Who Borrows from Whom? Market Segmentation in Consumer Loans

Rosina Rodríguez Olivera* Graciela Sanroman†

Abstract

This paper develops a model of credit market screening with heterogeneous lenders, borrower default risk, and regulatory interest rate caps. Two lenders compete by offering menus of loan contracts, and differ in their ability to enforce repayment and in their marginal lending costs. Borrowers are heterogeneous in their repayment probability and in their valuation of credit, both of which decline with borrower risk. In equilibrium, borrowers self-select across lenders according to type, generating endogenous market segmentation. We show that interest rate caps distort optimal contract menus, inducing bunching, flattening repayment schedules, and shifting the cutoff type that determines which borrowers are served by each lender. The model rationalizes two robust empirical regularities: (i) non-bank lenders concentrate at the regulatory ceiling, while banks offer strictly lower rates, and (ii) default rates are systematically higher among non-bank borrowers. We further characterize the welfare-maximizing cap, highlighting the regulator's trade-off between broader borrower coverage and sustaining lender participation. Our results underscore how enforcement heterogeneity interacts with borrower risk to shape the effectiveness and unintended consequences of interest rate regulation.

1. Introduction

Consumer credit markets are central to household finance and welfare. Access to credit enables households to smooth consumption, absorb shocks, and invest in opportunities. Over the past two decades, consumer credit has expanded dramatically across both advanced and emerging economies. This expansion has been characterized by high interest rates and

*Universidad de la República, Uruguay
†Universidad de la República, Uruguay

elevated default rates, raising concerns about efficiency and welfare. The growth reflects a larger supply of credit from traditional banks and, increasingly, the rise of *fintech* lenders and other non-traditional intermediaries. The coexistence of fintechs and banks has produced a highly heterogeneous supply side, with different lender types segmenting borrowers in novel ways. At the same time, regulators often impose interest-rate caps aimed at protecting borrowers from predatory practices. These caps interact with lender heterogeneity in ways that shape market segmentation, loan terms, and ultimately welfare. Understanding this interaction is essential for designing policies that protect borrowers while sustaining lender participation.

This paper develops a model of credit market screening with two competing lenders who face an interest rate cap imposed by a regulator. We incorporate two key asymmetries: (i) lenders vary in their efficiency of repayment collection and marginal cost of providing funds, and (ii) borrowers are heterogeneous in their repayment probability and in their valuation of credit.

Each lender offers a menu of contracts, consisting of a probability of approval and an interest rate. Borrowers are characterized by a uni-dimensional type in the unit interval, which represents their riskiness and jointly determines their valuation and their probability of repayment. In particular, we assume that the repayment probability and their valuation of a loan are decreasing in their riskiness level, introducing adverse selection. Borrowers self-select among the menu offered by the lenders based on their type and preferences. A regulator imposes an interest rate cap that constrains lenders' ability to screen borrowers efficiently. We characterize the equilibrium menus of contracts and analyze how asymmetries in enforcement shape the allocation of borrowers across lenders.

Our analysis highlights how regulatory caps induce distortions in contract design and lead to segmentation across lenders. In particular, we show that non-bank lenders with weaker enforcement tend to concentrate at the cap, while banks with stronger enforcement offer lower rates and attract safer borrowers. We derive the welfare-maximizing cap and show that the regulator faces a fundamental trade-off: a lower cap increases borrower coverage but discourages lender participation. These findings contribute to the ongoing debate on financial inclusion, consumer protection, and the unintended consequences of interest rate regulation. Our framework contributes by unifying borrower and lender heterogeneity, producing new predictions about segmentation, default, and discouragement.

Our framework generalizes classical screening models by combining heterogeneous lenders with repayment risk and regulatory constraints, offering a flexible foundation for evaluating credit market interventions in settings with heterogeneous agents on both sides of the market.

This framework contributes to the literature in three ways. First, it unifies insights from the consumption-smoothing and credit-rationing literatures by incorporating both borrower heterogeneity and lender diversity. Second, it generates new testable predictions about borrower–lender matching, credit terms, default, and discouragement patterns. Third, it provides a richer foundation for analyzing policy interventions—such as interest-rate caps, credit bureau regulations, or fintech entry—by evaluating their effects in a segmented market where both sides of the market interact strategically.

Related Literature. Our paper contributes to several strands of the literature in economics and finance.

First, it builds on the classic screening models of consumer credit markets under asymmetric information, such as Rothschild and Stiglitz [1978], Stiglitz and Weiss [1981] and Besanko and Thakor [1987]. The classical framework of Stiglitz and Weiss [1981] established that riskier projects promise higher expected returns but also a greater probability of default. Under asymmetric information, lenders cannot perfectly distinguish between risk types, so higher interest rates worsen the pool of applicants by discouraging safer borrowers and attracting riskier ones. This adverse selection mechanism, combined with moral hazard, implies that credit markets may not clear through prices alone, leading to credit rationing.

While the notions of return and risk are also relevant in the context of household demand for consumer credit, they rest on fundamentally different pillars. Our paper departs from the investment–project framework and instead provides a rationalization of consumer credit demand based on a simple intertemporal consumption model. In this setting, households are represented as a continuum of borrowers with heterogeneous loan valuations, whose credit demand arises endogenously from the trade-off between present consumption, future repayment, and potential penalties of default. This demand side interacts with multiple types of lenders endowed with distinct screening technologies, thereby generating segmentation in modern consumer credit markets. Furthermore, we incorporate heterogeneity on both sides of the market: not only are borrowers differentiated by repayment probability, but lenders also differ in enforcement capacity and cost efficiency. This dual heterogeneity generates novel implications for market segmentation and regulatory design.

Within the consumer-credit literature on endogenous default and information frictions, our work is related to Chatterjee et al. [2008], who develop a finite-horizon model where repayment histories signal borrower risk in other markets such as insurance, rationalizing the role of credit scores [Chatterjee et al., 2007]. Unlike their cross-market signaling channel, our focus is on within-market screening under lender heterogeneity and regulatory caps. Their mechanism is reputational, while ours is static and emphasizes segmentation and regulation, making the two approaches complementary. Our analysis is also related to recent work on search and approval in mortgage markets. Agarwal et al. [2024] develop a quantitative search model where application rejection alters borrower behavior, generating endogenous adverse selection and strategic complementarities in bank rate setting. Risky borrowers internalize high rejection probabilities and behave as if they face high search costs, sorting toward higher-priced lenders. While their focus is on search frictions and approval processes, ours emphasizes screening under lender heterogeneity and rate caps. Together, these papers highlight how competition in credit markets manifests not only through prices, but also through endogenous segmentation of borrowers shaped by enforcement, screening, and approval institutions.

A complementary strand of work highlights how borrowers themselves may engage in signaling. Kawai et al. [2022] study signaling in online peer-to-peer credit markets, showing that reserve interest rates reveal borrower creditworthiness and mitigate adverse selection. They estimate that adverse selection can destroy up to one-third of total surplus, much of which is restored through signaling. While their mechanism operates through borrower-side information disclosure, our focus is on lender heterogeneity and regulatory constraints. Together, these papers underscore multiple channels that shape adverse selection and equilibrium allocation in credit markets.

Second, our analysis relates to the literature on interest rate regulation and usury laws [e.g., Dehejia et al., 2012]. These studies document that caps are widespread but often controversial: while intended to protect borrowers from predatory lending, caps may reduce credit access, especially for high-risk borrowers. Our model provides a theoretical foundation for these trade-offs, showing precisely how caps alter contract menus and shift the borrower composition across lenders.

Third, we connect to recent work on heterogeneous lenders and the rise of non-bank financial intermediaries. Empirical evidence shows that banks and non-banks differ systematically in their cost structures, risk management technologies, and collection capabilities. We formalize such differences as variation in enforcement efficiency and show how they interact with regulatory caps to produce equilibrium patterns consistent with observed differences in interest rates and default outcomes. A growing literature emphasizes the role of financial technology in this evolution. Chu and Wei [2024] analyze a model where Fintech entrants compete with banks using alternative screening technologies. They show that superior Fintech screening may worsen allocative efficiency and even reduce welfare. In contrast, Fuster et al. [2019] document empirically that Fintech mortgage lenders process applications faster, respond more elastically to demand shocks, and increase refinancing efficiency, without higher defaults. Our model complements these insights by showing how lender heterogeneity—in enforcement as opposed to technology—interacts with regulatory caps to shape borrower segmentation and welfare.

Our work is also related to the macro–finance literature on consumer credit and default. In particular, Chatterjee et al. [2007] develop a dynamic model of unsecured credit with endogenous default, showing how heterogeneity in borrower risk and income shocks shapes equilibrium loan terms. Subsequent work, such as Chatterjee and Eyigungor [2012], incorporates limited commitment and richer credit contract environments to analyze the design of credit markets and bankruptcy institutions. These papers emphasize the dynamic interaction between household risk, default, and contract structure, whereas our focus is on static screening with lender heterogeneity and regulatory caps. In this sense, our model complements theirs by isolating how enforcement asymmetries and price regulation affect borrower sorting and market segmentation in equilibrium.

Finally, our work is related to the literature on screening in vertical oligopolies, such as Chade and Swinkels [2021], who analyze how competing principals design contracts in the presence of private information on the agent side. In a similar spirit, we study how competition between heterogeneous lenders leads to segmentation and cutoff rules in equilibrium, but extend the analysis by incorporating repayment enforcement heterogeneity and regulatory price constraints. Lester et al. [2019] develop a search-theoretic model that embeds adverse selection and nonlinear screening into a frictional market with imperfect competition. Their framework emphasizes how search frictions and market power jointly determine equilibrium menus, allocations, and welfare, and shows that more competition need not be welfare improving. By contrast, our analysis abstracts from search frictions and instead highlights how lender heterogeneity and regulatory caps shape contract design and borrower sorting. More broadly, our paper complements the literature on consumer protection and welfare in financial markets. While much of this work emphasizes behavioral biases and disclosure policies [e.g., Campbell, 2016], our focus is on the equilibrium consequences of price regulation in the presence of adverse selection and enforcement asymmetries. We show that a uniform cap may exacerbate risk concentration among non-bank lenders, highlighting the need for policy instruments that take into account the structural heterogeneity of credit markets.

The remainder of the paper is organized as follows: Section 2 introduces the model. Section 3 characterizes the optimal menu of financial products for lenders in the absence of regulation. Section 4 derives the optimal menu when a regulator is able to impose an interest rate cap. Section 5 derives the welfare optimal interest rate cap. Section 6 discusses the policy implications of the model. Section 7 discusses some extensions. Lastly, Section 8 concludes.

2. Model

We consider a market with two competing lenders and a continuum of heterogeneous borrowers. Lenders differ in cost and collection efficiency. Each lender $i \in \{A, B\}$ has a repayment collection efficiency $\rho_i \in (0, 1]$ and a per-type operating cost function $C_i : [0, 1] \to \mathbb{R}_+$ that is twice-continuously differentiable and strictly convex. Without loss of generality, assume that lender A has higher repayment collection efficiency, $\rho_A > \rho_B$. Assume that there is a regulator who imposes an upper bound $\bar{r}$ on the interest rate that any lender can charge.

Each lender offers a menu of contracts $\{(q_i(\theta), r_i(\theta))\}_{\theta\in[0,1]}$, where $q_i : [0,1] \to [0,1]$ represents the probability of the credit approval whereas $r_i : [0,1] \to \mathbb{R}_+$ denotes the interest rate associated to such credit. Assume that all credits are for the same amount and borrowers have the same initial endowment. The lender's objective is to maximize its expect profits subject to standard participation and incentive compatibility constraints.[1] Furthermore, for all $q \in [0,1]$, the higher-enforcement lender A has weakly lower marginal cost per unit of enforcement than B:

$$\frac{C_A'(q)}{\rho_A} \leq \frac{C_B'(q)}{\rho_B}, \qquad \text{with } \rho_A > \rho_B,\ C_i \in C^2,\ C_i'' > 0.$$

Borrowers are privately informed about their types, which determine both their valuation and their probability of repayment. Borrowers are indexed by their type $\theta \in [0,1]$, distributed according to the CDF F on $[0,1]$ with density f with full support. Borrowers' valuation is determined by a decreasing and concave function $v(\theta)$. The repayment probability $p(\theta)$ is decreasing in θ with $p(\theta) > 0$, capturing the notion that riskier borrowers are less creditworthy. Assume that f, p and v are continuously differentiable on $[0,1]$.

Payoffs. The utility a borrower of type θ obtains from a contract $(q_i(\hat{\theta}), r_i(\hat{\theta}))$ offered by lender i is given by:

$$U_i(\theta, \hat{\theta}) = q_i(\hat{\theta}) \left[v(\theta) - r_i(\hat{\theta}) \right].$$

The participation constraint of borrower of type θ requires that $U_i(\theta, \theta) := U_i(\theta) \geq 0$. The incentive compatibility constraint requires that $U_i(\theta) \geq U_i(\theta, \hat{\theta}))$ for all $\theta, \hat{\theta}$. Standard single-crossing assumptions and the envelope condition implies that $U_i'(\theta) = v'(\theta) q_i(\theta)$ a.e..

