Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Dynamic Credit-Collections Optimization

Naveed Chehrazi, Peter W. Glynn, Thomas A. Weber

informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 65, No. 6, June 2019, pp. 2737–2769
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Dynamic Credit-Collections Optimization

Naveed Chehrazi,[a] **Peter W. Glynn,**[b] **Thomas A. Weber**[c]

[a] Department of Information, Risk, and Operations Management, McCombs School of Business, The University of Texas at Austin, Austin, Texas 78705; [b] Department of Management Science and Engineering, Stanford University, Stanford, California 94305; [c] Economics and Strategy, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland

Contact: naveed.chehrazi@mccombs.utexas.edu, http://orcid.org/0000-0002-7750-6945 (NC); glynn@stanford.edu (PWG); thomas.weber@epfl.ch, http://orcid.org/0000-0003-2857-4754 (TAW)

Received: December 11, 2015
Revised: March 16, 2017; September 8, 2017
Accepted: January 11, 2018
Published Online in Articles in Advance: March 21, 2019

https://doi.org/10.1287/mnsc.2018.3070

Abstract. Based on a dynamic model of the stochastic repayment behavior exhibited by delinquent credit-card accounts in the form of a self-exciting point process, a bank can control the arrival intensity of repayments using costly account-treatment actions. A semi-analytic solution to the corresponding stochastic optimal control problem is obtained using a recursive approach. For a linear cost of treatment effort, the optimal policy in the two-dimensional (intensity, balance) space is described by the frontier of a convex action region. The unique optimal policy significantly reduces a bank's loss given default and concentrates the collection effort onto the best possible actions at the best possible times so as to minimize the sum of the expected discounted outstanding balance and the discounted cost of the collection effort, thus maximizing the net value of any given delinquent credit-card account.

History: Accepted by Noah Gans, stochastic models and simulation.

Keywords: account valuation • consumer credit • credit collections • singular control • self-exciting point process • stochastic optimization • control of Hawkes processes

Creditors have better memories than debtors.
—Benjamin Franklin

1. Introduction

The performance of numerous financial institutions critically depends on the efficient collection of outstanding unsecured consumer debt, a leading example of which is given by defaulted credit-card accounts. By the end of 2017, the aggregate revolving consumer credit in the United States well exceeded \$1 trillion (Federal Reserve Bank G.19, January 2018). At an average delinquency rate of 2.2% (FRB 2018), the resulting size of the delinquent debt pool of more than \$20 billion means that the issuing banks' exposure to nonperforming loans and their outstanding capital are highly sensitive to collection yields.

Based on the dynamic repayment model by Chehrazi and Weber (2015), we determine optimal collection strategies for overdue, so-called "delinquent" credit-card accounts. The repayment process, which for a given account specifies the timing and magnitude of the random repayments, is represented as a self-exciting point process in continuous time. The conditional arrival rate of future repayments (intensity) depends on the information revealed by past repayments and can be controlled by costly account-treatment actions. Both the magnitude and the timing of the account-treatment actions are subject to optimization. The cost of taking a given action is assumed to be a linear function of its impact on the point-process intensity. The bank's credit-collection problem is to determine a dynamic collection policy that maximizes the net present value of an account.

In an infinite-horizon setting, which for a bank and its affiliates amounts to a "going concern" for the account, the collection strategy becomes a time-invariant mapping from the action and holding subsets of the state space, characterized by the complement and the boundary of the two-dimensional inaction region, respectively, to the collection action set. Using the Bellman equation as a sufficient condition for optimality, we characterize the bank's optimal collection policy as a function of the "state" of an account in terms of its current repayment intensity and the remaining outstanding balance. The solution, which is obtained in quasi-closed form, exhibits a recursive structure in terms of the (maximal) number of repayments $i(w)$ for the outstanding balance w to drop below a balance threshold below which active collections no longer add value. The maximal number of iterations $i(w)$ can be precomputed for any account state as a function of the marginal account-treatment cost and the (empirically observed) account sensitivity with respect to collection actions.

The results have broad implications for collections management. First, given an identified model, a collections manager can use the optimal policy to determine the nature and timing of value-maximizing account-treatment actions. Second, the optimized (dynamic) net present value of an account implies at any point in time a threshold for early account settlement at less than 100% of the outstanding balance. Third, using the expected value of treated accounts in defaults, the bank can provide more precise estimates for the expected loss given default (LGD), which, in turn, enhances the bank's compliance with capital-reserve requirements as introduced in the Basel II accords. Moreover, the dependence of the minimum expected LGD on account attributes (and holder characteristics) may be used at the underwriting stage to curb excess exposure to risky clients and enhance the quality of the bank's loan portfolio. Finally, the results suggest new experiments for banks to better identify and test particular collection actions that show promise in improving the account workout.

1.1. Literature

1.1.1. Credit Management. The management of consumer credit addresses the three main phases in the life of an account: underwriting, treatment, and recovery (Rosenberg and Gleit 1994). Underwriting decisions concerning how much credit to grant (including no credit) usually depend on discriminant analysis and associated scoring methods (Bierman and Hausman 1970, Thomas 2000, Thomas et al. 2005). Treatment of an account in the absence of default on payments includes decisions about charge authorization, promotions, annual percentage rates, and credit-line extensions—be they positive or negative (Trench et al. 2003). In particular, credit risk can be reassessed repeatedly over the lifetime of an account (Crook et al. 2007). Finally, after a consumer defaults on a loan with respect to a specified repayment schedule, the credit-issuing bank enters recovery mode by putting the account in credit collections. The optimization of the associated account-treatment actions is the subject of this paper.

1.1.2. Credit Collections. In a seminal contribution, Mitchner and Peterson (1957) examine the optimal pursuit duration of credit collections as a stopping problem in continuous time. Based on historical data, they assume that the probability of repayment is decreasing in the pursuit time and find that it is optimal to stop collections when the expected repayment amount upon "conversion" of the account (from nonpayer to payer) equals the cost of pursuit measured in dollars per time unit. The authors realized that information is being revealed by the history of the repayment behavior:

> The problem of the collection of charged-off loans bears an interesting analogy to games such as certain types of poker in which betting takes place several times prior to the completion of a hand. In order to stay in the game it is necessary to continue betting. After each round of cards has been dealt, the player has additional information on which to base his decision as to whether to throw in his hand or to continue paying in order to stay in the game. (pp. 525–526)

Based on linear pursuit costs, Mitchner and Peterson (1957) estimated a gain of 33% in account repayment net of collection costs from applying their strategy. They further noted that in

> a fully rigorous treatment, the entire sequence of intervals between payments, as well as the payment amounts, should be taken into account as part of the over-all stochastic process. However, such a procedure would complicate the problem enormously. (pp. 537–538)

With the aid of recent advances in the theory of stochastic control, we formulate and solve the credit-collection problem in continuous time while allowing both the repayment amounts and inter-arrival times to be random and to influence each other.

After a 15-year hiatus of results, Liebman (1972) considers a discrete-time Markov decision problem based on transition probabilities conditional on characteristics such as account age, volume class (binary), and experience class (binary), which can be solved numerically using value iteration. With finer resolution of account characteristics, this approach suffers from the curse of dimensionality, thus diverting attention to numerical analysis and resulting in a lack of structural insights. Since the 1970s, the credit-collection problem remained again dormant in the literature until the Basel Committee on Banking Supervision (2004) introduced capital-reserve requirements based on LGD, which for delinquent credit-card debt corresponds to the expected discounted remaining outstanding balance plus the expected cost of collections. In a discrete-time Markov decision framework, Almeida Filho et al. (2010) examine the optimal timing of a given sequence of account-treatment actions, but their analysis does not condition the actions on account-specific repayment behavior. Our model is based on Chehrazi and Weber (2015), who propose a repayment model in terms of self-exciting point processes, which can be identified using the full variety of data sources available to the issuing bank. While their study focuses on the prediction of repayment behavior conditional on a sequence of account treatment actions, we are concerned here with the problem of determining an optimal sequence of treatment actions, including the timing of these interventions. To the best of our knowledge this is the first continuous-time treatment of the dynamic credit-collection problem.

From a behavioral viewpoint, consumers may be more impatient with respect to payoff delays in the short run than in the long run, which may lead to dynamically inconsistent preferences and a certain "present bias" (Laibson 1997, Bertaut et al. 2009).[1] Although the self-excitation feature in the assumed repayment process might at first sight suggest the contrary, our model makes no explicit assumption about debtors' preferences. Indeed, we show that a self-exciting point process is qualitatively consistent with a Bayesian variant of our model (examined in Appendix C). Our model does not explicitly consider the debtor's dynamic choice problem, involving the trade-off between consumption, debt repayment, and the possibility of personal bankruptcy, which might drive the debtor's repayment decisions. There is in fact at best weak evidence on consumers' strategizing personal bankruptcy decisions (Bertaut and Haliassos 2006). Moreover, the lack of debtors' bargaining power and banks' reputational cost of engaging in systematic debt renegotiations (decreasing its power to commit to future collections) both tend to reduce the salience of a strategic approach to the collection problem unlike situations of sovereign debt (Fernandez and Rosenthal 1990) or corporate debt (Roberts and Sufi 2009). Nevertheless, explicit strategic considerations may provide for interesting future extensions of this research.

