-1 annotation sample), averaged over the borrower's transcribed contact calls with the transcription inverse-probability weight. The sample comprises the borrowers in the headline recovery sample with at least one transcribed contact call (12,335 in-house and 19,277 third-party borrowers); Panel B uses the rated calls (12,038 and 18,116 borrowers). The reweighted column reweights the in-house borrowers to the covariate distribution of the third-party borrowers in this sample. Groups (1)–(4) of Panel A repeat the codes of Appendix Table B11, here including intimidation, which is estimable in both sectors on contact calls; group (5) lists the codes defined only for contact calls: whether the caller asks the contact to pass on a message or have the borrower call back, asks the contact about the borrower's whereabouts or situation, explains why the contact is being called, asks the contact to pay on the borrower's behalf, and whether the caller states the debt to the contact (disclosure). The Benjamini–Hochberg adjustment is applied within column and panel. Coefficients are in bold when the adjusted $p$-value is below 0.01 and underlined when it is below 0.05.

Table B14: Social-contact calling and the contact's responses on calls to the borrower's social contacts

| | In-house | | | | | Third-party | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Coef. | $p$ | Adj. $p$ | Coef. (rewgt.) | Mean | Coef. | $p$ | Adj. $p$ |
| *Panel A: Contact's moves* | | | | | | | | | |
| (1) General | | | | | | | | | |
| Promise to pay (specific) | 0.010 | -0.002 | 0.563 | 0.919 | -0.001 | 0.010 | 0.000 | 0.943 | 0.943 |
| Promise to pay (vague) | 0.017 | -0.000 | 0.919 | 0.919 | -0.001 | 0.020 | -0.008 | 0.073 | 0.657 |
| Partial commitment | 0.002 | 0.000 | 0.730 | 0.919 | 0.000 | 0.001 | 0.000 | 0.756 | 0.943 |
| Requests relief | 0.007 | 0.001 | 0.465 | 0.919 | 0.001 | 0.009 | -0.000 | 0.924 | 0.943 |
| Cites hardship | 0.027 | 0.005 | 0.292 | 0.748 | 0.005 | 0.039 | 0.001 | 0.878 | 0.943 |
| Asks for information | 0.071 | 0.001 | 0.897 | 0.919 | 0.001 | 0.140 | -0.006 | 0.472 | 0.943 |
| Asks caller's identity | 0.161 | -0.022 | 0.009 | 0.054 | <u>-0.026</u> | 0.185 | 0.005 | 0.648 | 0.943 |
| Refuses | 0.018 | 0.001 | 0.736 | 0.919 | 0.003 | 0.018 | -0.001 | 0.721 | 0.943 |
| Stalls / avoids | 0.024 | -0.002 | 0.628 | 0.919 | -0.001 | 0.030 | -0.004 | 0.468 | 0.943 |
| Hostile / insults | 0.077 | <u>0.019</u> | 0.004 | 0.036 | 0.017 | 0.045 | 0.001 | 0.792 | 0.943 |
| Threatens to complain | 0.035 | 0.010 | 0.012 | 0.056 | 0.006 | 0.021 | 0.004 | 0.302 | 0.943 |
| Asks to stop contact | 0.077 | <u>0.023</u> | 0.001 | 0.024 | <u>0.020</u> | 0.059 | 0.004 | 0.461 | 0.943 |
| Complains about conduct | 0.102 | 0.008 | 0.333 | 0.748 | 0.004 | 0.090 | -0.002 | 0.758 | 0.943 |
| (2) Contact-specific | | | | | | | | | |
| Confirms knowing the borrower | 0.139 | 0.003 | 0.772 | 0.919 | 0.005 | 0.257 | -0.006 | 0.653 | 0.943 |
| Cannot reach the borrower | 0.190 | -0.002 | 0.877 | 0.919 | -0.004 | 0.190 | -0.006 | 0.674 | 0.943 |
| Denies knowing the borrower | 0.362 | 0.005 | 0.732 | 0.919 | 0.002 | 0.427 | 0.030 | 0.012 | 0.212 |
| Agrees to relay | 0.231 | -0.012 | 0.286 | 0.748 | -0.012 | 0.278 | -0.020 | 0.128 | 0.766 |
| Refuses to relay | 0.108 | 0.017 | 0.037 | 0.134 | 0.015 | 0.094 | 0.004 | 0.545 | 0.943 |
| *Panel B: Contact's emotions* | | | | | | | | | |
| Valence (mean) | -0.150 | **-0.035** | <0.001 | <0.001 | **-0.032** | -0.087 | -0.002 | 0.677 | 0.735 |
| Intensity (mean) | 1.220 | **0.055** | <0.001 | <0.001 | **0.050** | 1.123 | 0.003 | 0.735 | 0.735 |
| Any emotion: anger | 0.167 | **0.029** | 0.005 | 0.010 | <u>0.024</u> | 0.120 | 0.004 | 0.669 | 0.735 |
| Any emotion: fear | 0.016 | -0.003 | 0.432 | 0.518 | -0.002 | 0.018 | -0.007 | 0.041 | 0.244 |
| Any emotion: sadness | 0.049 | -0.002 | 0.746 | 0.746 | -0.003 | 0.051 | 0.005 | 0.352 | 0.703 |
| Any emotion: shame | 0.002 | 0.001 | 0.372 | 0.518 | 0.001 | 0.002 | -0.002 | 0.129 | 0.386 |

*Notes*: Same specification, sample, column layout, rescaling to a 20-percentage-point increase in %*social*, reweighted in-house column, and $p$-value conventions as Appendix Table B13. Outcomes are the contact's side of the caller's transcribed conversations with the borrower's social contacts during days 26–30: in Panel A, indicators for whether the move appears in the conversation, group (1) repeating the borrower moves of Appendix Table B12 (disputing the debt is not coded on contact calls) and group (2) listing the moves defined only for contact calls: whether the contact confirms knowing the borrower, says the borrower cannot be reached, denies knowing the borrower, and agrees or refuses to relay the message; in Panel B, the mean valence ($-2$ to 2) and intensity (1 to 5) of the contact's turns and indicators for whether any of the contact's turns expresses each emotion, defined for the conversations in which the contact speaks (12,209 in-house and 19,035 third-party borrowers). The Benjamini–Hochberg adjustment is applied within column and panel. Coefficients are in bold when the adjusted $p$-value is below 0.01 and underlined when it is below 0.05.