Lender i's expected repayment from a type-θ borrower is $\rho_i p(\theta) r_i(\theta)$, so that expected profit per borrower is

$$\pi_i(\theta) = \rho_i p(\theta) r_i(\theta) q_i(\theta) - C_i(q_i(\theta)).$$

Lender's problem. Using the indirect utility representation, it follows that:

$$r_i(\theta) = v(\theta) - \frac{U_i(\theta)}{q_i(\theta)},$$

[1] Higher q_i may also reflect better customer service, reputational advantages, or more favorable loan terms beyond the interest rate.

whenever $q_i(\theta) > 0$. Substituting into the profit function:

$$\begin{aligned}\pi_i(\theta) &= \rho_i p(\theta)\left(v(\theta) - \frac{U_i(\theta)}{q_i(\theta)}\right) q_i(\theta) - C_i(q_i(\theta)) \\ &= \rho_i p(\theta)\left(v(\theta) q_i(\theta) - U_i(\theta)\right) - C_i(q_i(\theta)).\end{aligned}$$

Under standard assumptions guaranteeing incentive compatibility, the utility function is differentiable and:

$$v'(\theta) q_i(\theta) = U_i'(\theta),$$

so that:

$$\pi_i(\theta) = \rho_i p(\theta)\left(\frac{v(\theta)}{v'(\theta)} U_i'(\theta) - U_i(\theta)\right) - C_i\left(\frac{U_i'(\theta)}{v'(\theta)}\right).$$

Using this transformation of the problem, and letting Θ_i denote the subset of borrowers who buy from lender i, each lender chooses an incentive-compatible utility schedule $U_i(\cdot)$ to maximize expected profits:

$$\max_{U_i(\cdot)} \int_{\Theta_i} \left[\rho_i p(\theta)\left(\frac{v(\theta)}{v'(\theta)} U_i'(\theta) - U_i(\theta)\right) - C_i\left(\frac{U_i'(\theta)}{v'(\theta)}\right)\right] f(\theta)\, \mathrm{d}\theta,$$

subject to:

$$\begin{aligned}
&\text{IC:} && U_i'(\theta) \leq 0,\\
&\text{IR:} && U_i(\theta) \geq 0,\\
&\text{Cap:} && r_i(\theta) = v(\theta) - \frac{v'(\theta) U_i(\theta)}{U_i'(\theta)} \leq \bar{r}.
\end{aligned}$$

Market Segmentation. Let $\Theta_A \subset [0,1]$ denote the set of borrower types who strictly prefer lender A, and let $\Theta_B \subset [0,1]$ denote the set of borrower types who strictly prefer lender B and let $\overline{\theta}_i$ be the smallest type in Θ_i.

Regulator's problem. Assume that there is a regulator who imposes an upper bound $\bar{r}$ on the interest rate that any lender can charge. The regulator chooses the interest rate cap $\bar{r}$ to maximize welfare, subject to the equilibrium constraints that determine contract menus, borrower sorting, and participation. In this context, total welfare is:

$$W(\bar{r}) = \sum_{i \in \{A,B\}} \int_{\theta \in \Theta_i} \left[U_i(\theta) + \pi_i(\theta)\right] d\theta.$$

The regulator's problem is then:

$$\max_{\bar{r}} W(\bar{r}) \quad \text{subject to all equilibrium constraints.}$$

This framework captures the interaction between asymmetric lender characteristics, borrower heterogeneity in both willingness to pay and default risk, and regulatory constraints on pricing. The interest rate cap introduces nontrivial distortions to both the structure of the menus and the selection patterns across lenders. We show how changes in $\bar{r}$ affect screening incentives, cross-subsidization, and the allocation of borrowers, highlighting the trade-offs between protection and efficiency in regulated credit markets.

Binding constraints. We first identify the binding constraints.

Lemma 1 (Binding Participation Constraint) *In any equilibrium, the participation constraint binds for the highest type served by each lender. That is,*

$$U_i(\overline{\theta}_i) = 0 \ \forall \ i \in \{A, B\}.$$

2.1 Rationalizing the Valuation Function

A central assumption in our framework is that the gross valuation function $v(\theta)$ is decreasing in borrower type θ. This property can be rationalized by linking it to the to the certainty-equivalent and the standard Euler equation governing intertemporal consumption under uncertainty and default. Intuitively, with prudence, higher-risk borrowers face both a lower repayment probability and greater consumption risk, which jointly reduce their certainty-equivalent willingness to pay for credit.

Formally, consider a two-period consumer who maximizes expected utility $u(c_0)+\beta\mathbb{E}[u(c_1)]$, with $u' > 0$, $u'' < 0$, and prudence $u''' > 0$. Borrower type θ parameterizes risk, with higher θ associated with a lower probability of repayment and greater dispersion in future consumption.

Suppose that the borrower receives a marginal loan of size $\varepsilon > 0$ at date 0 and faces uncertainty about repayment at date 1. If repayment occurs, which happens with probability $p(\theta)$, future consumption becomes

$$c_1^{\text{rep}} = c_1^0 - (1 + r)\varepsilon.$$

If default occurs, which happens with probability $1 - p(\theta)$, future consumption is

$$c_1^{\text{def}} = c_1^0 - \kappa\varepsilon,$$

where $\kappa \in [0, 1+r)$ is a constant utility-equivalent penalty. Expected utility is then

$$\mathbb{E}[u(c_1) \mid \theta] = p(\theta)u(c_1^{\text{rep}}) + (1 - p(\theta))u(c_1^{\text{def}}).$$

The ex-ante marginal utility impact of the loan can be written exactly as

$$\beta\, \mathbb{E}\Big[u'(c_1)\Big((1+r)\mathbf{1}_{\text{rep}} + \kappa\, \mathbf{1}_{\text{def}}\Big) \,\Big|\, \theta\Big] = \beta\Big[(1{+}r)p(\theta)\, \mathbb{E}\big[u'(c_1) \mid \text{rep}, \theta\big] {+} \kappa\big(1{-}p(\theta)\big)\, \mathbb{E}\big[u'(c_1) \mid \text{def}, \theta\big]\Big],$$

where rep and def denote the repayment and default events at date 1, respectively.

Under the standard "small-loan" first-order approximation,

$$\mathbb{E}\big[u'(c_1) \mid \text{rep}, \theta\big] \;\approx\; \mathbb{E}\big[u'(c_1) \mid \text{def}, \theta\big] \;\approx\; \mathbb{E}[u'(c_1) \mid \theta],$$

and linearity of expectation then yields

$$\beta\, \mathbb{E}[u'(c_1) \mid \theta]\, \Big[(1+r)p(\theta) + \kappa\big(1 - p(\theta)\big)\Big].$$

Using this, the marginal change in ex-ante utility from a loan of size ε is approximately

$$\frac{\Delta U(\varepsilon; \theta)}{\varepsilon} \approx\; u'(c_0) - \beta\mathbb{E}[u'(c_1) \mid \theta]\, \Big[(1+r)\, p(\theta) + \kappa\big(1 - p(\theta)\big)\Big].$$

We define the gross valuation function $v(\theta)$ as the certainty-equivalent marginal willingness to pay for loan access. Formally,

$$v(\theta) := \frac{1}{u'(c_0)}\, \frac{\partial}{\partial \varepsilon}\Big\{u(c_0 + \varepsilon) + \beta\big[p(\theta)u(c_1^0 - (1+r)\varepsilon) + (1 - p(\theta))u(c_1^0 - \kappa\varepsilon)\big]\Big\}\Big|_{\varepsilon=0}.$$

We then obtain

$$u'(c_0)\, v(\theta) \;\approx\; u'(c_0) \;-\; \beta\, \mathbb{E}[u'(c_1) \mid \theta]\, \Big[(1+r)\, p(\theta) + \kappa\big(1 - p(\theta)\big)\Big]$$

or

$$v(\theta) \;\approx\; 1 - \beta\, \frac{\mathbb{E}[u'(c_1) \mid \theta]}{u'(c_0)}\, \Big[(1+r)\, p(\theta) + \kappa\big(1 - p(\theta)\big)\Big].$$

This expression highlights why $v(\theta)$ naturally declines with borrower risk. First, with prudence, higher θ increases the dispersion of future consumption, raising $\mathbb{E}[u'(c_1)]$. Second, because $p'(\theta) < 0$, the effective repayment burden $(1{+}r)p(\theta){+}\kappa(1{-}p(\theta))$ is weakly increasing in θ. Taken together, these effects imply that the certainty-equivalent willingness to pay for loan access decreases with borrower type. Hence, the assumption $v'(\theta) < 0$ is consistent with standard intertemporal choice theory under risk and default, and provides a natural behavioral foundation for declining valuations in our screening model.

Note that if instead the penalty κ depends on θ, then the risk-adjusted repayment burden becomes $\tilde{R}(\theta) = (1+r)p(\theta) + \kappa(\theta)(1-p(\theta))$. In that case, $v(\theta)$ will fall even more sharply whenever $\kappa'(\theta) \geq 0$, since both a lower repayment probability and a harsher default penalty reinforce the decline. Conversely, if $\kappa(\theta)$ decreases in θ, the risk-adjusted obligation could flatten or partly offset the prudence effect. Nonetheless, $v(\theta)$ remains decreasing under mild and natural conditions. Specifically, differentiating the expression above gives

$$v'(\theta) \;\approx\; -\beta \left[\frac{\partial_\theta \mathbb{E}[u'(c_1) \mid \theta]}{u'(c_0)}\, \tilde{R}(\theta) + \frac{\mathbb{E}[u'(c_1) \mid \theta]}{u'(c_0)}\, \tilde{R}'(\theta) \right].$$

Thus, a sufficient condition for $v'(\theta) < 0$ is

$$\frac{\partial_\theta \mathbb{E}[u'(c_1) \mid \theta]}{\mathbb{E}[u'(c_1) \mid \theta]}\, \tilde{R}(\theta) + \tilde{R}'(\theta) > 0.$$

This inequality states that the precautionary effect (an increase in expected marginal utility with θ due to prudence) is large enough to outweigh any decline in the risk-adjusted repayment burden $\tilde{R}(\theta)$. Under these mild conditions—which require only prudence and that risk increases with type—the assumption $v'(\theta) < 0$ is robust even when default penalties vary with borrower type.

3. Optimal menu without interest rate cap

Given the results on incentive compatibility and participation, we can restrict attention to utility functions $U_i(\cdot)$ that are differentiable and satisfy the envelope condition $U_i'(\theta) = v'(\theta)q_i(\theta)$, with participation binding at the highest type served. Thus, the lender's problem can be rewritten as follows. Letting $\Theta_i = [\underline{\theta}_i, \overline{\theta}_i] \subseteq [0,1]$ denote the set of types served by lender i, the lender chooses a utility function $U_i : \Theta_i \to \mathbb{R}_+$ satisfying $U_i(\overline{\theta}_i) = 0$ and monotone to maximize:

$$\max_{U_i(\cdot)} \int_{\underline{\theta}_i}^{\overline{\theta}_i} \left[\rho_i p(\theta) \left(\frac{v(\theta)}{v'(\theta)} U_i'(\theta) - U_i(\theta) \right) - C_i \left(\frac{U_i'(\theta)}{v'(\theta)} \right) \right] f(\theta)\, \mathrm{d}\theta,$$

Proposition 1 (Optimal Menu Without Interest Rate Cap) *In the absence of an interest rate cap, the optimal menu offered by lender i satisfies the following properties:*

1. ***Repayment Schedule:*** *The interest rate charged to type θ is:*

$$r_i(\theta) = v(\theta) - \frac{1}{q_i(\theta)} \int_{\theta}^{\bar{\theta}_i} q_i(s)v'(s)\mathrm{d}s.$$

2. ***Optimality Condition:*** *Let*

$$\Phi_i(\theta) \;:=\; \frac{1}{f(\theta)} \int_{\underline{\theta}_i}^{\theta} \rho_i \, p(t) \, f(t) \, dt.$$

The optimal approval schedule is given by

$$\begin{aligned} q_i(\theta) = 0 \quad &\textit{when} \quad C_i'(0) \;\geq\; \rho_i \, p(\theta) \, v(\theta) \;-\; v'(\theta) \, \Phi_i(\theta), \\ q_i(\theta) = 1 \quad &\textit{when} \quad C_i'(1) \;\leq\; \rho_i \, p(\theta) \, v(\theta) \;-\; v'(\theta) \, \Phi_i(\theta), \\ q_i(\theta) \in (0,1) \quad &\textit{when} \quad C_i'\big(q_i(\theta)\big) \;=\; \rho_i \, p(\theta) \, v(\theta) \;-\; v'(\theta) \, \Phi_i(\theta). \end{aligned}$$

The sketch of the proof is as follows. Incentive compatibility ensures that the utility function is differentiable and monotone, with the approval schedule given by $q_i(\theta)v'(\theta) = U_i'(\theta)$. In the absence of regulatory constraints or bunching, strict monotonicity of q_i holds generically. Part (1) follows directly from the definition of repayment, which is determined by the envelope condition. Part (2) derives from applying the calculus of variations to the lender's objective. Taking the variational derivative and integrating by parts yields the equation that characterizes the optimal approval function $q_i(\cdot)$.

Borrowers with higher values of θ are riskier in the sense that their probability of repayment is lower. In the optimal screening arrangement, incentive compatibility requires that each borrower type strictly prefers the contract intended for it. If more favorable terms were offered to riskier borrowers, safer borrowers would have a strict incentive to deviate and select those contracts, thereby undermining the screening mechanism.

The single-crossing property of the payoff functions then implies a monotone allocation: borrowers with lower θ (safer borrowers) must be offered weakly more favorable terms (that is, either a higher approval rate $q_i(\theta)$ or a lower interest rate $r_i(\theta)$) while borrowers with higher θ (riskier borrowers) must face tighter contractual terms. Consequently, in equilibrium:

1. the approval probability $q_i(\theta)$ is weakly decreasing in θ, reflecting the lender's reduced willingness to extend credit as repayment risk increases; and

2. the interest rate $r_i(\theta)$ is weakly increasing in θ, both to compensate for the higher expected default losses and to preserve the separation of borrower types.