1.1.3. Control of Jump Processes. The repayment process, which is described by the dynamics of the account state in the (intensity, balance) space, is a marked point process. In the intensity dimension, these dynamics are a self-exciting (Hawkes) jump process (see Hawkes 1971) with an additional predictable process that models account-treatment actions. The optimal schedule of treatment actions is a deterministic state-feedback law, but the implemented policy is ex ante stochastic as the account-state evolution (corresponding to the holder's repayment behavior) is random. Applications of self-exciting point processes in a financial setting originated with the description of counterparty risk (Jarrow and Yu 2001). More recently, they include models of financial contagion (Aït-Sahalia et al. 2015) and nonfinancial phenomena, such as terrorism (Porter and White 2012) or consumer response to online advertising (Xu et al. 2014). Our description of the evolution of self-exciting point processes is based on the transform analysis by Duffie et al. (2000) and Chehrazi and Weber (2015).

The optimal control of Markov jump processes for general (finite-dimensional) action and state spaces is examined by Pliska (1975) under the assumption that transition times are exponentially distributed and the corresponding state-dependent transition rates ("intensities" in our terminology) can be influenced by actions. While in that setting any transition rate remains constant between transitions, in our formulation the repayment intensity continues to evolve and can be actively controlled at any time. As a result, consistent with collections practice, account-treatment actions are effectively decoupled from waiting for repayments: they may be chosen so as to augment transition rates between state transitions. From this perspective, the collection problem is about the control of piecewise deterministic Markov processes (PDPs); the latter were introduced by Davis (1984) as a class of nondiffusion processes with applications in queueing (Dai 1995), R&D investment (Posner and Zuckerman 1990), and change-time detection (Bayraktar and Ludkovski 2009). Sufficient optimality conditions for systems with piecewise deterministic state dynamics can be obtained by exploiting martingale/submartingale properties of the value function.[2] Optimality conditions can often be cast in terms of quasi-variational inequalities (QVIs), a technique pioneered by Bensoussan and Lions (1982); these types of sufficient optimality conditions may be solved numerically (see Feng and Muthuraman 2010 for an example with diffusion).[3] An operator-based approach by Costa and Davis (1989) avoids the meticulous care needed when establishing existence and regularity of the value function as a viscosity solution of the QVIs.

The nature of solutions to optimal control problems, with and without diffusion, generally depends on the cost of control. With a fixed cost of taking an action, infrequent impulse-control interventions become optimal. Scarf (1960) shows that an (s, S) policy solves a firm's inventory-control problem: it is optimal to do nothing until the current stock drops below a trigger level s, at which point it is best to replenish the inventory to the level S. This type of policy, which features an inaction region in the state space, is optimal under a variety of assumptions on the stochasticity of demand (Harrison et al. 1983, Bensoussan et al. 2005). The same type of solution obtains in the context of holding an optimal amount of cash versus interest-bearing financial instruments (Bar-Ilan 1990) and optimal technology-investment decisions (Bar-Ilan and Maimon 1993). In the absence of fixed costs for taking an action, the solutions to stochastic optimization problems for diffusion processes (with or without jumps) usually feature instantaneous or barrier controls, where it is optimal to move the state continuously or hold it at the boundary ("barrier") of the action region (Dixit 1991). The implied three regions in the state space (inaction, action, and holding/transient) appear naturally in the control of systems with regime shifts (Guo et al. 2005). Henderson and Hobson (2008) find an analytical solution to an asset-selling/consumption problem that involves a diffusion without jumps, where it may be optimal to use both instantaneous controls and discrete impulse controls. Yang and Zhang (2005) consider a portfolio investment problem for an insurance company with a cash flow that is governed by a jump-diffusion process resulting from the uncertain arrival of claims. In their setting, the problem terminates when the insurance company is ruined; by contrast, the credit-collection problem ends once the outstanding balance has been repaid in full. The solution to the credit-collection problem presented here is exact and can in principle be obtained in closed form although the resulting expressions involve hypergeometric functions (as in Davis and Zervos 1998), which do not provide good intuition and are, therefore, omitted.

This work is closest to Bayraktar and Ludkovski (2009) who track the state of an unobservable Markov jump process by observing a compound Poisson process whose local characteristics (i.e., its intensity and mark distributions) depend on the state of the unobservable Markov jump process. Specifically, given a prior belief over the initial state of the unobservable Markov jump process, they obtain piecewise deterministic dynamics for the updated beliefs via Bayes' rule and, based on this, determine a tracking policy that maximizes an agent's expected net benefit. In principle, our problem may be cast into this formulation by viewing the repayment process as an observable compound Poisson process whose intensity is the unobservable Markov jump process. Our intensity can then be viewed as a Bayesian estimate of the true (unobservable) repayment intensity. Although the dynamics of our intensity are not exactly the same as in Bayraktar and Ludkovski (2009), it does have a similar structure for it features a positive jump when a repayment is received (exhibiting self-excitation) and declines exponentially during periods of inactivity. Our problem differs because our objective is to influence and not to track the state of an unobservable Markov jump process (corresponding to the repayment intensity), and therefore, our control variable is different. However, the solution technique is similar in that (exact) restrictions of the value function are obtained iteratively, indexed by the number of remaining potential interventions, which, in turn, are determined by the number of possible future contacts with the frontier of the action region conditional on minimum-size repayment events.

1.2. Outline

The remainder of this paper is organized as follows. In Section 2, we introduce the model for the repayment process and formulate the collection problem as a stochastic optimal control problem. Section 3 provides a sufficient optimality condition and then proceeds to the construction of a unique analytical solution to the collection problem for arbitrary initial data. In Section 4, we examine properties and practical aspects of the optimal collection strategy. Section 5 concludes. A summary of notation is provided in Table B.1 in the appendix.

2. Model

A credit-card account that misses the repayment deadline on its outstanding balance by more than a prespecified time period (e.g., 30 days) is called "delinquent" and placed in collections. To optimize the collection-treatment actions, consider such a delinquent account with outstanding balance w that enters the collection process at time $t=0$ to remain there for $t \geq 0$ until full repayment is received by the bank.[4]

2.1. Repayment Process

The holder of a delinquent account makes random repayments $Z_i \geq 0$ at random times $T_i > 0$ for i in $\{1, 2, \ldots\}$ until the balance is paid in full, at which point the account exits the collection process. As in Chehrazi and Weber (2015), the holder's repayment behavior is viewed as a self-exciting marked point process $(T_i, Z_i\colon i \geq 1)$ with piecewise-deterministic conditional arrival rate $\lambda(t)$ for $t \geq 0$ termed repayment intensity. This formulation is consistent with the Bayesian dynamics obtained by Bayraktar and Ludkovski (2009), where $\lambda(t)$ corresponds to the best estimate of the actual (unobserved) repayment intensity when the bank dynamically updates its belief on the holder's intentions to repay the debt (see Appendix C for analytical details).

Specifically, in the absence of any collection attempt ("account-treatment action"), the account holder's conditional repayment rate $\lambda(t+s)$ for $s \geq 0$ is given by the flow $\lambda(t+s) = \varphi(s, \lambda(t)) \triangleq \lambda_\infty + (\lambda(t) - \lambda_\infty)e^{-\kappa s}$ as long as there is no repayment in the time interval $(t, t+s]$. The mean reversion rate κ determines the speed with which the repayment intensity approaches its long-run steady state λ_∞; the latter generally depends on the macroeconomic environment.[5] In particular, when starting at $\lambda(t) > \lambda_\infty$, the arrival rate $\varphi(s, \lambda(t))$ is exponentially decreasing, dissipating the memory of past repayment behavior or account-treatment actions. Thus, if $\mathcal{N}(t)$ denotes the number of repayments on $[0, t]$, the conditional probability of the next repayment time $T_{\mathcal{N}(t)+1}$ to exceed $t+s$ is

$$P(T_{\mathcal{N}(t)+1} > t + s \mid \mathcal{F}_t) = \exp\left(-\int_0^s \varphi(\varsigma, \lambda(t))\, d\varsigma\right), \quad (1)$$

where $\mathcal{F}_t$, as an element of the standard information filtration $\mathbb{F}$,[6] contains all available information at time t. These dynamics are illustrated in Figure 2(b); they are consistent with Bayes' rule—a period of inactivity lowers the bank's estimate of the unobserved repayment intensity.

At any repayment time T_i, the holder's outstanding balance $W(T_i)$ diminishes by the received repayment amount Z_i, so $W(T_i) = W(T_i^-) - Z_i$. In line with institutional practice, we assume that, over the course of the collection process, the issuer does not accrue interest on the remaining outstanding balance. As a result, $W(t)$ remains constant between consecutive repayment events. The repayment sequence $(Z_i\colon i \geq 1)$ is obtained from a $[0,1]$-valued, independent and identically distributed (i.i.d.) sequence $(R_i\colon i \geq 1)$ via $Z_i = W(T_i^-)R_i = W(T_{i-1})R_i$. Each R_i follows the distribution $F_R\colon [0,1] \to [0,1]$ and corresponds to the percentage of the remaining outstanding balance (i.e., $W(T_{i-1})$), received with the repayment Z_i; the expected relative repayment is

$$\bar{r} \triangleq \mathbb{E}[R_i] = \int_0^1 r\, dF_R(r). \tag{2}$$

At any repayment event i, the holder's repayment intensity increases by $\delta_{10} + \delta_{11}R_i$, so $\lambda(T_i) = \lambda(T_i^-) + \delta_{10} + \delta_{11}R_i$. As detailed in Chehrazi and Weber (2015), the parameters δ_{10} and δ_{11} measure the holder's willingness to repay (i.e., the sensitivity of $\lambda(t)$ with respect to $\mathcal{N}(t) \triangleq \sum_i \mathbf{1}_{\{T_i \leq t\}}$) and ability to repay (i.e., the sensitivity of $\lambda(t)$ with respect to $\mathcal{R}(t) \triangleq \sum_i R_i \mathbf{1}_{\{T_i \leq t\}}$), respectively. For a given account, these two dimensions of the repayment process pertain to the number of observed repayments and the relative magnitude of these repayments: while frequent repayments tend to indicate a high willingness to repay, the account holder's ability to repay might still be quite low, for example, in the case where the holder's checks are bouncing or the repayments do not add up to a significant portion of the outstanding balance. Again, the positive jump introduced by a repayment is consistent with Bayes' rule: observing an event increases the bank's estimate of the repayment intensity (see Figure 2(b)).

Remark 1. In the absence of any account-treatment action, the "autonomous" repayment dynamics follow an affine, self-exciting point process (Hawkes process). In that case, the repayment intensity satisfies the stochastic differential equation (SDE)

$$d\lambda(t) = \kappa(\lambda_\infty - \lambda(t))dt + \delta_1^\top dJ(t), \tag{3}$$

where $J(t) \triangleq [\mathcal{N}(t), \mathcal{R}(t)]^\top$ denotes the repayment process with $\mathcal{N}(t)$ a counting process for the number of repayments and $\mathcal{R}(t)$ a process describing the relative magnitude of repayments. The parameter vector $\delta_1 \triangleq [\delta_{10}, \delta_{11}]^\top$ captures the account holder's willingness and ability to repay. The model parameters can be identified using estimation techniques developed by Chehrazi and Weber (2015), who also show the empirical significance of the self-excitation term for predicting consumers' repayment behavior based on a data set from the credit-card industry (see Appendix D for details).

2.2. Collection Problem

At any time in the collection process, the card issuer (bank) or a designated party (collection agency) can take account-treatment actions, usually in the form of attempts to collect all or part of the outstanding debt. These actions, in the form of establishing contact, negotiating a repayment plan, or filing a lawsuit, often do not lead to an immediate repayment but merely increase the likelihood of receiving a (partial) repayment in the future. The process $A = (A(t)\colon t \geq 0)$ with $A(0) = 0$ encapsulates the bank's collection strategy; it is nondecreasing, left-continuous (predictable), and adapted to $\mathbb{F}$. Taking a collection action produces a thrust in the dynamics of the repayment intensity,

$$d\lambda(t) = \kappa(\lambda_\infty - \lambda(t))dt + \delta_1^\top dJ(t) + \delta_2 dA(t), \tag{4}$$

where the sensitivity δ_2 describes the responsiveness of the account to its treatment. The change of A is viewed as a proxy for the collection effort, $dA(t)$, necessary to carry out the corresponding account-treatment action.[7] The cost of collection is assumed to be linear with constant marginal cost $c > 0$, consistent with a collection effort measured in hours spent on an account's workout. The bank's *collection problem* is to find an admissible collection strategy $A^* = (A^*(t)\colon t \geq 0) \in \mathcal{A}$ that solves

$$v^*(\lambda, w) = \min_{A \in \mathcal{A}} \mathbb{E}\left[\int_0^\infty e^{-\rho s}\, dW(s) + c\int_0^\infty e^{-\rho s}\, dA(s)\,\middle|\, \lambda(0) = \lambda, W(0) = w\right], \tag{P}$$

where $\mathcal{A}$ is the set of left-continuous, nondecreasing, adapted processes. An optimal collection strategy produces the minimum *net expected loss* $w + v^*(\lambda, w)$, consisting of the unrecovered outstanding balance plus collection cost given the initial intensity λ and initial balance w. Conversely, $\bar{V}^*(\lambda, w) \triangleq -v^*(\lambda, w)$ represents the (optimized) expected economic value of the outstanding loan, net of collection costs.

Although the bank's collection problem is formulated over an infinite horizon for a "collectable" account of positive value, the optimal pursuit time is endogenously bounded; see Section 3.2.2. Whether the pursuit time should be exogenously bounded or not depends on the specifics of a country's statute of limitations[8] for unsecured consumer debt or else on bank-internal guidelines. An infinite horizon reflects a going concern for an account and naturally accommodates optimal stopping problems, for example, by including considerations of early settlement.

3. Optimal Account Treatment

The optimal solution to the collection problem (P) consists of a sequence of account-treatment actions. Treatment actions are taken when the account's repayment intensity falls below an intensity threshold determined by the account's current outstanding balance and its characteristics. The intensity threshold reflects a minimal arrival likelihood for repayments that needs to be maintained throughout an active collection process. All else equal, it increases in an account's outstanding balance: when more money is owed, the bank needs to ensure a greater likelihood of repayment and is, therefore, more willing to invest in an active treatment of the account. In addition, for high-balance accounts, more forceful treatment actions (e.g., threatening with a lawsuit) are warranted. In our formulation, implementing an account-treatment action (such as establishing first-party contact, entering negotiations for a payment plan, or filing a lawsuit) can increase the repayment intensity to a certain level (by introducing an intensity jump) and maintain the intensity at this level (by providing a positive intensity thrust) as long as the action remains in effect. In cases where the initial repayment intensity, at the start of the collection process, is below the intensity threshold, an optimal account-treatment action leads to an intensity jump. The impact of the first action includes the concomitant status change of an account when entering the collection process, reflecting also the fact that the account holder's attention has been captured. Subsequent actions do not induce such jumps because they are taken at the optimal time. Instead they tend to maintain a designated level of intensity until the next repayment event triggers a shift in the outstanding balance and repayment intensity, upon which the bank relaxes the collection effort until the repayment likelihood continuously drops to the (now lower) intensity threshold once more. The collection process terminates as soon as the outstanding balance reaches an *economic balance threshold* below which account treatment is no longer economically viable. From a mathematical perspective, the (unique) optimal collection strategy consists of an impulse control at the time of placement ($t = 0$), followed by a series of extended treatment intervals with continuous control. Note that an arbitrary collection strategy can consist of more than one impulse, for example, because subsequent actions are taken when the repayment intensity has already fallen below the optimal threshold (and consequently below the intensity levels these actions can maintain). Next, we provide a detailed analysis and formal description of the solution to the collection problem (P).

3.1. Sufficient Optimality Condition

Without loss of generality, any admissible collection strategy $A \in \mathscr{A}$ can be represented in the form

$$A(t) = \int_{[0,t]} E(s)\,ds + \sum_{k=0}^{\infty} \Delta A_k \mathbf{1}_{\{\vartheta_k < t\}}, \quad t \geq 0, \qquad (5)$$

where the adapted nonnegative process $E(t)$ describes an infinitesimal extended collection effort applied on a time interval and where the predictable jumps $\Delta A_k \triangleq A(\vartheta_k^+) - A(\vartheta_k)$ at $\mathbb{F}$-stopping times ϑ_k for $k \geq 0$ (with $\vartheta_{k+1} > \vartheta_k$) correspond to concentrated collection efforts, each of which has a sizeable impact on the account's repayment likelihood. In this formulation, an account-treatment action (such as establishing first-party contact, entering negotiations for a payment plan, or filing a lawsuit) is modeled by a fixed intensity level $\hat{\lambda}$; see Section 3.2.3. If the current intensity $\lambda(t)$ is less than $\hat{\lambda}$, implementing that action increases the current intensity to this level through a jump of size $\hat{\lambda} - \lambda(t)$. For example, initiating negotiation for a repayment plan or filing a lawsuit can induce a jump in the repayment intensity as these actions may change the account holder's priorities. While in effect, this action maintains the intensity at $\hat{\lambda}$ by means of the continuous thrust $\kappa(\hat{\lambda} - \lambda_\infty)$, which captures the effect of the action while it is implemented, for example, until an agreement for a repayment plan is reached or a court order is obtained.[9] We allow for an action's induced intensity level $\hat{\lambda}$ to depend on account-specific information. For example, for a given current intensity $\lambda(t)$, establishing and maintaining first-party contact through phone calls, emails, or text messages can cause a sizeable intensity transition for "high-quality" accounts while it may have virtually no impact on the repayment intensity of "low-quality" accounts. For low-quality accounts, achieving the same effect may require a stronger and more expensive treatment action, such as filing a lawsuit.[10] In our formulation, we assume that any intensity level can, in principle, be attained and maintained as long as the bank is able to absorb the corresponding collection cost. To arrive at an actual practical implementation of an "optimal" collection strategy A^*, it is generally necessary to determine a workable approximation using the bank's available collection toolbox. Effective policy-approximation techniques for a given finite set of feasible account interventions are the subject of a future empirical study; see Appendix D for further details.

Given the representation of admissible collection strategies in Equation (5), a sufficient optimality condition for a solution to the collection problem (P) is obtained using a standard martingale argument; see Endnote 2.