Any deviation from these monotonicity patterns would violate the incentive compatibility constraints: for example, an increasing $q_i(\theta)$ or a decreasing $r_i(\theta)$ would induce safer borrowers to mimic riskier ones, making the allocation infeasible. Monotonicity thus emerges as a necessary structural property of the optimal contract, ensuring self-selection and preserving the feasibility of the screening equilibrium. This is formalized in Corollary 1.

Corollary 1 (Monotonicity of the optimal contract) *In the optimal contract, $q_i(\theta)$ is weakly decreasing on $[\underline{\theta}_i, \overline{\theta}_i]$ whereas $r_i(\theta)$ is weakly increasing on such interval.*

Sorting Across Lenders. To determine how borrowers are allocated between the two lenders in equilibrium, we first establish a *single-crossing property* in the difference between their approval schedules. The logic is standard in screening environments: if the difference $q_A(\theta) - q_B(\theta)$ is weakly increasing in borrower type θ, then by the envelope condition the difference in borrower utilities

$$\Delta U(\theta) = U_A(\theta) - U_B(\theta)$$

is also single crossing in θ.

Lemma 2 (Single crossing of approval schedules) *The function $\theta \mapsto q_A(\theta) - q_B(\theta)$ is weakly decreasing on the union of types served in equilibrium.*

The single-crossing property implies that borrowers' preferences over lenders can change at most once as θ increases, giving rise to a unique *cutoff type* θ^* that separates the two market segments. Let θ^* denote the marginal borrower who is indifferent between the two lenders, defined by

$$U_A(\theta^*) = U_B(\theta^*) \quad \Leftrightarrow \quad q_A(\theta^*)\big[v(\theta^*) - r_A(\theta^*)\big] = q_B(\theta^*)\big[v(\theta^*) - r_B(\theta^*)\big].$$

Equivalently,

$$v(\theta^*) = \frac{q_A(\theta^*) r_A(\theta^*) - q_B(\theta^*) r_B(\theta^*)}{q_A(\theta^*) - q_B(\theta^*)}.$$

In equilibrium, all borrowers with $\theta < \theta^*$ are served by one lender and all borrowers with $\theta \geq \theta^*$ are served by the other. Accordingly, for each $i \in \{A, B\}$, the set Θ_i of types served by lender i takes the form

$$\Theta_i = \left[\underline{\theta}_i, \overline{\theta}_i\right],$$

where $\underline{\theta}_i \in \{0, \theta^*\}$ and $\overline{\theta}_i \in \{\theta^*, 1\}$.

While the single-crossing property guarantees the existence of a unique cutoff, it does not identify *which* lender serves the higher-risk borrowers. To pin this down, we show that under our maintained assumptions there is a *top-type feasibility dominance*: at sufficiently low risk, lender A can profitably serve borrowers whereas lender B cannot.

Lemma 3 (Low-type dominance) *At sufficiently low risk, lender B cannot serve the type profitably while lender A can.*

This property follows from the monotonicity of $p(\theta)$: for sufficiently low types, the repayment probability is high enough to satisfy A's break-even condition, while B's fails under the effective cost–efficiency ordering at safer types.

Proposition 2 (Sorting by enforcement) *There exists a unique cutoff type $\theta^* \in [0,1]$ such that*

$$\Theta_A = [\underline{\theta}_A, \theta^*) \quad \textit{and} \quad \Theta_B = [\theta^*, \overline{\theta}_B].$$

Competition in this environment arises not from Bertrand price undercutting but from the endogenous determination of the borrower segments that each lender serves. In the model, both lenders offer screening menus subject to participation and incentive compatibility constraints. Because lender A is more efficient in enforcement, it can profitably serve safer (low-θ) borrowers, while lender B focuses on riskier (high-θ) ones. This generates a unique cutoff type θ^* such that borrowers below θ^* strictly prefer lender A and those above θ^* strictly prefer lender B. The cutoff emerges endogenously from the interaction of approval schedules $q_i(\theta)$ and repayment rates $r_i(\theta)$, and rules out multiple switching points or overlapping service regions. Thus, market segmentation is the equilibrium manifestation of competition: each lender optimally adjusts its contract menu given the other's, and the resulting partition of borrowers reflects their relative enforcement efficiencies

4. Optimal menu with interest rate cap

Consider now the case in which the regulator introduces the interest rate cap.

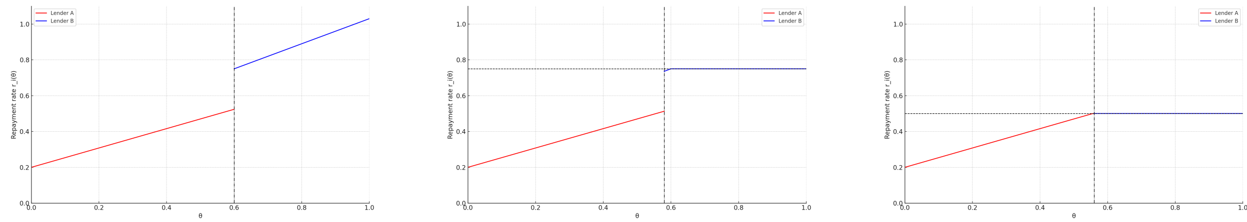
Proposition 3 (Characterization of the Optimal Menu) *In the optimal contract offered by lender i, the utility schedule $U_i(\theta)$ is convex and satisfies the following properties:*

1. ***Interest Rate Cap and Bunching:*** *There exists a (possibly empty) subset $\mathcal{B}_i \subseteq \Theta_i$ such that for all $\theta \in \mathcal{B}_i$, the interest rate cap binds:*

$$v(\theta) - \frac{U_i(\theta)}{q_i(\theta)} = \bar{r}.$$

   On this set, the optimal menu exhibits bunching: distinct types receive the same interest rate, and the lender adjusts quality to maintain incentive compatibility.

2. ***Strict Screening Outside the Cap:*** *Outside the bunching region $\mathcal{B}_i$, the menu is strictly screening and characterized by Proposition 1.*



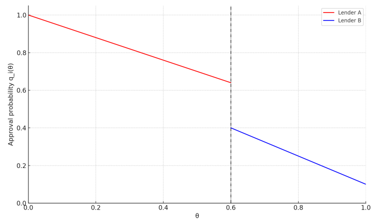
(a) Cap doesn't bind (b) Cap bind for B (c) Cap binds for both firms.

Figure 1: Interest rate schedules under different regulatory cap scenarios.

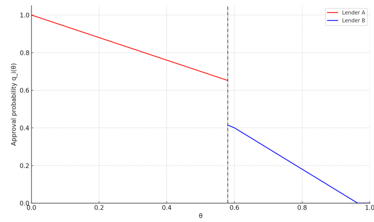
The result follows from standard arguments in optimal control and mechanism design. Property (1) arises from the fact that the interest rate cap may prevent the lender from fully separating borrower types through pricing. When the cap binds, the lender is constrained from increasing repayment further, so all types $\theta \in \mathcal{B}_i$ must be offered the same interest rate. To preserve incentive compatibility in this region, the lender adjusts $q_i(\theta)$, resulting in bunching. Property (2) reflects the standard screening logic in unconstrained regions: types are fully separated, and Proposition 1 characterizes the optimal approval schedule whenever the cap constraint is slack.

The interest rate cap $\bar{r}$ may bind for one or both lenders, depending on its tightness. If $\bar{r}$ is not too restrictive, both lenders are able to offer strictly screening menus with full separation of borrower types. However, if the cap binds for lender B, it induces bunching among higher types: since borrowers cannot be charged higher repayment, the lender must distort the probability of approval downward to preserve incentive compatibility. If the cap is sufficiently tight, it may lead to pooling of types, drive one lender out of the market, or significantly reduce total borrower surplus. In general, the cap flattens contract menus, reduces the extent of sorting, and shifts the cutoff type θ^* that determines which lender a borrower selects.

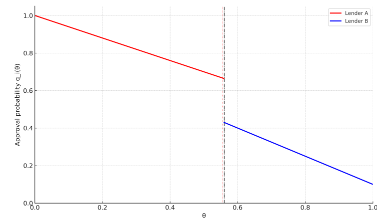
Figures 1 and 2 capture the central mechanism by which interest rate caps distort the structure of credit contracts. When the cap is slack, lenders fully separate borrower types through smoothly increasing repayment rates and decreasing approval probabilities. Once the cap becomes binding, however, this separation breaks down: repayment schedules flatten at the ceiling, and lenders are forced to screen borrowers through reduced approval instead of higher charges. These distortions compress differences across borrower types, reallocate risk toward the less efficient lender, and shift the cutoff type that determines who borrows from whom. In this way, the figures illustrate how regulation transforms competitive screening into a segmented market where bunching, exclusion, and misallocation naturally emerge.



(a) Cap doesn't bind



(b) Cap bind for B



(c) Cap binds for both lenders.

Figure 2: Approval rate schedules under different regulatory cap scenarios.

5. Welfare optimal interest cap

We now characterize the interest rate cap $\bar{r}^*$ that maximizes total welfare, accounting for its effects on contract menus, borrower sorting, and lender participation. The regulator's objective is to choose the cap that optimally balances access to credit, borrower surplus, and lender profitability, given the distortions induced by the cap on screening and selection.

Regulator's Problem. Let $\theta^*(\bar{r})$ denote the marginal borrower who is indifferent between lenders under interest cap $\bar{r}$. Then, total welfare as a function of the cap is given by:

$$W(\bar{r}) = \int_0^{\theta^*(\bar{r})} [U_A(\theta;\bar{r}) + \pi_A(\theta;\bar{r})] \, d\theta + \int_{\theta^*(\bar{r})}^1 [U_B(\theta;\bar{r}) + \pi_B(\theta;\bar{r})] \, d\theta.$$

The regulator chooses $\bar{r} \in [0, \bar{r}_{\max}]$ to maximize $W(\bar{r})$, subject to all equilibrium constraints: that is, the incentive compatibility and participation constraints, the interest rate cap constraint, the borrower sorting condition and the lender non-negative profit condition.

Optimal Cap. We next derive the first-order condition characterizing $\bar{r}^*$.

Proposition 4 (Optimal Interest Rate Cap) *Suppose the equilibrium menu is unique and continuously differentiable in $\bar{r}$, and that $\theta^*(\bar{r})$ is differentiable. Then the optimal cap $\bar{r}^*$ satisfies the first-order condition:*

$$\frac{dW(\bar{r})}{d\bar{r}} = 0.$$

The first-order condition reflects three components:

1. The *direct effect* of relaxing the cap on the menus and utilities offered to borrowers currently served by each lender.

2. The *profit effect* through changes in lender margins due to loosening or tightening the rate constraint.

3. The *extensive margin effect* due to re-sorting of the marginal borrower θ^*, who may switch lenders in response to a change in the cap.

If the cap is too low, credit access is restricted and menus are distorted by bunching. If the cap is too high, lenders serve excessively risky borrowers at high repayment rates, increasing default and harming borrower welfare. The optimal cap $\bar{r}^*$ balances these competing forces to maximize overall surplus.

6. Discussion

This section discusses how the model rationalizes two robust empirical regularities in credit markets with interest rate regulation and heterogeneous lenders:

1. non-bank lenders tend to charge the maximum interest rate permitted by regulation, whereas banks typically offer lower rates; and

2. non-bank lenders exhibit systematically higher default rates than banks.

We show that both patterns naturally emerge from the model's equilibrium, reflecting the interaction between lender heterogeneity, borrower risk, and regulatory caps. We also address the conditions under which non-banks remain active in equilibrium, compare the relative size and profitability of lenders, and consider the broader policy implications of the analysis.

These regularities can be traced back to fundamental differences in lenders' enforcement capacity. Banks, endowed with greater repayment efficiency—through superior screening technologies, stronger legal enforcement, or reputational advantages—are able to profitably serve safer borrowers at lower rates. Non-banks, by contrast, face limited collection power and consequently design flatter menus that cluster at the regulatory ceiling.

6.1 Non-bank participation and clustering at the regulatory cap

Proposition 3 shows that the interest rate cap may bind on a (possibly nonempty) set of borrower types, depending on the lender's characteristics and the tightness of the cap. When the cap binds, lenders cannot fully separate borrower types through pricing. To maintain incentive compatibility, they respond by reducing approval probabilities (bunching), which flattens repayment rates at the regulatory ceiling and distorts contract terms over the binding region.

Because less efficient lenders face higher effective marginal costs of serving risky borrowers, they require higher rates to break even and are therefore more likely to be constrained by the cap. In equilibrium, the optimal contract offered by a low-efficiency lender typically features a binding interest rate constraint over a wide range of borrower types, leading them to set rates at the ceiling for most, if not all, clients. By contrast, high-efficiency lenders—banks in the empirical counterpart—can profitably offer credit at lower rates to safer borrowers. Their menus involve strictly interior rates below the cap, particularly in the region where screening is active. Hence, the model predicts a pattern consistent with the data: non-bank lenders cluster at the regulatory ceiling, while banks offer more dispersed and typically lower rates.

Although non-banks are dominated in efficiency, they remain active in equilibrium because they serve a segment of the borrower population that banks optimally choose not to cover. Banks, endowed with superior enforcement capacity, could in principle supply credit to the entire market. However, extending credit to riskier borrowers would require raising rates for safer clients as well, thereby eroding informational rents and reducing profitability in their core portfolio. It is therefore optimal for banks to restrict their menus to lower-risk borrowers, leaving the riskier segment to non-banks. As a result, non-banks specialize in the upper tail of the risk distribution. Their lower enforcement efficiency makes them more likely to be constrained by the regulatory cap, but their comparative advantage lies precisely in their willingness to extend credit to borrowers that banks deliberately exclude.

6.2 Borrower Sorting and Default Rates across lenders

The model embeds adverse selection through a decreasing repayment probability function: riskier borrowers are more likely to default, and equilibrium borrower–lender matching is determined by endogenous sorting based on the menus offered by each lender.

Because of their limited collection efficiency, non-bank lenders must design flatter menus that implicitly subsidize high-risk borrowers within the cap constraint. In equilibrium, this strategy attracts a disproportionate share of borrowers from the upper tail of the risk distribution. Banks, by contrast, exploit their higher enforcement capacity to profitably serve lower-risk borrowers while optimally excluding excessively risky types. The resulting cutoff that separates lender portfolios is interior whenever both lenders are active, with safer borrowers allocated to banks and riskier borrowers self-selecting into non-bank contracts.

Taken together, these equilibrium patterns generate systematic differences in borrower composition and repayment outcomes. Non-banks concentrate at the regulatory ceiling and serve a pool with higher default incidence, while banks offer lower and more dispersed rates to safer clients. Since repayment probabilities decline with borrower risk, the model predicts that non-bank portfolios will exhibit substantially higher default rates. These two empirical regularities—non-banks clustering at the ceiling and experiencing higher defaults—are thus complementary manifestations of a single mechanism: banks' incentive to preserve favorable terms for safer borrowers leads them to exclude riskier clients, who are then absorbed by non-banks under the binding cap.

6.3 Profits, market shares and policy implications

The analysis underscores the importance of accounting for enforcement heterogeneity when designing interest rate regulation. A uniform cap compresses contract menus and reallocates riskier borrowers toward lenders with weaker enforcement capacity. This segmentation preserves access but simultaneously amplifies default externalities and erodes overall credit quality.

In equilibrium, both banks and non-banks can earn positive profits, but their size and composition differ. Banks capture the larger, safer segment of the market by offering lower rates supported by stronger enforcement technologies, while non-banks concentrate on a smaller, riskier pool of borrowers. Non-banks' profitability does not stem from efficiency advantages but from tailoring contracts to borrowers whom banks optimally exclude. Their smaller market share reflects this specialization in serving borrowers who would otherwise lack access to credit.

From a policy perspective, these findings highlight a fundamental trade-off. A uniform interest rate cap sustains the coexistence of heterogeneous lenders but at the cost of concentrating risk among non-banks. Differentiated caps, subsidies for enforcement technologies, or complementary regulatory instruments could mitigate these distortions by reducing adverse sorting and improving credit quality. More broadly, the model suggests that the presence of non-banks should not be viewed exclusively as evidence of predatory practices, but rather as the equilibrium outcome of structural segmentation in which banks optimally relinquish the riskiest borrowers.

7. Extensions

7.1 Fixed costs

We extend the baseline framework by allowing each lender $i \in \{A, B\}$ to face fixed enforcement costs in addition to variable costs and repayment efficiency. In particular, in addition to the per-type operating cost function $C_i(q)$ and repayment efficiency parameter ρ_i, each lender is subject to: (i) a per-loan activation cost $f_i \geq 0$, incurred whenever $q_i(\theta) > 0$; and (ii) a per-default fixed enforcement cost $\rho_i^F \geq 0$, which is paid only in default states.

Note that the repayment efficiency parameter ρ_i measures the proportional fraction of repayments that can be effectively collected, scaling revenues smoothly across all types. By contrast, the activation cost f_i and the per-default fixed cost ρ_i^F capture nonconvex enforcement frictions: f_i arises whenever a loan is originated, while ρ_i^F is triggered only in default states. The activation cost f_i reflects overhead expenses of initiating a loan (screening, paperwork, monitoring), while the per-default fixed cost ρ_i^F reflects the legal and administrative burden of pursuing a delinquent borrower. Accordingly, ρ_i affects marginal efficiency continuously, whereas (f_i, ρ_i^F) create discrete extensive-margin cutoffs and wedges that grow in importance for high-risk borrowers.

Lender's problem. In the presence of fixed costs, lender i's profits per type θ can therefore be expressed as

$$\pi_i(\theta) = \rho_i p(\theta) r_i(\theta) q_i(\theta) - C_i(q_i(\theta)) - (1 - p(\theta))\rho_i^F q_i(\theta) - f_i \mathbf{1}\{q_i(\theta) > 0\}.$$

As in the baseline, repayment is given by

$$r_i(\theta) = v(\theta) - \frac{U_i(\theta)}{q_i(\theta)} \quad \text{whenever } q_i(\theta) > 0,$$

and incentive compatibility implies $U_i'(\theta) = v'(\theta)q_i(\theta)$ with $U_i(\theta_i) = 0$ at the highest type served. We continue to define

$$\Phi_i(\theta) := \frac{1}{f(\theta)} \int_{\underline{\theta}_i}^{\theta} \rho_i p(t) f(t)\, dt.$$

The characterization of the optimal menu with fixed costs is analogous to that in the baseline model. The only change is that the optimality conditions include, for each type, an additional constraint: expected profits (net of variable costs) must at least cover the fixed cost of approving a loan and must also account for the contingent enforcement cost in default.

Proposition 5 (Optimal menu with fixed costs) *In any optimal menu,*

$$\begin{aligned}
q_i(\theta) = 0 \text{ when } C_i'(0) \;&\geq\; \rho_i p(\theta) v(\theta) - v'(\theta)\Phi_i(\theta),\\
q_i(\theta) = 1 \text{ when } C_i'(1) + (1 - p(\theta))\rho_i^F \;&\leq\; \rho_i p(\theta) v(\theta) - v'(\theta)\Phi_i(\theta),\\
q_i(\theta) \in (0,1),\; r_i(\theta) < \bar{r} \text{ when } C_i'(q_i(\theta)) + (1 - p(\theta))\rho_i^F &= \rho_i p(\theta) v(\theta) - v'(\theta)\Phi_i(\theta),\\
q_i(\theta) \in (0,1),\; r_i(\theta) = \bar{r} \text{ when } C_i'(q_i(\theta)) + (1 - p(\theta))\rho_i^F &= \rho_i p(\theta) v(\theta) - v'(\theta)\Phi_i(\theta) + \lambda_i(\theta),
\end{aligned}$$

where $\lambda_i(\theta) \geq 0$ is the multiplier on the cap constraint. Moreover, if $q_i(\theta) > 0$, then

$$\rho_i p(\theta)\big(v(\theta)q_i(\theta) - U_i(\theta)\big) - C_i(q_i(\theta)) - (1 - p(\theta))\rho_i^F q_i(\theta) \;\geq\; f_i.$$

Note that the proof of Proposition 5 is analogous to the baseline case, and the additional term $(1 - p(\theta))\,\rho_i^F$ acts as a contingent marginal cost, steepening the effective cost schedule—especially for higher-θ types. This observation has a direct implication for the allocation under the cap. In particular, the presence of fixed costs implies that some types may be unserved (i.e., $q_i(\theta) = 0$) even within the bunching region. Define, within the lender's market Θ_i, the cap–binding set $B_i = \{\theta \in \Theta_i : \; r_i(\theta) = \bar{r}\}$ and the served set $S_i = \{\theta \in \Theta_i : \; q_i(\theta) > 0\}$. Moreover, whenever $B_i \neq \emptyset$ we define

$$\theta_c^i \;:=\; \inf B_i,$$

i.e., the smallest type in Θ_i at which the cap binds. Note that monotonicity implies that $B_i = [\theta_c^i, \sup \Theta_i]$. Lemma 4 shows that only sufficiently risky types are potentially excluded in equilibrium, with the extent of exclusion determined by the fixed-cost parameters.

Lemma 4 (Terminal exclusion in the cap region) *If $B_i \neq \emptyset$, then there exists $\tilde{\theta}_i \in B_i$ such that $S_i \cap B_i = [\theta_c^i, \tilde{\theta}_i)$ and $q_i(\theta) = 0$ for all $\theta > \tilde{\theta}_i$.*

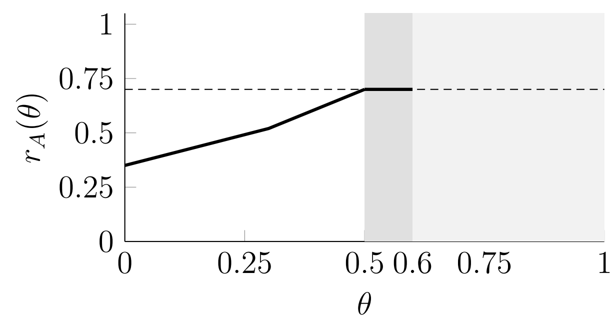
Figure 3 illustrates how fixed costs affect the optimal menus, illustrating how sufficiently risky types may be excluded from the market. The top panels plot $r_i(\theta)$ and show that the schedule rises with type until it reaches the cap and then remains flat at $\bar{r}$ for all higher served types. The bottom panels plot $q_i(\theta)$ and show a weakly decreasing approval schedule that, with fixed costs, may hit zero once at a cutoff inside B_i, beyond which the lender does not lend. Types outside Θ_i are shaded to emphasize market segmentation. Hence, the right column illustrates exclusion for lender B, while the left column shows the no-exclusion case for lender A.

No exclusion. To identify conditions under which exclusion does not arise, define for $\theta \in B_i$ the capped linear coefficient

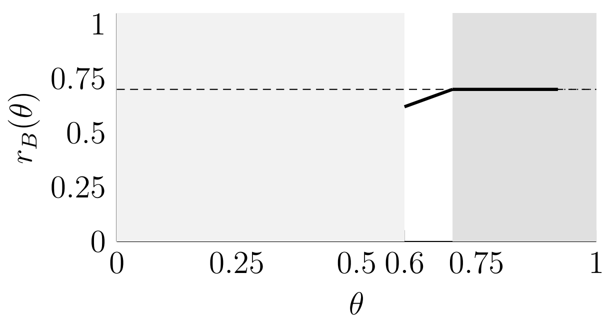
$$\alpha_i(\theta) := \rho_i p(\theta)\,\bar{r} - (1 - p(\theta))\,\rho_i^F.$$

Inside B_i the cap fixes $r_i(\theta) = \bar{r}$ for all served types, so the per–type net margin at approval level q equals $\alpha_i(\theta)q - C_i(q)$ and must cover the cost f_i. Let $\underline{\alpha}_i := \inf_{\theta \in B_i} \alpha_i(\theta)$. A simple uniform requirement then rules out exclusion: if

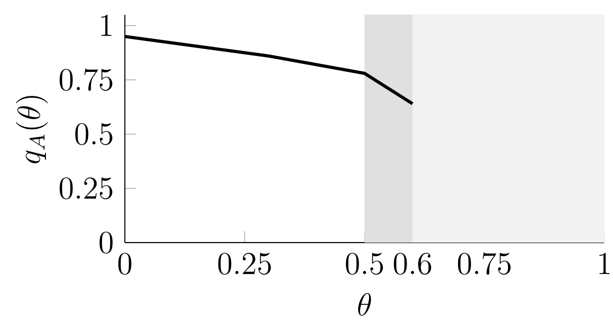
$$\max_{q \in [0,1]} \{\underline{\alpha}_i q - C_i(q)\} \;\geq\; f_i,$$



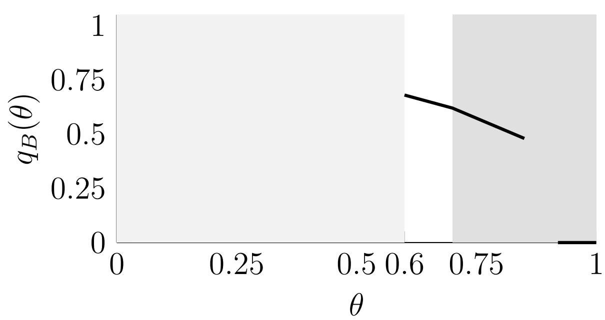
(a) Lender A is constrained by the cap.



(b) Lender B is also constrained by the cap.



(c) Lender A serves all types below θ^*.



(d) Lender B excludes sufficiently risky types.

Figure 3: Each panel is restricted to the lender's own Θ_i and shades types outside it. Darker bands mark $B_i = \{\theta \in \Theta_i : \; r_i(\theta) = \bar{r}\}$.

there exists $q^\star > 0$ with $\underline{\alpha}_i q^\star - C_i(q^\star) \geq f_i$, and since $\alpha_i(\theta) \geq \underline{\alpha}_i$ for all $\theta \in B_i$ it follows that $\alpha_i(\theta)q^\star - C_i(q^\star) \geq f_i$ pointwise on B_i, implying $q_i(\theta) > 0$ throughout B_i.

Lemma 5 (Sufficient condition for no exclusion) *If* $\max_{q\in[0,1]}\{\underline{\alpha}_i q - C_i(q)\} \geq f_i$*, then* $q_i(\theta) > 0$ *for all* $\theta \in B_i$*.*

Formally, within the cap-binding set the interest rate is fixed at the regulatory cap, so prices cannot adjust and all screening occurs through quantities. For any borrower type in this set, approving an additional unit yields a cap-adjusted marginal revenue: the lender's collection efficiency multiplied by the probability of repayment and by the cap, minus the probability of default multiplied by the fixed enforcement cost per default. Net surplus at a given approval level equals this cap-adjusted marginal revenue times the approved quantity, minus variable operating costs, and it must cover the per-loan activation cost. No exclusion obtains if a uniform activation test is satisfied: there exists a single approval level—strictly between zero and full approval—such that, even for the worst type in the cap-binding set (the type with the lowest cap-adjusted marginal revenue), the resulting net surplus is at least as large as the activation cost. When this condition holds, the same approval level is profitable for every other type in the cap-binding set (since their cap-adjusted marginal revenue is no lower), and the lender therefore approves a positive amount for all such types. In particular, no borrower type in the cap-binding set is excluded.