Theorem 1. *Let $v^*: \mathbb{R}_+^2 \to \mathbb{R}$ be a continuous, almost everywhere (a.e.) differentiable function that satisfies the Bellman equation*

$$0 = \min\Big\{\min_{a>0}\{v^*(\lambda + \delta_2 a, w) + ca - v^*(\lambda, w)\},$$
$$\min_{\epsilon \geq 0}(\mathcal{D}_\epsilon v^*)(\lambda, w)\Big\}, \quad (\lambda, w) \geq 0, \qquad (6)$$

and the boundary condition $\lim_{\lambda\to\infty} v^*(\lambda, w) = -w$, where for any $\epsilon \geq 0$ and for any differentiable function $v: \mathbb{R}_+^2 \to \mathbb{R}$

the integro-differential operator $\mathcal{D}_\epsilon$ is defined by

$$(\mathcal{D}_\epsilon v)(\lambda,w) \triangleq [\kappa(\lambda_\infty-\lambda)+\delta_2\epsilon]\partial_1 v(\lambda,w)-\rho v(\lambda,w) + \lambda\mathbb{E}[v(\lambda+\delta_{10}+\delta_{11}R, w(1-R))-v(\lambda,w)-wR]+c\epsilon, \quad (\lambda,w)\geq 0. \qquad (7)$$

Then v^* is the value function of the collection problem (P).

The Bellman equation (6) establishes the optimality of a candidate value function by testing it for any possible improvement. Specifically, the first term of the minimand on the right-hand side expresses the best net reduction of the expected loss as a result of any *positive* discrete effort.[11] The second term of this minimand corresponds to the best expected net reduction of the objective for any admissible continuous effort. For $\epsilon=0$, this term (when set to zero) guarantees that a candidate value function is consistent with the autonomous evolution of the intensity (see Equation (3)) when no account-treatment action is applied. To guarantee that the value function is optimal, it is sufficient that neither of the two terms is negative, so their minimum on the right-hand side of Equation (6) must vanish. Intuitively, the Bellman equation ensures that neither a discrete nor a continuous collection effort can lead to an improvement in the expected net loss and, in addition, that, if inactivity is optimal, then the account-state evolution is guaranteed to be autonomous. We also note that a solution to the Bellman equation is, by construction, unique.

Remark 2. At an optimum, the value of the first term in the variational identity (6), when it exists, can generally be achieved using one of several (and possibly a continuum of) different simultaneous discrete collection efforts. Because the loss is bounded by the outstanding balance w, any optimal discrete effort $a^*(\lambda,w)$ is naturally bounded by $w/c<\infty$. By the Berge maximum theorem, the set of solutions to a parameterized optimization problem over a compact set (here $[0,w/c]$) with continuous objective function is compact-valued (Berge 1963, p. 116). In particular, there exists a maximal optimal discrete collection effort,

$$a^*(\lambda,w)=\max\left\{a\geq 0: a\in\arg\min_{\hat{a}\geq 0}\{v^*(\lambda+\delta_2\hat{a},w)+c\hat{a}\}\right\}, \quad (\lambda,w)\geq 0, \qquad (8)$$

which satisfies the Bellman equation (6) when it is strictly positive. This solution is chosen here to exclude an uncountable number of possible sequences of quasi-simultaneous discrete interventions, all of which are theoretically equivalent to implementing the maximal intensity jump $a^*(\lambda,w)$. This is in agreement with our view that, in practice, an account-treatment action (such as establishing first-party contact, entering negotiations for a payment plan, or filing a lawsuit) can only increase the repayment intensity to a certain level from a currently smaller level $\lambda(t)$; taking the same action multiple times would not lead to a larger intensity jump.

Corollary 1. *Equation (6) partitions the attainable account states $(\lambda,w)\geq 0$ (i.e., the "state space" $\mathbb{R}^2_+$) into three disjoint regions:*

(i) *an "action region,"*

$$\mathcal{A}^* \triangleq \left\{(\lambda,w)\in\mathbb{R}^2_+: a^*(\lambda,w)>0 \text{ and } \min_{\epsilon\geq 0}(\mathcal{D}_\epsilon v^*)(\lambda,w)\geq 0\right\}, \qquad (9)$$

where $v^*(\lambda+\delta_2 a^*(\lambda,w),w)+ca^*(\lambda,w)-v^*(\lambda,w)=0$;

(ii) *a "holding region,"*

$$\mathcal{H}^* \triangleq \{(\lambda,w)\in\mathbb{R}^2_+: a^*(\lambda,w)=0,\ \lambda>\lambda_\infty, \text{ and } (\epsilon\geq 0\Rightarrow(\mathcal{D}_\epsilon v^*)(\lambda,w)=0)\}; \qquad (10)$$

(iii) *a "continuation region,"*

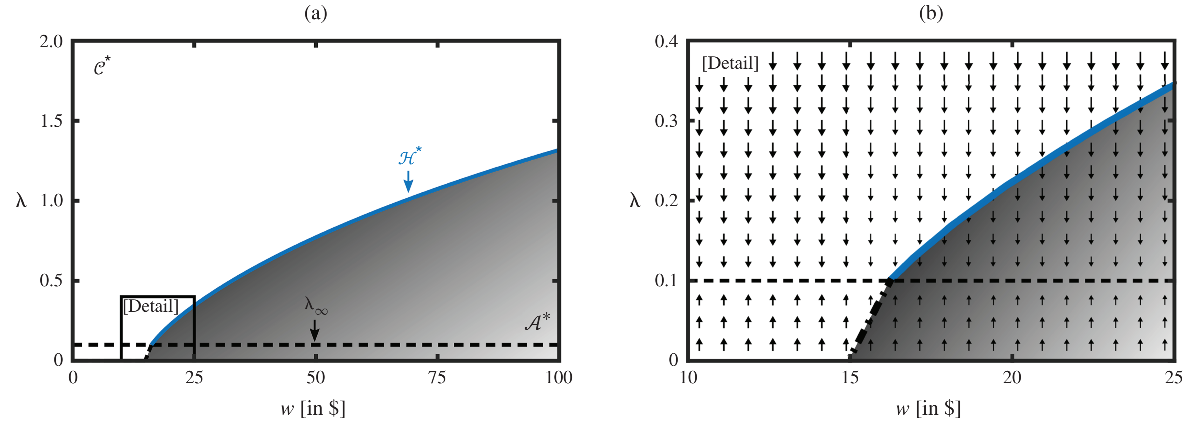
$$\mathcal{C}^* \triangleq \{(\lambda,w)\in\mathbb{R}^2_+: a^*(\lambda,w)=(\mathcal{D}_0 v^*)(\lambda,w)=0, \text{ and } (\lambda\geq\lambda_\infty,\epsilon>0\Rightarrow(\mathcal{D}_\epsilon v^*)(\lambda,w)>0)\}. \qquad (11)$$

Thus, $\mathcal{A}^*\cup\mathcal{H}^*\cup\mathcal{C}^*=\mathbb{R}^2_+$.

Given any two regions, the third can be obtained as their complement in the state space $\mathbb{R}^2_+$. These regions are illustrated in Figure 1(a). The action region $\mathcal{A}^*$ comprises all points of the state space at which an immediate discrete collection effort is best. The holding region $\mathcal{H}^*$ is formed by points on the boundary of the action region $\mathcal{A}^*$, from which—in the absence of any collection effort—the account state would drift immediately into $\mathcal{A}^*$;[12] see Figure 1(b). From within $\mathcal{A}^*$, the optimal discrete collection effort moves the account state to the holding region $\mathcal{H}^*$ where it becomes best to apply a continuous collection effort so as to keep the debtor engaged, thus maintaining the current repayment intensity. Finally, the complement of $\mathcal{A}^*\cup\mathcal{H}^*$ is the continuation (or inaction) region $\mathcal{C}^*$ where collection actions are economically inefficient.

Remark 3. (i) If $(\lambda,w)\in\mathcal{A}^*$, then by Equation (8) in Remark 2 it is $(\lambda+\delta_2 a^*(\lambda,w),w)\notin\mathcal{A}^*$: applying the (maximal) optimal discrete collection effort leads to a state outside of the action region. Indeed, Figure 2 shows a situation where an initial account state lies strictly inside the action region $\mathcal{A}^*$ and where the optimal discrete collection effort moves the account state to the holding region. (ii) For a corner solution, where $a^*(\lambda,w)=0$ by the first-order necessary optimality condition: $\partial_1 v^*(\lambda,w)\geq-\hat{c}$, where $\hat{c}\triangleq c/\delta_2$. In particular, since $(\mathcal{D}_\epsilon v^*-\mathcal{D}_0 v^*)(\lambda,w)=(\partial_1 v^*(\lambda,w)+\hat{c})(\delta_2\epsilon)$, if $\partial_1 v^*(\lambda,w)=-\hat{c}$, then necessarily $(\mathcal{D}_\epsilon v^*)(\lambda,w)=0$ for all

Figure 1. (a) Partition of the State Space; (b) Vector Field Corresponding to the Autonomous Intensity Dynamics Between Two Repayment Events

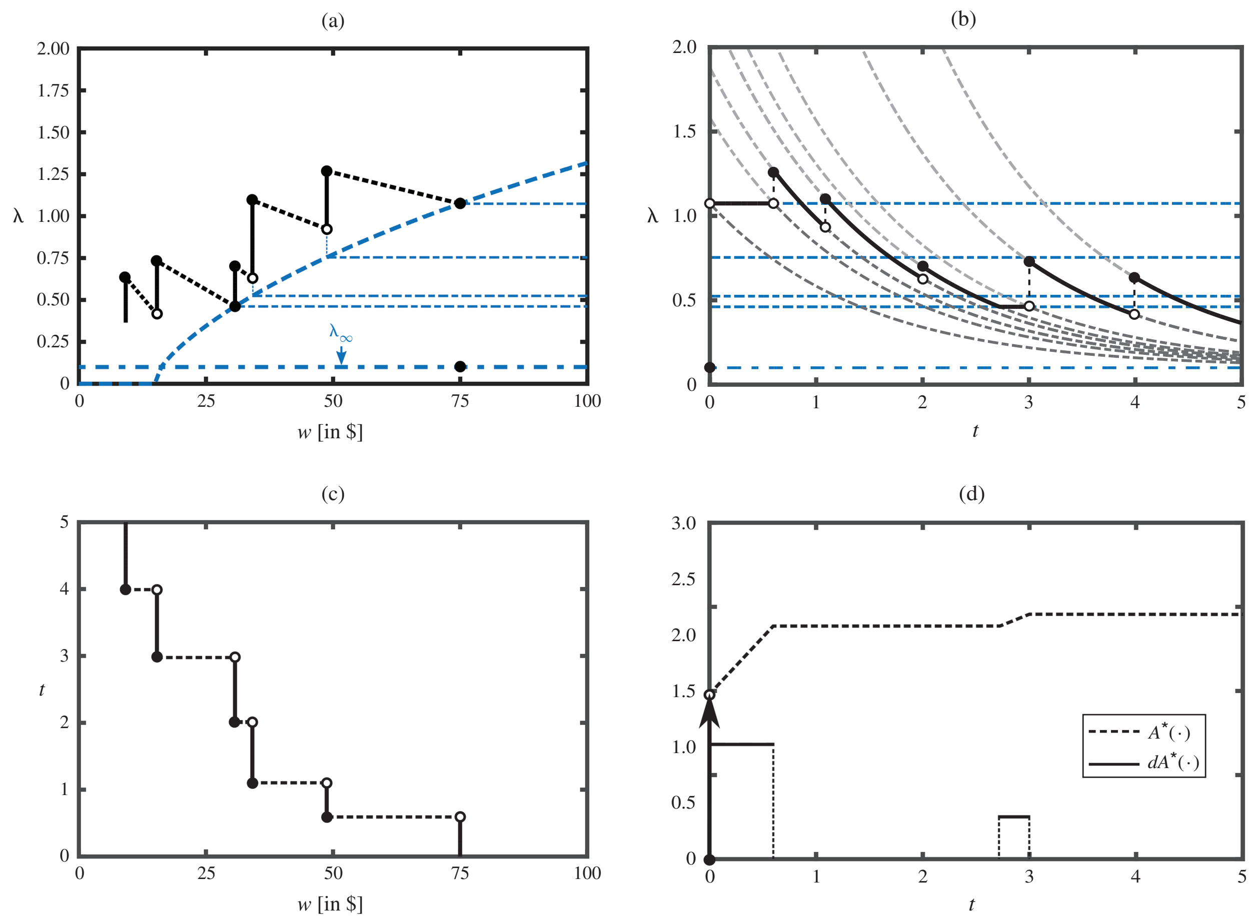


$\epsilon \geq 0$, so $(\lambda, w) \in \mathcal{H}^*$, provided $\lambda > \lambda_\infty$. By choosing the continuous collection effort

$$\varepsilon^*(\lambda, w) \triangleq (\kappa/\delta_2)(\lambda - \lambda_\infty)\mathbf{1}_{\{(\lambda,w)\in\mathcal{H}^*\}}, \qquad (12)$$

the account state in the holding region remains in the holding region $\mathcal{H}^*$ until a repayment is received since, by construction, $\kappa(\lambda_\infty - \lambda) + \delta_2\varepsilon^*(\lambda, w) = 0$. This is depicted in Figure 2 where for $(\lambda(t), w) \in \mathcal{H}^*$ a minimal continuous collection effort is applied to keep the value of the intensity $\lambda(t)$ fixed and to prevent the account state $(\lambda(t), w)$ from entering into the action region $\mathcal{A}^*$. Finally, for $\partial_1 v^*(\lambda, w) > -\hat{c}$, clearly $(\mathcal{D}_0 v^*)(\lambda, w) \geq 0$ implies that $(\mathcal{D}_\epsilon v^*)(\lambda, w) > 0$, for any $\epsilon > 0$.

Figure 2. Sample Paths Under Optimal Policy for the Collection Problem (P): (a) Account State; (b) Repayment Intensity; (c) Outstanding Balance; (d) Collection Effort (Discrete and Continuous)



3.2. Constructing the Value Function

To construct the value function v^*, it is possible to use a recursive approach, which, for any given initial account state (λ, w), converges in a finite number of steps and provides in each iteration an exact solution restricted to a subset of the attainable subsequent account states. For this, Section 3.2.1 derives a semi-analytic solution to the collection problem (P) in the absence of any collection activity, referred to as "autonomous account value." In Section 3.2.2, we consider a "terminal collection problem." Its solution determines "terminal" account states, for which the "autonomous account value" satisfies the Bellman equation (6). Any state trajectory $(\lambda(t), W(t))$ reaches a terminal state with probability one in finite time. Hence, the solution in this part of the state space is referred to as the "terminal account value." The latter can be precomputed and, thus, initializes the recursion. Section 3.2.3 introduces a family of "sustained extensions," indexed by the holding intensities $\hat{\lambda} \geq \lambda_\infty$, to extend the validity of a value function to generate its next iteration by finding the (unique) optimal holding intensity. Finally, Section 3.2.4 establishes the optimality of the recursion and provides expressions for the optimal collection strategy. The recursion extends the solution of the collection problem (P) from terminal account states to the entire state space.

3.2.1. Autonomous Account Value. In the absence of any account-treatment action, the stochastic evolution of the repayment intensity follows the SDE (3), and the collection cost vanishes. Hence, the value of the untreated account in (P) becomes (minus) the expected discounted value of the autonomous repayment sequence $(Z_i\colon i \geq 1)$,

$$u(\lambda, w) = \mathbb{E}\left[-\sum_{i=1}^{\infty} e^{-\rho T_i} Z_i \,\middle|\, \lambda(0) = \lambda,\, W(0) = w\right] = \mathbb{E}\left[\int_0^\infty e^{-\rho s}\, dW(s) \,\middle|\, \lambda(0) = \lambda, W(0) = w\right], \quad (13)$$

for all $(\lambda, w) \geq 0$.

Theorem 2. *For any $(\lambda, w) \geq 0$, the autonomous account value can be represented in the form*

$$u(\lambda, w) = -\left(1 - \rho \int_0^\infty \exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\, dt\right) w, \quad (14)$$

where (α, β): $\mathbb{R}_+ \to \mathbb{R}^2$ uniquely solves the initial-value problem

$$\dot{\alpha}(t) = -\kappa\alpha(t) + \int_0^1 [1 - (1 - r) \cdot \exp(-(\delta_{10} + \delta_{11} r)\, \alpha(t))]\, dF_R(r), \quad \alpha(0) = 0, \quad (15)$$

$$\dot{\beta}(t) = \lambda_\infty\, \alpha(t), \quad \beta(0) = 0, \quad (16)$$

for all $t \geq 0$.

The representation of the autonomous account value in Equation (14) decomposes the effects of time value of money (via the discount factor $e^{-\rho t}$), the self-excitation of the arrival intensity (via the discount factor $e^{-\lambda\alpha(t)}$), and its reversion toward the long-run steady state λ_∞ (via the discount factor $e^{-\kappa\beta(t)}$). For example, if repayments can be made only in full (so F_R becomes a Dirac distribution at $r = 1$) and there is no mean reversion (so $\kappa = 0$), then the repayment intensity is constant, $\lambda(t) \equiv \lambda$, and $(\alpha(t), \beta(t)) = (t, \lambda_\infty t^2/2)$, which by Equation (14) implies that $u(\lambda, w) = -(\lambda w)/(\rho + \lambda)$.

Corollary 2. *The autonomous account value in Equation (13) is such that*

(i) $(\mathcal{D}_0 u)(\lambda, w) = 0$ *on* $\mathbb{R}_+^2$ *with the boundary condition* $\lim_{\hat{\lambda}\to\infty} u(\hat{\lambda}, w) = -w$;

(ii) $\partial_1 u(\lambda, w), \partial_2 u(\lambda, w), \partial_{12} u(\lambda, w) < 0 = \partial_{22} u(\lambda, w) < \partial_{11} u(\lambda, w)$ *on* $\mathbb{R}_+ \times \mathbb{R}_{++}$;

(iii) $\partial_2 u(\lambda, 0), \partial_{12} u(\lambda, 0) < 0 = u(\lambda, 0) = \partial_1 u(\lambda, 0) = \partial_{11} u(\lambda, 0) = \lim_{\hat{\lambda}\to\infty} \partial_1 u(\hat{\lambda}, w)$, *and* $\lim_{\hat{w}\to\infty} u(\lambda, \hat{w}) = \lim_{\hat{w}\to\infty} \partial_1 u(\lambda, \hat{w}) = -\infty$.