Comparative statics. Because the fixed–cost parameters determine which borrowers are excluded, we study how the exclusion boundary responds to changes in these parameters. Within the cap–binding set B_i the interest rate is fixed at $\bar{r}$, so screening operates entirely through quantities. For $\theta \in B_i$, the default–contingent term $(1 - p(\theta))\,\rho_i^F$ lowers $\alpha_i(\theta)$ more for riskier types, while the per–loan cost f_i imposes a uniform threshold: at any approval level q, the capped margin $\alpha_i(\theta)q - C_i(q)$ must cover f_i. As θ rises (so $p(\theta)$ falls), both forces depress $q_i(\theta)$; once $q_i(\theta)$ reaches zero at a cutoff $\tilde{\theta}_i \in B_i$, incentive–compatibility implies a monotone $q_i(\cdot)$, so approval cannot resume at higher types and exclusion beyond $\tilde{\theta}_i$ is terminal.

Lemma 6 (Comparative statics of terminal exclusion) *The terminal cutoff* $\tilde{\theta}_i$ *is weakly decreasing in* ρ_i^F *and in* f_i*. Equivalently, the terminal exclusion region weakly expands as either parameter increases.*

A formal intuition for Lemma 6 is as follows. Inside B_i the price is fixed at the cap, so the lender's choice for each type reduces to a one–dimensional quantity problem with an activation constraint: approve $q \in [0, 1]$ only if the cap–adjusted surplus at that type exceeds the per–loan cost, and otherwise do not lend. Increasing the per–default fixed cost

ρ_i^F lowers the cap–adjusted surplus pointwise for every type, and it does so more sharply at higher risk levels because the probability of default is larger there. This contraction in surplus tightens the activation constraint exactly where approval is already most difficult, thereby removing the highest–risk served types and shifting the terminal boundary $\tilde{\theta}_i$ toward lower types. Increasing the activation cost f_i raises the threshold uniformly across types; holding everything else fixed, fewer types now clear the activation test, so the served set inside B_i shrinks from the top for the same reason. Because incentive compatibility implies that the set of served types within B_i is a connected lower interval, any reduction in the set of types satisfying the activation constraint can only move the upper endpoint of that interval downward. Hence $\tilde{\theta}_i$ is weakly decreasing in both ρ_i^F and f_i, and the terminal exclusion region weakly expands as either parameter increases.

Welfare. In the presence of fixed costs, total welfare is given by

$$W(\bar{r}) = \sum_{i\in\{A,B\}} \int_{\Theta_i} \Big[v(\theta)\, q_i(\theta) - C_i\big(q_i(\theta)\big) - (1-p(\theta))\, \rho_i^F\, q_i(\theta) - f_i\, \mathbf{1}\{q_i(\theta)>0\} \Big]\, f(\theta)\, d\theta.$$

The welfare optimal interest cap with this cost structure is characterized in Proposition 6.

Proposition 6 (Optimal interest-rate cap with fixed costs) *Suppose the equilibrium menu is differentiable in $\bar{r}$ on each Θ_i, the type distribution places no atoms on B_i, and the sorting cutoff θ^* and the terminal cutoffs $\tilde{\theta}_i$ vary smoothly with $\bar{r}$. Let $\lambda_i(\theta)\geq 0$ denote the multiplier on $r_i(\theta)\leq \bar{r}$, and define*

$$w_i(\theta;\bar{r}) := v(\theta)q_i(\theta) - C_i(q_i(\theta)) - (1-p(\theta))\rho_i^F q_i(\theta) - f_i\, \mathbf{1}\{q_i(\theta)>0\}.$$

Then

$$\frac{dW(\bar{r})}{d\bar{r}} = \underbrace{\sum_i \int_{B_i\cap S_i} \lambda_i(\theta)\, f(\theta)\, d\theta}_{\textit{direct effect of relaxing the binding cap}} + \underbrace{\sum_i w_i(\tilde{\theta}_i^-;\bar{r})\, f(\tilde{\theta}_i)\, \frac{d\tilde{\theta}_i}{d\bar{r}}}_{\textit{effect of moving the terminal approval boundary}} + \underbrace{\Delta_{sort}(\bar{r})}_{\textit{effect of shifting the market cutof}}$$

In particular, any interior optimum $\bar{r}^\star$ satisfies $\frac{dW(\bar{r}^\star)}{d\bar{r}}=0$ together with the usual regularity and second–order conditions; otherwise the optimum lies at a boundary of the feasible cap.

The three terms have a clear interpretation. The first term aggregates, over types who are both served and constrained by the cap, the gain from a marginal relaxation of the cap; because $\lambda_i(\theta)\geq 0$, this contribution is nonnegative. The second term reflects how a change in $\bar{r}$ alters the size of the served set within the cap-binding region: a higher cap raises the cap-adjusted margin (since $\partial\alpha_i(\theta)/\partial\bar{r} = \rho_i p(\theta) > 0$), shifts the terminal cutoff $\tilde{\theta}_i$ weakly outward, and changes welfare at a rate equal to the welfare density at that boundary times the rate at which the boundary moves; this term is weakly nonnegative whenever $w_i(\tilde{\theta}_i^-; \bar{r}) \geq 0$. The third term captures reallocation at the market cutoff, with a local representation

$$\Delta_{\text{sort}}(\bar{r}) = \big[w_B(\theta^*; \bar{r}) - w_A(\theta^*; \bar{r})\big]\, f(\theta^*)\, \frac{d\theta^*}{d\bar{r}},$$

so its sign depends on which lender delivers higher welfare at the margin and on whether the cutoff moves toward that lender as the cap changes.

These elements imply the following policy guidance. If, over the range where the cap binds, welfare at the terminal boundary is nonnegative for each lender and the reallocation effect is favorable (i.e., $\Delta_{\text{sort}}(\bar{r}) \geq 0$ on a set of positive measure), then $\frac{dW(\bar{r})}{d\bar{r}} > 0$ and welfare rises as the cap is relaxed until the constraint ceases to bind; the optimal cap is therefore the least restrictive (nonbinding) one. If boundary welfare is nonpositive and the reallocation effect is unfavorable, then $\frac{dW(\bar{r})}{d\bar{r}} < 0$ on the binding range and the most restrictive feasible cap is optimal. In intermediate cases the three forces offset each other, and the welfare-maximizing cap is interior and determined by the first-order condition in Proposition 6; its level depends on the shadow values $\lambda_i(\theta)$, the boundary welfare $w_i(\tilde{\theta}_i^-; \bar{r})$, the responsiveness of the terminal cutoffs $d\tilde{\theta}_i/d\bar{r}$, and the product $\big[w_B(\theta^*; \bar{r}) - w_A(\theta^*; \bar{r})\big]\,(d\theta^*/d\bar{r})$ at the market margin.

7.2 Interest Rate-Dependent Repayment Probabilities

We now extend the framework by allowing the repayment probability to depend not only on borrower type but also on the contractual interest rate. Formally, let $p(\theta, r)$ denote the repayment probability of a borrower of type $\theta \in [0, 1]$ facing an interest rate r. We assume that p is continuously differentiable, strictly decreasing in θ, and weakly decreasing in r. This modification reflects the idea that riskier borrowers are less likely to repay and that higher repayment burdens increase the incidence of default.

The borrower's problem remains unchanged. Given a contract $(q_i(\hat{\theta}), r_i(\hat{\theta}))$, a borrower of type θ obtains

$$U_i(\theta, \hat{\theta}) = q_i(\hat{\theta})\big[v(\theta) - r_i(\hat{\theta})\big].$$

As in the baseline model, incentive compatibility implies the envelope condition

$$U_i'(\theta) = v'(\theta) q_i(\theta),$$

with participation binding at the highest type served. Hence, approval rates remain weakly decreasing in type and repayment rates weakly increasing. The modification arises on the supply side: lender i's expected profit from serving type θ now takes the form

$$\pi_i(\theta) = \rho_i \, p(\theta, r_i(\theta)) \, r_i(\theta) \, q_i(\theta) - C_i(q_i(\theta)),$$

where ρ_i denotes enforcement efficiency. Substituting the utility representation, profits can be expressed as

$$\pi_i(\theta) = \rho_i \, p(\theta, r_i(\theta)) \Big(\frac{v(\theta)}{v'(\theta)} U_i'(\theta) - U_i(\theta) \Big) - C_i \Big(\frac{U_i'(\theta)}{v'(\theta)} \Big),$$

with $r_i(\theta) = v(\theta) - U_i(\theta) v'(\theta) / U_i'(\theta)$. The lender maximizes the expected value of this expression over the set of types it serves, subject to monotonicity.

The next result characterizes the structure of optimal contracts in this extended setting without an interest rate cap.

Proposition 7 (Optimal Menu with Rate-Dependent Repayment) *In the absence of an interest rate cap, the optimal contract offered by lender i satisfies:*

1. *For any type θ in the support of Θ_i, repayment is given by*

$$r_i(\theta) = v(\theta) - \frac{1}{q_i(\theta)} \int_{\theta}^{\bar{\theta}_i} q_i(s) v'(s) \, ds.$$

2. *Defining*

$$\Phi_i(\theta) = \frac{1}{f(\theta)} \int_{\underline{\theta}_i}^{\theta} \rho_i p(t, r_i(t)) f(t) \, dt.$$

the approval schedule $q_i(\theta)$ is characterized by

$$\begin{cases} q_i(\theta) = 0 & \text{if } C_i'(0) \;\geq\; \rho_i \Big(p(\theta, r_i(\theta)) v(\theta) - v'(\theta) \Phi_i(\theta) + \\ q_i(\theta) = 1 & \text{if } C_i'(1) \;\leq\; \rho_i \Big(p(\theta, r_i(\theta)) v(\theta) - v'(\theta) \Phi_i(\theta) + \\ C_i'(q_i(\theta)) = \rho_i \Big(p(\theta, r_i(\theta)) v(\theta) - v'(\theta) \Phi_i(\theta) + \Xi_i(\theta) \Big) & \text{if } q_i(\theta) \in (0, 1), \end{cases}$$

where

$$\Xi_i(\theta) = \frac{v(\theta)}{v'(\theta)} \frac{d}{d\theta} \left(\frac{r_i(\theta) p_r(\theta, r_i(\theta))}{v'(\theta)} \right)$$

is a correction term reflecting the sensitivity of repayment to the interest rate.

This proposition confirms that the structure of the optimal menu remains as in the baseline model. Repayment schedules continue to be determined by the envelope condition, while approval rates are pinned down by a pointwise first-order condition. The only change is the presence of the correction term $\Xi_i(\theta)$, which captures the fact that interest rates affect repayment probabilities as well as revenues. In equilibrium, lenders must balance the benefit of charging a higher rate against the associated decline in repayment likelihood.

Corollary 2 (Monotonicity) *In the optimal contract, the approval schedule $q_i(\theta)$ is weakly decreasing in type, while the repayment schedule $r_i(\theta)$ is weakly increasing.*

Thus the ordering of contract terms across types is preserved: safer borrowers are granted greater access to credit at lower repayment rates, while riskier borrowers face reduced approval probabilities and higher repayment burdens. Allowing repayment probabilities to depend on interest rates does not alter this fundamental property.

Proposition 8 (Sorting with Rate-Dependent Repayment) *There exists a unique cutoff type θ^* such that*

$$\Theta_A = [\theta_A, \theta^*), \qquad \Theta_B = [\theta^*, \theta_B].$$

The more efficient lender serves the safer borrowers, while the less efficient lender serves the riskier borrowers.

This result demonstrates that the central sorting logic of the model is robust to allowing repayment to depend on the interest rate. The more efficient lender specializes in safer borrowers, and the less efficient lender specializes in riskier borrowers, ensuring market segmentation along a unique cutoff. Taken together, the extension shows that the main qualitative results of the baseline model continue to hold. Borrower incentives and the envelope condition remain unchanged, so monotonicity is preserved; lenders continue to sort borrowers according to efficiency; and the only substantive difference is that the kernel of the first-order condition includes the adjustment term $\Xi_i(\theta)$, reflecting the trade-off between higher interest revenues and lower repayment probabilities. This modification enriches the quantitative characterization of optimal menus without altering the qualitative structure of equilibrium.

Having characterized the structure of optimal menus, we now turn to the welfare implications of allowing repayment probabilities to depend on both borrower type and contractual interest rates. The key question is whether the dependence of $p(\theta, r)$ on r alters the evaluation of aggregate borrower surplus, lender profits, and the effects of regulatory interventions such as interest rate caps. Because higher rates raise revenues but lower repayment probabilities, they create an internal distortion relative to the benchmark where repayment depends only on borrower type. Nonetheless, the envelope structure of borrower utility ensures that borrower surplus is unaffected in form: higher rates reduce utility directly, while incentive compatibility continues to determine approval through the slope of the utility schedule.

We now analyze the effect of imposing an exogenous cap $\bar{r}$ on repayment rates. As in the baseline case, if the cap is not binding, the equilibrium menu remains unchanged. If the cap binds, then repayment is evaluated at the capped level $p(\theta, \bar{r})$ for all borrowers in the bunching set.

Proposition 9 (Interest Rate Caps with Rate-Dependent Repayment) *Suppose a cap $\bar{r}$ binds for some set $B_i \subseteq \Theta_i$. Then for all $\theta \in B_i$,*

$$r_i(\theta) = \bar{r}, \qquad p(\theta, r_i(\theta)) = p(\theta, \bar{r}).$$

In this region, approval probabilities adjust downward, repayment schedules are flat, and bunching occurs exactly as in the baseline model.