By the preceding result, the expected loss, $w + u(\lambda, w)$, of an untreated account is decreasing and convex in the repayment intensity λ while it is increasing and linear in the outstanding balance w; see Figures 3(c) and 3(d). Moreover, the expected loss exhibits decreasing differences in the sense that a positive change in the repayment intensity decreases the expected loss more for higher outstanding balances; see Figure 3(b).

Remark 4. The solution (α, β): $\mathbb{R}_+ \to \mathbb{R}^2$ to the initial-value problem (15)–(16) is increasing and such that $0 \leq \alpha(t) \leq (1 - e^{-\kappa t})/\kappa$ and $0 \leq \beta(t) \leq (\lambda_\infty/\kappa^2)(\kappa t - (1 - e^{-\kappa t}))$, for all $t \geq 0$. Thus, by Equation (14), the autonomous account value can be bounded from below for all $(\lambda, w) \in [\lambda_\infty, \infty) \times \mathbb{R}_+$:

$$u(\lambda, w) \geq -\left(1 - \frac{\rho e^{-(\lambda - \lambda_\infty)/\kappa}}{\rho + \lambda_\infty}\right) w. \quad (17)$$

The last inequality implies a lower bound for expected loss given default in the absence of active collections,

$$\text{LGD} = w + u(\lambda, w) \geq \left(\frac{\rho w}{\rho + \lambda_\infty}\right) e^{-(\lambda - \lambda_\infty)/\kappa}, \quad (18)$$

which may be used to determine the capital at risk in the issuing bank's delinquent account portfolio.

3.2.2. Terminal Account Value. An immediate implication of Corollary 2(i) is that $u(\lambda, w)$ is the value function of the collection problem (P) for any account state (λ, w) in a *subset* of the continuation region $\mathcal{C}^*$, for which the trajectory $(\lambda(t), W(t))_{t\geq 0}$ with $(\lambda(0), W(0)) = (\lambda, w)$ is guaranteed to remain in the continuation region. For such account states, Remark 3(i) implies that the collection problem (P) simplifies to

$$v_0^*(\lambda, w) = \min_{a_0 \geq 0}\{u(\lambda + \delta_2 a_0, w) + c a_0\}, \quad \text{(TCP)}$$

where, upon expending the optimal effort $a_0^*(\lambda, w)$, the bank obtains the autonomous value function at the new intensity. We refer to this problem as the *terminal collection problem* since—no matter if the outstanding debt is paid in full or not—it will never be optimal to attempt active collections again.

Note that since $\partial_{12}u < 0 \le \partial_{11}u$ by Corollary 2, the minimand of (TCP) is supermodular in (a_0, λ) and submodular in (a_0, w) so that the discrete terminal collection effort $a_0^*(\lambda, w)$ is decreasing in λ and increasing in w. Hence, a trivial discrete terminal effort $a_0^*(\lambda, w) = 0$ implies that the optimal account treatment remains trivial for larger intensities and smaller balances; that is, $a_0^*(\hat\lambda, \hat w) = 0$ for all $(\hat\lambda, \hat w)$ with $\hat\lambda \ge \lambda$ and $\hat w \le w$. On the other hand, any nontrivial $a_0^*(\lambda, w) > 0$ satisfies the necessary (and here also sufficient) optimality condition by Fermat,

$$\partial_1 u(\lambda + \delta_2 a_0^*(\lambda, w), w) = -\hat c, \tag{19}$$

where $\hat c = c/\delta_2$, as in Remark 3, denotes the effective marginal collection cost, that is, the marginal collection cost c relative to the sensitivity δ_2 of the intensity process with respect to the collection effort. Thus, the smallest balance $\underline{w}$ above which the terminal collection effort remains nontrivial for *some* positive λ is such that $\partial_1 u(0, \underline{w}) = -\hat c$, that is, by Theorem 2,

$$\underline{w} \triangleq \left(\int_0^\infty \alpha(t)\exp[-\rho t - \kappa\beta(t)]\,dt\right)^{-1}\left(\frac{\hat c}{\rho}\right).$$

For any w exceeding the *minimal actionable balance* $\underline{w}$, consider the intensity threshold $\lambda_0^*(w)$, below which the discrete terminal collection effort is nontrivial. By the optimality condition (19), $\lambda_0^*(w)$ is the unique solution of $\partial_1 u(\lambda, w) = -\hat c$ for $w > \underline{w}$. Note that, similar to $a_0^*(\lambda, w)$, submodularity of u implies that $\lambda_0^*(w)$ is (strictly) increasing in w with $\lim_{w\to\infty}\lambda_0^*(w) = \infty$.[13] Defining $\lambda_0^*(w) \triangleq 0$ for $w \le \underline{w}$, we can write the terminal value function in (TCP) as

$$v_0^*(\lambda, w) = \begin{cases} u(\lambda, w), & \text{if } \lambda \ge \lambda_0^*(w), \\ u(\lambda + \delta_2 a_0^*(\lambda, w), w) + c a_0^*(\lambda, w), & \text{otherwise}, \end{cases} \tag{20}$$

where for $\lambda \le \lambda_0^*(w)$, it is $\lambda + \delta_2 a_0^*(\lambda, w) = \lambda_0^*(w)$. Let w_0^* be the unique balance w at which the intensity threshold $\lambda_0^*(w)$ is equal to the long-run steady-state intensity λ_∞; that is, $\lambda_0^*(w_0^*) = \lambda_\infty$. Then, by Corollary 1, for balances $w \in [0, w_0^*]$ we conclude that the holding region $\mathcal{H}^*$ is empty, and starting from $\lambda \ge \lambda_0^*(w)$, $\lambda(t)$ never becomes smaller than $\lambda_0^*(w)$. Consequently, in this region, the solution to the terminal collection problem (TCP) coincides with the solution to the (general) collection problem (P).

Theorem 3. *For any $(\lambda, w) \in \mathbb{R}_+ \times [0, w_0^*]$, the value function $v^*(\lambda, w)$ of the collection problem (P) is equal to the value function $v_0^*(\lambda, w)$ of the terminal collection problem (TCP).*

The following properties follow directly from Corollary 2.

Corollary 3. *For any $(\lambda, w) \in \mathbb{R}_+ \times [0, w_0^*]$, $v_0^*(\lambda, w)$ is C^1 and has second-order derivatives a.e. Moreover,*
(i) $0 \ge v_0^*(\lambda, w) \ge -w = \lim_{\hat\lambda\to\infty} v_0^*(\hat\lambda, w)$;
(ii) $-\hat c \le \partial_1 v_0^*(\lambda, w) \le 0 = \lim_{\hat\lambda\to\infty}\partial_1 v_0^*(\hat\lambda, w)$;
(iii) $\partial_2 v_0^*(\lambda, w), \partial_{22} v_0^*(\lambda, w), \partial_{12} v_0^*(\lambda, w) \le 0 \le \partial_{11} v_0^*(\lambda, w)$.

Based on the fact that, by Theorem 3 for a sufficiently small outstanding balance the collection problem (P) can be reduced to a terminal collection problem, the optimal collection strategy can be inferred from the value function v_0^* characterized in Equation (20).

Corollary 4. *For account states in $\mathbb{R}_+ \times [0, w_0^*]$, the continuation, holding, and action regions are given by*

$$\mathcal{C}_0^* = \mathbb{R}_+ \times [0, \underline{w}] \cup [\lambda_0^*(w), \infty) \times (\underline{w}, w_0^*], \quad \mathcal{H}_0^* = \varnothing, \quad \text{and} \quad \mathcal{A}_0^* = [0, \lambda_0^*(w)) \times (\underline{w}, w_0^*],$$

respectively. The optimal collection strategy for an account in such states, that is, the optimal solution to the collection problem (P) is $A^*(t) \equiv a_0^*(\lambda, w)$ for $t > 0$.

The intuition behind Theorem 3 and the optimal collection strategy in Corollary 4 is that for accounts with an outstanding balance below w_0^* it is *either* economically inefficient to attempt collection *or else* worthwhile to make only a single collection attempt (e.g., by establishing contact via a phone call, an email, or a letter to remind the holder about the overdue balance). Expending further collection effort is not advised since the additional collection cost will exceed the expected recovery. Figure 3(a) depicts $\underline{w}$ and w_0^* for the case studied in Section 4.

3.2.3. Sustained Extensions. For any $\lambda, \hat\lambda \ge \lambda_\infty$ with $\lambda \ge \hat\lambda$, consider the cumulative distribution function of a "sustained" intensity process according to SDE (3), which starts at $\lambda(0) = \lambda$, is bounded from below by $\hat\lambda$, and proceeds until the next repayment event at a random time $T \ge 0$,

$$\mathbb{P}(T \le t \mid \lambda(0) = \lambda) = F_{\lambda,\hat\lambda}(t) \triangleq 1 - \exp[-\Phi_{\lambda,\hat\lambda}(t)], \quad t \ge 0,$$

where $\theta(\lambda, \hat\lambda) \triangleq -(1/\kappa)\ln((\hat\lambda - \lambda_\infty)/(\lambda - \lambda_\infty)) \ge 0$ is the critical time $\hat t \ge 0$ at which $\lambda(\hat t) = \varphi(\hat t, \lambda) = \hat\lambda$, and

$$\Phi_{\lambda,\hat\lambda}(t) \triangleq \int_0^t \max\{\varphi(s, \lambda), \hat\lambda\}\,ds = \begin{cases} \lambda_\infty t + (\lambda - \varphi(t, \lambda))/\kappa, & \text{if } 0 \le t \le \theta(\lambda, \hat\lambda), \\ \lambda_\infty\theta(\lambda, \hat\lambda) + (\lambda - \hat\lambda)/\kappa + \hat\lambda(t - \theta(\lambda, \hat\lambda)), & \text{otherwise}. \end{cases}$$

Since $\Phi_{\lambda,\hat\lambda}$ is strictly increasing in $(\lambda, \hat\lambda)$, a decrease in the initial value λ or a decrease in the lower bound $\hat\lambda$

produces a first-order stochastically dominant shift of the cumulative distribution function $F_{\lambda,\hat{\lambda}}$. The payment distribution is stochastically delayed if its intensity starts at a lower value or is sustained at a smaller lower bound.

Lemma 1. *The expected present value of a unit repayment under a sustained intensity process is*

$$\mathbb{E}[e^{-\rho T} \mid \lambda(0)=\lambda] = \int_{\hat{\lambda}}^{\lambda} Q(\lambda, l)\frac{\rho + l}{\kappa(l-\lambda_\infty)}\,dl + Q(\lambda,\hat{\lambda}), \tag{21}$$

where, for $\lambda \geq \hat{\lambda} \geq \lambda_\infty$,

$$Q(\lambda,\hat{\lambda}) = \frac{\hat{\lambda}}{(\rho+\hat{\lambda})}\left(\frac{\hat{\lambda}-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\rho+\lambda_\infty)/\kappa}\exp\left(-\frac{\lambda-\hat{\lambda}}{\kappa}\right) \tag{22}$$

represents $\mathbb{E}[e^{-\rho T} \mid T \geq \theta(\lambda,\hat{\lambda}), \lambda(0)=\lambda]\cdot\mathbb{P}(T \geq \theta(\lambda,\hat{\lambda}) \mid \lambda(0)=\lambda)$, *that is, the expected present value of a unit repayment while the intensity is being sustained at* $\hat{\lambda}$.

Given the current state $(\lambda(0), W(0)) = (\lambda, w) \in \mathbb{R}_+^2$, let $(\lambda(T), W(T))$ be the state that can be reached *after* the next repayment at T. Moreover, let $v(\lambda(T), W(T))$ be (minus) the expected net account value right after the repayment at T by following an admissible collection strategy $A \in \mathscr{A}$ over $[T,\infty)$. Consider now an extension of this policy that prescribes maintaining the repayment intensity above $\hat{\lambda}$ until the repayment time T. After a repayment, the original collection strategy A (which leads to the value function v) is applied. For $\lambda \geq \hat{\lambda}$, by Lemma 1, this "sustained-extension strategy" lowers the expected loss because

$$\begin{aligned}&\mathbb{E}[\bar{v}(\lambda(T^-), W(T^-))e^{-\rho T} \mid \lambda(0)=\lambda, W(0)=w]\\ &\quad= \int_{\hat{\lambda}}^{\lambda} \bar{v}(l,w)\,Q(\lambda,l)\frac{l+\rho}{\kappa(l-\lambda_\infty)}\,dl + \bar{v}(\hat{\lambda},w)Q(\lambda,\hat{\lambda}),\end{aligned} \tag{23}$$

where

$$\bar{v}(l,w) \triangleq \int_0^1 (v(l+\delta_{10}+\delta_{11}r, (1-r)w) - rw)\,dF_R(r) \tag{24}$$

is (minus) the expected value of the account in state (l,w) given a (partial) repayment. To maintain the intensity at $\hat{\lambda}$ after having reached it, the bank needs to expend a continuous collection effort $E(s) = (\kappa/\delta_2)\cdot(\hat{\lambda}-\lambda_\infty)$ for $s \geq \theta(\lambda,\hat{\lambda})$. The expected present value of the cost for sustaining the repayment intensity at $\hat{\lambda}$ until the next repayment is, therefore,

$$\int_{\theta(\lambda,\hat{\lambda})}^{\infty}\left(c\int_{\theta(\lambda,\hat{\lambda})}^{s} e^{-\rho\varsigma}E(\varsigma)\,d\varsigma\right)dF_{\lambda,\hat{\lambda}}(s) = \hat{c}\kappa\frac{\hat{\lambda}-\lambda_\infty}{\hat{\lambda}}Q(\lambda,\hat{\lambda}). \tag{25}$$

Hence, given any outstanding balance $w \geq 0$, the expected (minus) net account value under the sustained-extension strategy can be described by the *sustained-extension operator*

$$\begin{aligned}(\mathcal{S}_{\hat{\lambda}}v)(\lambda,w) \triangleq &\int_{\hat{\lambda}}^{\lambda} \bar{v}(l,w)\,Q(\lambda,l)\frac{l+\rho}{\kappa(l-\lambda_\infty)}\,dl\\ &+\left[\bar{v}(\hat{\lambda},w) + \hat{c}\kappa\frac{\hat{\lambda}-\lambda_\infty}{\hat{\lambda}}\right]Q(\lambda,\hat{\lambda}), \quad \lambda \geq \hat{\lambda}.\end{aligned} \tag{26}$$

For $\lambda < \hat{\lambda}$, a sustained-extension strategy requires an immediate intensity jump of size $a(\lambda,\hat{\lambda}) = (\hat{\lambda}-\lambda)/\delta_2$, moving the repayment intensity from λ to $\hat{\lambda}$, followed by the continuous collection effort $E(s) = \kappa/\delta_2(\hat{\lambda}-\lambda_\infty)$ for $s \geq 0$ to hold the intensity at $\hat{\lambda}$, so that the sustained-extension operator is defined as

$$(\mathcal{S}_{\hat{\lambda}}v)(\lambda,w) \triangleq ca(\lambda,\hat{\lambda}) + \left[\bar{v}(\hat{\lambda},w) + \hat{c}\kappa\frac{\hat{\lambda}-\lambda_\infty}{\hat{\lambda}}\right]Q(\hat{\lambda},\hat{\lambda}), \quad \lambda < \hat{\lambda}. \tag{27}$$

The next result establishes invariance properties of $\mathcal{S}_{\hat{\lambda}}$.

Lemma 2. *Assume that* $\hat{\lambda} \geq \lambda_\infty$ *and let* $v\colon \mathbb{R}_+^2 \to \mathbb{R}_-$ *be a* C^1*-function that has second-order derivatives a.e. Moreover, assume that* $\partial_1 v, \partial_2 v, \partial_{12} v, \partial_{22} v \leq 0$. *Then the sustained extension of* v,

$$\hat{v}(\lambda,w;\hat{\lambda}) \triangleq (\mathcal{S}_{\hat{\lambda}}v)(\lambda,w),$$

is such that

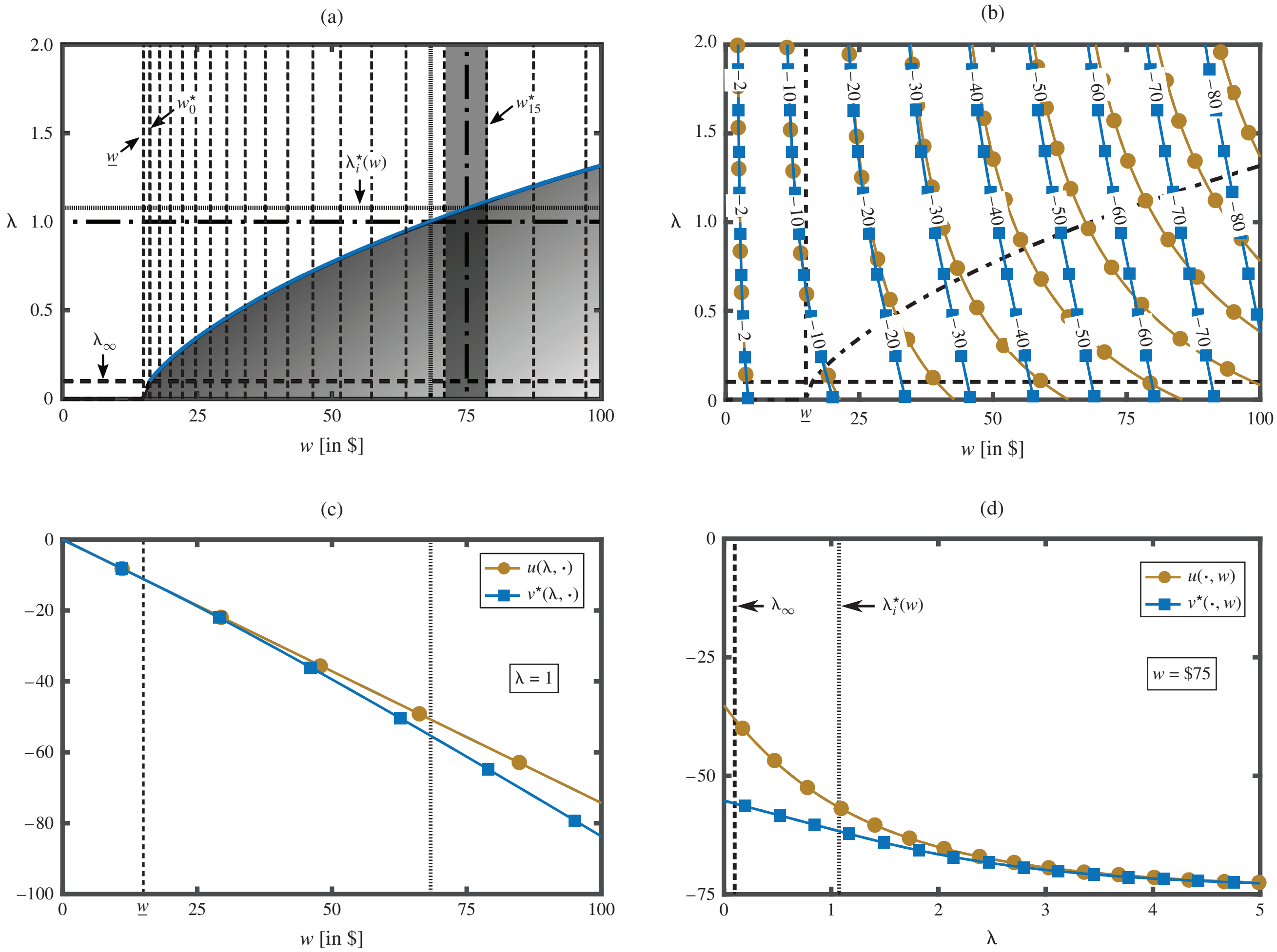
(i) $\hat{v}\colon \mathbb{R}_+^2 \to \mathbb{R}_-$ *remains* C^1 *and is an a.e. twice differentiable function such that* $\partial_1\hat{v}, \partial_2\hat{v}, \partial_{12}\hat{v}, \partial_{22}\hat{v} \leq 0$;