The introduction of rate-dependent repayment therefore leaves the qualitative effects of caps unchanged. When the cap is slack, it is irrelevant. When binding, it creates flat repayment schedules and downward distortions in approval. The only difference is that repayment probabilities are evaluated at $p(\theta, \bar{r})$, rather than at the type-dependent baseline $p(\theta)$.

From a welfare perspective, the presence of $p(\theta, r)$ modifies the trade-off associated with caps. On the one hand, lower capped rates reduce borrower repayment obligations, which directly increases borrower surplus. On the other hand, they also increase repayment probabilities, which raises expected revenues for lenders and reduces the enforcement costs associated with defaults. These effects reinforce each other, so that the welfare gain from a binding cap is generally larger when repayment depends on the interest rate than when it does not. However, as in the baseline case, the cap reduces lender flexibility in screening borrowers and therefore distorts approval downward in the bunching set.

Corollary 3 (Welfare under Binding Caps) *If a cap $\bar{r}$ binds on a non-empty set of types, then total welfare relative to the uncapped optimum is determined by the balance between (i) increased repayment probabilities, which raise lender revenues and reduce enforcement costs, and (ii) reduced screening efficiency, which lowers the allocation of credit. The presence of $p_r(\theta, r) < 0$ amplifies the first effect and leaves the second effect unchanged.*

This corollary highlights that the welfare implications of caps are strengthened in the extended model: because repayment probabilities now rise when interest rates are constrained, caps have an additional beneficial effect relative to the baseline. At the same time, the efficiency costs of bunching remain present, so the overall welfare consequences of caps remain ambiguous in general.

In summary, allowing repayment probabilities to depend on both borrower type and the contractual interest rate preserves the main qualitative results of the baseline model while enriching its quantitative structure. Borrower utility and incentive compatibility remain governed by the envelope condition, ensuring that approval probabilities are weakly decreasing and repayment rates weakly increasing in type. Sorting across lenders is unaffected: there continues to exist a unique cutoff type such that the more efficient lender serves the safer borrowers and the less efficient lender serves the riskier borrowers.

The central difference lies in the lender]s optimality condition. Whereas in the baseline case the kernel of the first-order condition depends only on type, here it also includes an adjustment term that reflects the sensitivity of repayment to the interest rate. This term formalizes the trade-off faced by lenders: raising the interest rate increases revenues per successful repayment but simultaneously lowers the probability of repayment. The correction term captures the net effect of this trade-off on the optimal approval schedule.

Finally, the analysis of regulation through interest rate caps carries over directly. When the cap is slack, it is irrelevant; when binding, it generates flat repayment schedules and bunching exactly as in the baseline model. The only modification is that repayment probabilities are evaluated at $p(\theta, \bar{r})$. From a welfare perspective, this feature strengthens the positive effects of caps, since lower rates increase repayment probabilities as well as borrower surplus. The efficiency cost of bunching, however, remains unchanged. In sum, the extension confirms that the qualitative predictions of the baseline model are robust, while introducing an additional and realistic channel through which contract terms affect repayment behavior.

8. Conclusion

This paper develops a screening model of credit markets with heterogeneous lenders, borrower default risk, and an interest rate cap imposed by a regulator. We show how differences in enforcement capacity shape contract design, borrower sorting, and market segmentation. In equilibrium, lenders with weaker enforcement are more likely to be constrained by the cap, which pushes them to offer flat menus at the regulatory ceiling and to attract riskier borrowers. By contrast, lenders with stronger enforcement offer more flexible menus at lower rates and attract safer borrowers.

Our welfare analysis identifies the central trade-off faced by regulators: tighter caps expand borrower coverage but reduce lender participation and distort screening, while looser caps maintain participation but shift riskier borrowers into the market at high repayment burdens. The welfare-maximizing cap balances these opposing forces, ensuring both access and sustainability.

The model rationalizes two key empirical regularities—non-bank lenders charging at the cap and experiencing higher default rates than banks—and provides a unified framework for analyzing the consequences of interest rate regulation. More broadly, our results highlight that uniform caps are a blunt instrument: by ignoring lender heterogeneity, they can induce inefficient borrower sorting and amplify default risk.

Future research could extend the framework along several dimensions: by incorporating dynamic interactions where default today affects future credit access; by analyzing alternative regulatory tools such as differentiated caps, collateral requirements, or subsidies for enforcement technology; and by calibrating the model with empirical data to quantify the welfare consequences of alternative policies. These extensions would further inform the design of regulation in credit markets where both lenders and borrowers are heterogeneous.

References

Sumit Agarwal, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao. Searching for approval. *Econometrica*, 92(4):1195–1231, 2024.

Adrien Auclert. Monetary policy and the redistribution channel. *American Economic Review*, 109(6):2333–2367, 2019. URL `https://www.aeaweb.org/articles?id=10.1257/aer.20160111`.

Adrien Auclert, Matthew Rognlie, and Ludwig Straub. The intertemporal keynesian cross. *Quarterly Journal of Economics*, 136(4):2335–2403, 2021. URL `https://academic.oup.com/qje/article/136/4/2335/6297841`.

Robert Bartlett, Adair Morse, Richard Stanton, and Nancy Wallace. Consumer-lending discrimination in the fintech era. *Journal of Financial Economics*, 143(1):30–56, 2022.

David Besanko and Anjan V Thakor. Competitive equilibrium in the credit market under asymmetric information. *Journal of Economic Theory*, 42(1):167–182, 1987.

John Y Campbell. Restoring rational choice: The challenge of consumer financial regulation. *American Economic Review*, 106(5):1–30, 2016.

Christopher D. Carroll. Buffer-stock saving and the life cycle/permanent income hypothesis. *Quarterly Journal of Economics*, 112(1):1–55, 1997. doi: 10.1162/003355397555109.

Christopher D. Carroll, Edmund Crawley, Jiri Slacalek, Kiichi Tokuoka, and Matthew N. White. Sticky expectations and consumption dynamics. *Working/Pub. versions in the 2020s*. URL `https://www.econ2.jhu.edu/people/ccarroll/public/`. Integrates inattentive expectations into buffer-stock models; see author pages for latest published version.

Hector Chade and Jeroen Swinkels. Screening in vertical oligopolies. *Econometrica*, 89(3): 1265–1311, 2021.

Satyajit Chatterjee and Burcu Eyigungor. Maturity, indebtedness, and default risk. *American Economic Review*, 102(6):2674–2699, 2012.

Satyajit Chatterjee, Dean Corbae, Makoto Nakajima, and José-Víctor Ríos-Rull. A quantitative theory of unsecured consumer credit with risk of default. *Econometrica*, 75(6): 1525–1589, 2007.

Satyajit Chatterjee, Dean Corbae, and Jose-Victor Rios-Rull. A finite-life private-information theory of unsecured consumer debt. *Journal of Economic Theory*, 142(1): 149–177, 2008.

Raj Chetty, John N. Friedman, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team. How did covid-19 and stabilization policies affect spending and employment? a new real-time economic tracker. *Quarterly Journal of Economics*, 135(4): 2575–2611, 2020. URL `https://opportunityinsights.org/paper/tracker/`.

Yinxiao Chu and Jianxing Wei. Fintech lending and credit market competition. *Journal of Financial and Quantitative Analysis*, 59(5):2199–2225, 2024.

Angus Deaton. Saving and liquidity constraints. *Econometrica*, 59(5):1221–1248, 1991. doi: 10.2307/2938366.

Rajeev Dehejia, Heather Montgomery, and Jonathan Morduch. Do interest rates matter? credit demand in the dhaka slums. *Journal of Development Economics*, 97(2):437–449, 2012.

Milton Friedman. *A Theory of the Consumption Function.* Princeton University Press, Princeton, NJ, 1957. doi: 10.1515/9781400879766.

Andreas Fuster, Matthew Plosser, Philipp Schnabl, and James Vickery. The role of technology in mortgage lending. *The review of financial studies*, 32(5):1854–1899, 2019.

Robert E. Hall. Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence. *Journal of Political Economy*, 86(6):971–987, 1978. doi: 10.1086/260724.

Tullio Jappelli. Who is credit constrained in the u.s. economy? *The Quarterly Journal of Economics*, 105(1):219–234, 1990. doi: 10.2307/2937825.

Tullio Jappelli. Credit rationing: Empirical evidence from survey data. *Oxford Economic Papers*, 46(3):517–538, 1994. doi: 10.1093/oxfordjournals.oep.a042137.

Tullio Jappelli, Marco Pagano, and Marco Di Maggio. Information sharing in credit markets. *Journal of Money, Credit and Banking*, 30(2):351–377, 1998. doi: 10.2307/2601110.

Greg Kaplan, Giovanni L. Violante, and Justin Weidner. The wealthy hand-to-mouth. *Brookings Papers on Economic Activity*, (Spring):77–153, 2014. URL https://www.brookings.edu/bpea-articles/the-wealthy-hand-to-mouth/.

Greg Kaplan, Benjamin Moll, and Giovanni L. Violante. Monetary policy according to hank. *American Economic Review*, 108(3):697–743, 2018. URL https://www.aeaweb.org/articles?id=10.1257/aer.20160042.

Kei Kawai, Ken Onishi, and Kosuke Uetake. Signaling in online credit markets. *Journal of Political Economy*, 130(6):1585–1629, 2022.

Benjamin Lester, Ali Shourideh, Venky Venkateswaran, and Ariel Zetlin-Jones. Screening and adverse selection in frictional markets. *Journal of Political Economy*, 127(1):338–377, 2019.

Franco Modigliani and Richard Brumberg. Utility analysis and the consumption function: An interpretation of cross-section data. In Kenneth K. Kurihara, editor, *Post-Keynesian Economics*. Rutgers University Press, New Brunswick, NJ, 1954.

John Mondragon. Credit supply, prices, and non-price mechanisms in the mortgage market. Federal Reserve Bank of San Francisco, 2024.

Sangmin Oh, Ishita Sen, and Ana-Maria Tenekedjieva. Pricing of climate risk insurance: Regulation and cross-subsidies. *Journal of Finance, Forthcoming*, 2021.

Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. In *Uncertainty in economics*, pages 257–280. Elsevier, 1978.

Joseph E Stiglitz and Andrew Weiss. Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3):393–410, 1981.

A. Appendix: Proofs

Proof. Lemma 1. Each borrower type θ accepts a contract only if her utility is non-negative, i.e., $U_i(\theta) \geq 0$ for all $\theta \in [0,1]$. Consider lender $i \in \{A, B\}$, who serves types in $[\underline{\theta}_i, \overline{\theta}_i]$. Suppose by contradiction that $U_i(\overline{\theta}_i) > 0$. Lender i could strictly reduce utility for type $\overline{\theta}_i$ by increasing the interest rate or lowering the probability of approval (while maintaining incentive compatibility), thereby increasing profits without violating participation or incentive compatibility. This contradicts optimality. Hence, $U_i(\overline{\theta}_i) = 0$ for all $i \in \{A, B\}$. ■

Proof. Proposition 1. We solve the relaxed version of the lender's problem without an interest rate cap. Incentive compatibility implies that $q_i(\theta)v'(\theta) = U_i'(\theta)$, and the repayment function is given by:

$$r_i(\theta) = v(\theta) - \frac{U_i(\theta)}{q_i(\theta)}.$$

Lender i's expected profit in terms of $q_i(\theta)$ can be written as:

$$\int_{\underline{\theta}_i}^{\overline{\theta}_i} \left[\rho_i p(\theta) \left(v(\theta)q_i(\theta) - \int_{\theta}^{\overline{\theta}_i} q_i(s)v'(s)\,ds\right) - C_i(q_i(\theta))\right] f(\theta)d\theta.$$

Define

$$\Phi_i(\theta) \;:=\; \frac{1}{f(\theta)} \int_{\underline{\theta}_i}^{\theta} \rho_i\, p(t)\, f(t)\, dt.$$

Then any interior solution $q_i(\theta) \in (0,1)$ satisfies the pointwise first-order condition

$$C_i'\big(q_i(\theta)\big) \;=\; \rho_i\, p(\theta)\, v(\theta) \;-\; v'(\theta)\, \Phi_i(\theta).$$

Therefore, the optimal approval schedule is given by

$$\begin{aligned} q_i(\theta) = 0 \quad &\text{when} \quad C_i'(0) \;\geq\; \rho_i\, p(\theta)\, v(\theta) \;-\; v'(\theta)\, \Phi_i(\theta), \\ q_i(\theta) = 1 \quad &\text{when} \quad C_i'(1) \;\leq\; \rho_i\, p(\theta)\, v(\theta) \;-\; v'(\theta)\, \Phi_i(\theta), \\ q_i(\theta) \in (0,1) \quad &\text{when} \quad C_i'\big(q_i(\theta)\big) \;=\; \rho_i\, p(\theta)\, v(\theta) \;-\; v'(\theta)\, \Phi_i(\theta). \end{aligned}$$

■

Proof. Corollary 1. The approval function q_i is weakly decreasing by incentive-compatibility. To see that the interest rate is weakly increasing in θ, define

$$I_i(\theta) := -\int_{\theta}^{\overline{\theta}_i} q_i(s)v'(s)\,ds \quad \geq 0.$$

Differentiating $r_i(\theta)$ with respect to θ yields

$$r_i'(\theta) = \frac{I_i(\theta)q_i'(\theta)}{q_i(\theta)^2}.$$

Because $I_i(\theta) \geq 0$ and $q_i'(\theta) \leq 0$ by incentive compatibility, it follows that $r_i'(\theta) \geq 0$ for all $\theta \in \Theta_i$ with $q_i(\theta) > 0$. ■

Proof. Lemma 2. Fix any interval on which both lenders serve borrowers and their approval schedules are interior and differentiable. Incentive compatibility implies

$$U_i'(\theta) = v'(\theta)q_i(\theta) \quad \text{and} \quad U_i(\overline{\theta}_i) = 0,$$

so for the utility difference $\Delta U(\theta) := U_A(\theta) - U_B(\theta)$ we have

$$\Delta U'(\theta) = v'(\theta)\big(q_A(\theta) - q_B(\theta)\big).$$

Because $v'(\theta) < 0$, the sign of $\Delta U'(\theta)$ is the opposite of the sign of $q_A(\theta) - q_B(\theta)$. Hence, to show that $q_A(\theta) - q_B(\theta)$ is weakly decreasing, it suffices to show that $\Delta U'(\theta)$ is weakly increasing in θ on any region where both menus are interior.

For interior types, the first–order condition gives

$$C_i'\big(q_i(\theta)\big) = \rho_i p(\theta) v(\theta) - v'(\theta)\Phi_i(\theta), \qquad \Phi_i(\theta) = \frac{1}{f(\theta)} \int_{\underline{\theta}_i}^{\theta} \rho_i p(t) f(t)\, dt,$$

with $C_i'' > 0$. Differentiating the FOC and using $\Phi_i'(\theta) = \rho_i p(\theta)$ yields

$$C_i''\big(q_i(\theta)\big)\, q_i'(\theta) = \rho_i p'(\theta) v(\theta) \;-\; v''(\theta)\, \Phi_i(\theta).$$

Because $\rho_A > \rho_B$ and $\Phi_A(\theta) \geq \Phi_B(\theta)$ for all θ (both scale with ρ_i), while $C_i'' > 0$, we obtain

$$\big(q_A'(\theta) - q_B'(\theta)\big) \;\leq\; 0 \quad \text{whenever both menus are interior.}$$

Therefore, on any such region $q_A(\theta) - q_B(\theta)$ is weakly decreasing in θ.

On regions where one lender is at a boundary ($q_i \in \{0, 1\}$), monotonicity of q_i implies $q_i'(\theta) \leq 0$ in the sense of distributions, so the difference remains weakly decreasing across the junctions. Finally, because $U_i(\overline{\theta}_i) = 0$ for each served set $[\underline{\theta}_i, \overline{\theta}_i]$, the utility difference is anchored at the upper endpoints, and the single–crossing of $\Delta U(\theta)$ follows from the weak decrease of $q_A - q_B$ together with $\Delta U'(\theta) = v'(\theta)(q_A - q_B)$. Hence $\theta \mapsto q_A(\theta) - q_B(\theta)$ is weakly decreasing on the union of types served in equilibrium. ■

Proof. Lemma 3. Fix a type θ in a neighborhood of the lower end of the support, with θ close to 0 (safer borrowers). An interior approval $q_i(\theta) \in (0, 1)$ is characterized pointwise by

$$C_i'\big(q_i(\theta)\big) \;=\; \rho_i\, p(\theta)\, v(\theta) \;-\; v'(\theta)\, \Phi_i(\theta), \qquad \Phi_i(\theta) \;=\; \frac{1}{f(\theta)} \int_{\underline{\theta}_i}^{\theta} \rho_i p(t) f(t)\, dt.$$

Consider the marginal profitability at $q = 0$. By convexity ($C_i'' > 0$), a necessary condition for any profitable approval is that the right-hand side exceed $C_i'(0)$ at the candidate type:

$$\underbrace{\rho_i\, p(\theta)\, v(\theta) \;-\; v'(\theta)\, \Phi_i(\theta)}_{=:\; M_i(\theta)} \;>\; C_i'(0).$$

As $\theta \downarrow \underline{\theta}_i$, we have $\Phi_i(\theta) \to 0$ by definition, so $M_i(\theta) \to \rho_i\, p(\theta)\, v(\theta)$. By continuity of p and

v, for θ sufficiently small we can make $M_i(\theta)$ arbitrarily close to $\rho_i p(\theta)v(\theta)$.

By the maintained ordering of effective marginal costs,

$$\frac{C'_A(0)}{\rho_A} \le \frac{C'_B(0)}{\rho_B} \qquad \text{with } \rho_A > \rho_B,$$

so there exists $\bar{\theta} > 0$ such that for all $\theta \in [0, \bar{\theta}]$,

$$\rho_A\, p(\theta)\, v(\theta) \;-\; C'_A(0) \;>\; 0 \quad \text{while} \quad \rho_B\, p(\theta)\, v(\theta) \;-\; C'_B(0) \;\le\; 0.$$

Using $\Phi_i(\theta) \approx 0$ near the lower endpoint, it follows that for such θ,

$$M_A(\theta) - C'_A(0) > 0 \quad \Rightarrow \quad \text{lender } A \text{ can profitably approve a positive } q_A(\theta),$$

whereas

$$M_B(\theta) - C'_B(0) \le 0 \quad \Rightarrow \quad \text{lender } B \text{ cannot profitably approve any } q_B(\theta) > 0.$$

Hence, for sufficiently low risk (small θ), lender A can profitably serve the type, while lender B cannot. ■

Proof. Proposition 2. Let $\Delta U(\theta) := U_A(\theta) - U_B(\theta)$. By incentive compatibility and the envelope condition,

$$\Delta U'(\theta) = v'(\theta)\big(q_A(\theta) - q_B(\theta)\big).$$

Since $\theta \mapsto q_A(\theta) - q_B(\theta)$ is weakly decreasing on the union of types served in equilibrium. Since $v'(\theta) < 0$, it follows that $\Delta U'(\theta)$ is weakly increasing in θ. Hence, ΔU has the single-crossing property.

At sufficiently low risk, lender A can profitably serve borrowers while lender B cannot. On such a neighborhood, either only A serves or, where both serve, A offers weakly higher utility. In either case there exists θ_L sufficiently small such that $\Delta U(\theta_L) \ge 0$.

At sufficiently high risk, feasibility and monotonicity imply that the less efficient lender B specializes in higher-risk types whenever both lenders are active. In particular, if both lenders serve some interval near the top of the support, B must (weakly) dominate there in borrower utility; otherwise safer borrowers would prefer to mimic higher-risk contracts, violating incentive compatibility and lender optimality. Thus, there exists θ_H close to 1 such that $\Delta U(\theta_H) \le 0$.

By continuity of utilities in θ and the intermediate value theorem, there exists at least one $\theta^* \in [\theta_L, \theta_H]$ such that $\Delta U(\theta^*) = 0$. By the single-crossing property, this cutoff is unique. Therefore, borrowers with $\theta < \theta^*$ strictly prefer A, and those with $\theta > \theta^*$ strictly prefer B;

at θ^* they are indifferent. Writing the served sets with explicit endpoints,

$$\Theta_A = [\underline{\theta}_A, \theta^*) \quad \text{and} \quad \Theta_B = [\theta^*, \overline{\theta}_B],$$

where $U_i(\overline{\theta}_i) = 0$. ■

Proof. Proposition 3. We analyze the constrained optimization problem faced by lender i.

1. Interest Rate Cap and Bunching. Define the set $\mathcal{B}_i := \{\theta \in \Theta_i : r_i(\theta) = \bar{r}\}$. On this set, the repayment constraint binds, meaning the lender is unable to raise the interest rate to extract more surplus from the borrower. Since the interest rate is fixed across all types in $\mathcal{B}_i$, but types differ in their marginal valuation of q_i, the lender must distort the approval schedule $q_i(\theta)$ to preserve incentive compatibility. In particular, higher types in $\mathcal{B}_i$ must receive lower q_i to deter mimicking by lower types. This implies that the optimal menu features bunching within $\mathcal{B}_i$: the repayment is flat, and the variation in contracts is solely along the approval dimension.

2. Strict Screening Outside the Cap. Outside the region where the cap binds (i.e., for $\theta \notin \mathcal{B}_i$), the lender is unconstrained in setting repayment, and hence solves a standard screening problem. For any interior types $\theta \notin \mathcal{B}_i$, for which the cap constraint is slack, the optimal contract is as in Proposition 1. ■

Proof. Proposition 4. Let $W(\bar{r})$ denote total welfare under cap $\bar{r}$. Differentiating $W(\bar{r})$ with respect to $\bar{r}$ yields:

$$\begin{aligned}
\frac{dW}{d\bar{r}} &= \frac{d}{d\bar{r}} \left(\int_{\theta^*(\bar{r})}^{1} [U_B(\theta) + \pi_B(\theta)]\, d\theta + \int_0^{\theta^*(\bar{r})} [U_A(\theta) + \pi_A(\theta)]\, d\theta \right) \\
&= \int_{\theta^*}^{1} \left(\frac{\partial U_B(\theta)}{\partial \bar{r}} + \frac{\partial \pi_B(\theta)}{\partial \bar{r}} \right) d\theta + \int_0^{\theta^*} \left(\frac{\partial U_A(\theta)}{\partial \bar{r}} + \frac{\partial \pi_A(\theta)}{\partial \bar{r}} \right) d\theta \\
&\quad + [U_A(\theta^*) - U_B(\theta^*) + \pi_A(\theta^*) - \pi_B(\theta^*)] \cdot \frac{d\theta^*}{d\bar{r}}.
\end{aligned}$$

At the cutoff θ^* borrowers are indifferent, so $U_A(\theta^*) = U_B(\theta^*)$; hence the boundary term simplifies to

$$\left[\pi_A(\theta^*) - \pi_B(\theta^*)\right] \frac{d\theta^*}{d\bar{r}}.$$

Note that

$$\frac{\partial U_i(\theta)}{\partial \bar{r}} = \big(v(\theta) - r_i(\theta)\big) \frac{\partial q_i(\theta)}{\partial \bar{r}} - q_i(\theta) \frac{\partial r_i(\theta)}{\partial \bar{r}}, \tag{1}$$

$$\frac{\partial \pi_i(\theta)}{\partial \bar{r}} = \rho_i p(\theta) \left[q_i(\theta) \frac{\partial r_i(\theta)}{\partial \bar{r}} + r_i(\theta) \frac{\partial q_i(\theta)}{\partial \bar{r}} \right] - C_i'\big(q_i(\theta)\big) \frac{\partial q_i(\theta)}{\partial \bar{r}}. \tag{2}$$

Let $\mathcal{B}_i(\bar{r}) := \{\theta \in \Theta_i : \ r_i(\theta; \bar{r}) = \bar{r}\}$ be the (a.e.) binding set. For small changes in $\bar{r}$, we have that

$$\frac{\partial r_i(\theta)}{\partial \bar{r}} = \mathbf{1}_{\{\theta \in \mathcal{B}_i(\bar{r})\}} \quad \text{a.e.}, \qquad \frac{\partial q_i(\theta)}{\partial \bar{r}} \ \geq \ 0 \ \text{ a.e.}$$

Summing (1)–(2) gives the local welfare effect at type θ:

$$\frac{\partial}{\partial \bar{r}}\big(U_i(\theta) + \pi_i(\theta)\big) = q_i(\theta)\,\big(\rho_i p(\theta) - 1\big)\,\mathbf{1}_{\{\theta \in \mathcal{B}_i(\bar{r})\}} \\ + \Big[\big(v(\theta) - r_i(\theta)\big) \ + \ \rho_i p(\theta)\, r_i(\theta) \ - \ C_i'\big(q_i(\theta)\big)\Big]\,\frac{\partial q_i(\theta)}{\partial \bar{r}}.$$

Therefore,

$$\frac{dW}{d\bar{r}} = \sum_{i \in \{A,B\}} \int_{\Theta_i} \Big\{ q_i(\theta)\,\big(\rho_i p(\theta) - 1\big)\,\mathbf{1}_{\{\theta \in \mathcal{B}_i(\bar{r})\}} \\ + \Big[\big(v(\theta) - r_i(\theta)\big) + \rho_i p(\theta)\, r_i(\theta) - C_i'\big(q_i(\theta)\big)\Big]\,\frac{\partial q_i(\theta)}{\partial \bar{r}} \Big\}\, d\theta \\ + \big[\pi_A(\theta^*) - \pi_B(\theta^*)\big]\,\frac{d\theta^*}{d\bar{r}}.$$

The first integral term captures the direct effect of relaxing the cap on types where it binds: it raises posted rates (a transfer from borrowers to lenders) but destroys surplus proportional to default risk, at rate $q_i(\rho_i p - 1) \leq 0$. The second integral term captures the induced adjustment in approval (screening distortion): relaxing the cap weakly increases q_i, and the bracketed coefficient is the marginal social surplus of approval at type θ. The last term is the reallocation effect from the induced shift in the cutoff $\theta^*(\bar{r})$.

An interior optimum $\bar{r}^\star$ thus satisfies the first–order condition $dW/d\bar{r}\big|_{\bar{r}=\bar{r}^\star} = 0$: the marginal welfare gain from relaxing the cap on the binding region and from the induced expansion of approval (and reallocation toward the more efficient enforcer) is exactly offset by the marginal welfare loss due to higher prices under default risk and any additional screening distortions. ∎

Proof. Proposition 5. The proof is analogous to the one from Proposition 3. ∎

Proof. Lemma 4. Since $q_i(\theta)$ is weakly decreasing, $S_i \cap B_i = \{\theta \in B_i : q_i(\theta) > 0\}$ is a lower interval. Otherwise, there would exist $\theta_1 < \theta_2 < \theta_3$ in B_i with $q_i(\theta_1) > 0$, $q_i(\theta_2) = 0$, $q_i(\theta_3) > 0$, contradicting monotonicity. Let $\tilde{\theta}_i$ be the upper endpoint; then $(\tilde{\theta}_i, \overline{\theta}_i] \cap B_i$ is the (possibly empty) exclusion region. ∎

Proof. Lemma 5.By definition of B_i, the interest rate is fixed at $\bar{r}$ on B_i, so for any $\theta \in B_i$ and any $q \in [0,1]$ the per–type net surplus at approval level q equals

$$\alpha_i(\theta)\, q - C_i(q), \qquad \text{where } \ \alpha_i(\theta) = \rho_i p(\theta)\, \bar{r} - (1 - p(\theta))\, \rho_i^F.$$

Let $\alpha_i := \inf_{\theta\in B_i} \alpha_i(\theta)$. If

$$\max_{q\in[0,1]} \{\alpha_i q - C_i(q)\} \;\geq\; f_i,$$

then there exists $q^\star \in (0,1]$ such that $\alpha_i q^\star - C_i(q^\star) \geq f_i$. Since $\alpha_i(\theta) \geq \alpha_i$ for all $\theta \in B_i$, we have

$$\alpha_i(\theta)\, q^\star - C_i(q^\star) \;\geq\; \alpha_i q^\star - C_i(q^\star) \;\geq\; f_i \qquad \text{for every } \theta \in B_i.$$

Therefore, approving $q^\star > 0$ is profitable for every $\theta \in B_i$, which implies $q_i(\theta) > 0$ throughout B_i. Hence there is no exclusion in the cap–binding region, as claimed. ■

Proof. Lemma 6. Inside the cap–binding set B_i we have $r_i(\theta) = \bar{r}$, so screening operates only through quantities. For $\theta \in B_i$ and any $q \in [0,1]$, define the cap–adjusted per–type margin

$$\alpha_i(\theta) \;:=\; \rho_i p(\theta)\, \bar{r} \;-\; (1 - p(\theta))\, \rho_i^F,$$

so that the net surplus at approval level q equals $\alpha_i(\theta)\, q - C_i(q)$ and lending is optimal only if this surplus covers the activation cost f_i. Equivalently, define

$$\Gamma_i(\theta; \rho_i^F, f_i) \;:=\; \max_{q\in[0,1]} \{\alpha_i(\theta)\, q - C_i(q)\} \;-\; f_i.$$

Then, within B_i, the served set is $S_i \cap B_i = \{\theta \in B_i : \Gamma_i(\theta; \rho_i^F, f_i) \geq 0\}$. By monotonicity of $q_i(\cdot)$, this set is a connected lower interval $[\theta_i^c, \tilde{\theta}_i)$, where $\tilde{\theta}_i$ is the terminal cutoff characterized by $\Gamma_i(\tilde{\theta}_i; \rho_i^F, f_i) = 0$ when the exclusion region is nonempty.

We claim that Γ_i is (i) weakly decreasing in θ, (ii) strictly decreasing in ρ_i^F, and (iii) strictly decreasing in f_i. Indeed, since $p'(\theta) < 0$ on B_i, we have

$$\frac{\partial \alpha_i(\theta)}{\partial \theta} = \rho_i p'(\theta)\, \bar{r} \;-\; (-p'(\theta))\, \rho_i^F = p'(\theta)\left(\rho_i \bar{r} + \rho_i^F\right) \;\leq\; 0,$$

so $\alpha_i(\theta)$ is weakly decreasing in θ; as the pointwise supremum of affine functions in $\alpha_i(\theta)$, $\max_q\{\alpha_i(\theta) q - C_i(q)\}$ is weakly decreasing in θ, hence Γ_i is weakly decreasing in θ. Moreover,

$$\frac{\partial \alpha_i(\theta)}{\partial \rho_i^F} = -(1 - p(\theta)) \;\leq\; 0, \qquad \frac{\partial \Gamma_i}{\partial f_i} = -1 < 0,$$

so Γ_i is strictly decreasing in each parameter ρ_i^F and f_i.

Because $\Gamma_i(\cdot; \rho_i^F, f_i)$ is weakly decreasing in θ and continuous (by continuity of p and C_i), the terminal cutoff $\tilde{\theta}_i$ is the unique solution of $\Gamma_i(\theta; \rho_i^F, f_i) = 0$ when it exists. An increase in either ρ_i^F or f_i shifts $\Gamma_i(\theta; \cdot, \cdot)$ downward pointwise, so the zero–crossing moves (weakly) to a lower θ. Formally, by the monotone comparative statics of threshold crossings,

$$\frac{\partial \tilde{\theta}_i}{\partial \rho_i^F} \leq 0 \qquad \text{and} \qquad \frac{\partial \tilde{\theta}_i}{\partial f_i} \leq 0,$$

with weak inequalities allowing for corners where the exclusion region is empty or touches the boundary.

Hence the terminal exclusion region $(\tilde{\theta}_i, \overline{\theta}_i] \cap B_i$ weakly expands as either ρ_i^F or f_i increases, i.e., $\tilde{\theta}_i$ is weakly decreasing in each parameter. ■

Proof. Proposition 6. Fix $\bar{r}$ and let $\{q_i(\cdot), r_i(\cdot)\}_{i\in\{A,B\}}$ be an optimal menu. For each lender i, write the contribution to welfare as

$$W_i(\bar{r}) = \int_{\Theta_i} w_i(\theta;\bar{r})\, f(\theta)\, d\theta, \quad w_i(\theta;\bar{r}) := v(\theta)q_i(\theta) - C_i(q_i(\theta)) - (1-p(\theta))\rho_i^F q_i(\theta) - f_i\, \mathbf{1}\{q_i(\theta) > 0\}.$$

Consider the Lagrangian for lender i's problem aggregated over types in Θ_i, placing the multiplier $\lambda_i(\theta) \geq 0$ on the cap constraint $r_i(\theta) \leq \bar{r}$ in the form $\bar{r} - r_i(\theta) \geq 0$, and let the (type-by-type) multipliers on the remaining constraints be collected in $\Gamma_i(\theta)$. Under the stated differentiability and no-atom assumptions, the Milgrom–Segal envelope theorem and dominated convergence give

$$\frac{dW_i(\bar{r})}{d\bar{r}} = \int_{\Theta_i} \lambda_i(\theta)\, f(\theta)\, d\theta \;+\; \text{boundary terms from moving cutoffs.}$$

Complementary slackness implies $\lambda_i(\theta) = 0$ whenever the cap does not bind ($r_i(\theta) < \bar{r}$) or the type is not served ($q_i(\theta) = 0$). Hence the interior (envelope) contribution is

$$\int_{B_i \cap S_i} \lambda_i(\theta)\, f(\theta)\, d\theta.$$

We now identify the boundary terms. First, within B_i there is a terminal approval cutoff $\tilde{\theta}_i$ (possibly absent) at which q_i drops to zero. At this point $w_i(\cdot\,;\bar{r})$ has a jump from $w_i(\tilde{\theta}_i^-;\bar{r})$ to 0. By the Leibniz rule for integrals with moving discontinuities,

$$\frac{d}{d\bar{r}} \int_{\Theta_i} w_i(\theta;\bar{r})\, f(\theta)\, d\theta = \int_{\Theta_i} \frac{\partial w_i(\theta;\bar{r})}{\partial \bar{r}}\, f(\theta)\, d\theta \;+\; w_i(\tilde{\theta}_i^-;\bar{r})\, f(\tilde{\theta}_i)\, \frac{d\tilde{\theta}_i}{d\bar{r}},$$

where the first term on the right is exactly the envelope term already computed. This delivers the second component in the statement:

$$w_i(\tilde{\theta}_i^-;\bar{r})\, f(\tilde{\theta}_i)\, \frac{d\tilde{\theta}_i}{d\bar{r}}.$$

Second, the lenders' integration domains are split by the sorting cutoff $\theta^*(\bar{r})$. Differentiating

$$W_A(\bar{r}) = \int_{\theta_A}^{\theta^*(\bar{r})} w_A(\theta;\bar{r})\, f(\theta)\, d\theta, \qquad W_B(\bar{r}) = \int_{\theta^*(\bar{r})}^{\theta_B} w_B(\theta;\bar{r})\, f(\theta)\, d\theta,$$

and using the Leibniz rule with a moving limit yields the re-sorting term

$$\big[w_B(\theta^*;\bar{r}) - w_A(\theta^*;\bar{r})\big]\ f(\theta^*)\ \frac{d\theta^*}{d\bar{r}}\ =\ \Delta_{\text{sort}}(\bar{r}).$$

Summing the three contributions across lenders gives

$$\frac{dW(\bar{r})}{d\bar{r}} = \sum_i \int_{B_i \cap S_i} \lambda_i(\theta)\, f(\theta)\, d\theta\ +\ \sum_i w_i(\tilde{\theta}_i^-;\bar{r})\, f(\tilde{\theta}_i)\, \frac{d\tilde{\theta}_i}{d\bar{r}}\ +\ \Delta_{\text{sort}}(\bar{r}),$$

as claimed. The final statement about interior optima follows immediately: at any interior maximizer $\bar{r}^\star$, the one-sided derivative is zero and the usual regularity and second-order conditions apply; otherwise the maximizer lies at a boundary of the feasible set for $\bar{r}$. ■

Proof. Proposition 7. The result follows by applying the Euler–Lagrange condition to the variational problem

$$\max_{U_i(\cdot)} \int_{\Theta_i} \Big\{\rho_i\, p(\theta, r_i(\theta)) \Big(\tfrac{v(\theta)}{v'(\theta)} U_i'(\theta) - U_i(\theta)\Big) - C_i\Big(\tfrac{U_i'(\theta)}{v'(\theta)}\Big)\Big\} f(\theta)\, d\theta.$$

Letting

$$L(\theta, U, U') = \rho_i\, p(\theta, r(U, U'))\Big(\tfrac{v(\theta)}{v'(\theta)} U' - U\Big) - C_i\Big(\tfrac{U'}{v'}\Big),$$

the Euler–Lagrange equation requires

$$\frac{\partial(Lf)}{\partial U} - \frac{d}{d\theta}\frac{\partial(Lf)}{\partial U'} = 0.$$

Differentiating yields

$$\frac{\partial L}{\partial U} = \rho_i\left[p_r(\theta, r)\Big(-\tfrac{v'}{U'}\Big)\Big(\tfrac{v}{v'}U' - U\Big) - p(\theta, r)\right],$$

$$\frac{\partial L}{\partial U'} = \rho_i\left[p_r(\theta, r)\Big(\tfrac{Uv'}{(U')^2}\Big)\Big(\tfrac{v}{v'}U' - U\Big) + p(\theta, r)\tfrac{v}{v'}\right] - \tfrac{1}{v'}C_i'\Big(\tfrac{U'}{v'}\Big).$$

Substituting these expressions into the Euler–Lagrange equation and simplifying gives the stated condition. The repayment schedule follows from the envelope condition and the definition of borrower utility. ■

Proof. Corollary 2. By incentive compatibility, $U_i'(\theta) = v'(\theta)q_i(\theta)$ with $v'(\theta) < 0$ and $q_i(\theta) \geq 0$, so U_i is weakly decreasing in θ and q_i is weakly monotone by standard single-crossing (as in the baseline). Where q_i is interior, the first-order condition from Proposition 7 gives

$$C_i'\big(q_i(\theta)\big) = \rho_i(p(\theta, r_i(\theta))v(\theta) - v'(\theta)\Phi_i(\theta) + \Xi_i(\theta)), \quad \Phi_i(\theta) = \frac{1}{f(\theta)} \int_{\underline{\theta}_i}^{\theta} \rho_i p\big(t, r_i(t)\big) f(t)\, dt,$$

with $C_i'' > 0$. Differentiating both sides and using $C_i'' > 0$ implies $q_i'(\theta) \leq 0$ on interior regions; at boundaries $q_i \in \{0, 1\}$, monotonicity is preserved, so overall q_i is weakly decreasing.

For r_i, Proposition 7 (part 1) and the normalization $U_i(\overline{\theta}_i) = 0$ yield

$$r_i(\theta) = v(\theta) - \frac{1}{q_i(\theta)} \int_{\theta}^{\overline{\theta}_i} q_i(s)v'(s)\, ds, \quad I_i(\theta) := -\int_{\theta}^{\overline{\theta}_i} q_i(s)v'(s)\, ds \;\geq 0.$$

Differentiating where $q_i(\theta) > 0$,

$$r_i'(\theta) = \frac{I_i(\theta)\, q_i'(\theta)}{q_i(\theta)^2}.$$

Since $I_i(\theta) \geq 0$ and $q_i'(\theta) \leq 0$, it follows that $r_i'(\theta) \geq 0$ wherever $q_i(\theta) > 0$. This proves the claim. ■

Proof. Proposition 8. The argument is analogous to the baseline case. Given enforcement-cost dominance, the more efficient lender can profitably serve low-risk types at lower rates, while the less efficient lender caters to high-risk types. Since the difference in approval probabilities is decreasing in type, the existence of a unique cutoff follows. ■

Proof. Proposition 9. If the cap binds, lenders cannot raise $r_i(\theta)$ above $\bar{r}$. Hence the repayment probability is evaluated at the capped rate, $p(\theta, \bar{r})$. Since repayment cannot vary with type in the bunching set, incentive compatibility requires that the schedule $r_i(\theta)$ be constant there. The only margin of adjustment is approval, which responds to the modified marginal revenue condition with repayment fixed at $p(\theta, \bar{r})$. Thus the structure of bunching is identical to the baseline case. ■