(ii) $\hat{v}$ *is continuously differentiable with respect to* $\hat{\lambda}$. *Moreover,* $\partial_2\hat{v}$ *is decreasing in* $\hat{\lambda}$.

The sustained extension $\mathcal{S}_{\hat{\lambda}}v$ preserves most first- and second-order monotonicity properties of v. Note that, while for an arbitrary level $\hat{\lambda}$ the function $\hat{v}(\cdot;\hat{\lambda})$ is generically not convex in λ, as is shown in the next subsection, this second-order monotonicity is preserved at "the optimal level" when extending an (optimal) value function of the collection problem to a larger domain of account states. The intensity dynamics induced by the sustained-extension operator resemble the effects of an account-treatment action. As noted in Section 3.1, each account-treatment action can sustain a certain intensity level $\hat{\lambda}$. Implementing an account-treatment action increases the intensity to $\hat{\lambda}$, provided the current intensity is smaller, and it maintains the intensity at $\hat{\lambda}$ as long as the action remains in effect.

3.2.4. Optimal Account Value. The key idea for constructing the value function $v^*(\lambda,w)$ of the collection problem for *all* $(\lambda,w) \geq 0$ consists in recursively extending a known value function, valid for small balances, to successively include larger balances. As shown in

Figure 3. (a) Solution to the Collection Problem; (b)–(d) Comparison of Optimal and Autonomous Value



Section 3.2.2, the value function of the collection problem (P) is the terminal value function $v_0^*(\lambda, w)$ in Equation (20) for all $w \in [0, w_0^*]$. For any given outstanding balance $w > w_0^*$, the recursive construction of the value function v^* on $\mathbb{R}_+ \times [0, w]$ can be completed in a finite number of iterations, provided that repayments cannot become arbitrarily small.

Minimal-repayment assumption (MRA). *There exists a positive minimal (relative) repayment* $\underline{r} \in (0, 1)$ *such that the support of* F_R *is contained in* $[\underline{r}, 1]$.

In practice, MRA is satisfied because there is a natural lower bound $\underline{z}$ imposed by any discrete currency on nonzero absolute repayment amounts (e.g., $\underline{z} = \$0.01$), and one can set $\underline{r} = \underline{z}/w$.[14] For a given minimal relative repayment $\underline{r}$ and $w > w_0^*$, the integer

$$i(w) \triangleq \left\lceil \frac{\ln(w_0^*/w)}{\ln(1 - \underline{r})} \right\rceil \geq 1 \tag{28}$$

is an upper bound for the number of repayments until a positive outstanding balance is repaid up to the *economic balance threshold* w_0^*, below which it makes no economic sense to actively pursue collection. For any $i \geq 1$, let $w_i^* \triangleq w_0^*(1 - \underline{r})^{-i}$ denote the maximum outstanding balance that may require at most i repayments for the outstanding amount to drop to or below w_0^*, so $i(w) \leq i$ if and only if $w \leq w_i^*$; see Figure 3(a). Moreover, assume that the value function v_{i-1}^* of the collection problem (P) for account states in $\mathbb{R}_+ \times [0, w_{i-1}^*]$ is obtained and has the following characteristic properties.

Value properties. For a given $i \geq 1$, $v_{i-1}^*\colon \mathbb{R}_+ \times [0, w_{i-1}^*] \to \mathbb{R}$ is C^1 and has second-order derivatives a.e. In addition, v_{i-1}^* satisfies the following three "value properties":

(P1) $0 \geq v_{i-1}^*(\lambda, w) \geq -w = \lim_{\lambda\to\infty} v_{i-1}^*(\lambda, w)$;
(P2) $-\hat{c} \leq \partial_1 v_{i-1}^*(\lambda, w) \leq 0 = \lim_{\lambda\to\infty} \partial_1 v_{i-1}^*(\lambda, w)$;
(P3) $\partial_2 v_{i-1}^*(\lambda, w), \partial_{22} v_{i-1}^*(\lambda, w), \partial_{12} v_{i-1}^*(\lambda, w) \leq 0 \leq \partial_{11} v_{i-1}^*(\lambda, w)$.

Applying the sustained-extension operator $\mathcal{S}_{\hat{\lambda}}$ on v_{i-1}^* extends the definition of v_{i-1}^*. Let $\hat{\lambda} \geq \lambda_\infty$ and for any $(\lambda, w) \in \mathbb{R}_+ \times [w_{i-1}^*, w_i^*]$ set

$$v_i(\lambda, w; \hat{\lambda}) \triangleq (\mathcal{S}_{\hat{\lambda}} v_{i-1}^*)(\lambda, w). \tag{29}$$

The intensity level $\hat{\lambda}$, at which the sustained extension v_i is being held before the next repayment is subject to optimization, leading to the *i-th collection-continuation problem*,

$$\min_{\hat{\lambda}\geq\lambda_\infty} v_i(\lambda, w; \hat{\lambda}), \quad (\lambda, w) \in \mathbb{R}_+ \times [w^*_{i-1}, w^*_i]. \qquad (\text{CCP}_i)$$

By Equation (27), it is $v_i(\lambda, w; \hat{\lambda}) > v_i(\lambda, w; \lambda)$ for $\hat{\lambda}$ large enough so that the solution of (CCP$_i$) must be finite and must, therefore, also exist by the Weierstrass theorem (Bertsekas 1995, p. 540).

Moreover, since the first-order necessary optimality condition of the i-th collection-continuation problem (CCP$_i$), obtained by setting the derivative of $v_i(\lambda, w; \hat{\lambda})$ with respect to $\hat{\lambda}$ to zero,

$$f_i(\hat{\lambda}, w) \triangleq \hat{c}\frac{(\hat{\lambda}+\rho)^2}{\kappa(\rho+\lambda_\infty)} + \hat{c} + \frac{\rho}{\kappa(\rho+\lambda_\infty)}\bar{v}^*_{i-1}(\hat{\lambda}, w) + \frac{\hat{\lambda}(\hat{\lambda}+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_1\bar{v}^*_{i-1}(\hat{\lambda}, w) = 0, \qquad (30)$$

is independent of λ, it can be concluded that the optimal $\hat{\lambda}$ depends only on w.[15] Denoting by $\lambda^*_i(w)$ the $\hat{\lambda} \geq \lambda_\infty$ that solves the above optimality condition for $w \in [w^*_{i-1}, w^*_i]$ the following result establishes existence and uniqueness as well as monotonicity properties.

Theorem 4. *The intensity level $\lambda^*_i(w)$ exists as a unique solution to Equation (30); it is increasing in w, and $\lambda^*_{i-1}(w^*_{i-1}) = \lambda^*_i(w^*_{i-1})$ for any $i \geq 1$.*

Recall that the solution to the terminal collection problem (TCP) in Section 3.2.2 is the value function of the collection problem (P) for $w \in [0, w^*_0]$. The construction of the value function can now proceed by recursion for $i \geq 1$. Specifically, for any $(\lambda, w) \in \mathbb{R}_+ \times [0, w^*_i]$, let

$$v^*_i(\lambda, w) = \begin{cases} v_i(\lambda, w; \lambda^*_i(w)), & \text{if } w \in (w^*_{i-1}, w^*_i], \\ v^*_{i-1}(\lambda, w), & \text{otherwise.} \end{cases} \qquad (31)$$

Building on Lemma 2(i), setting the holding intensity $\hat{\lambda}$ equal to $\lambda^*_i(w)$ yields the value properties.

Lemma 3. *The function v^*_i satisfies the value properties (P1)–(P3) on $\mathbb{R}_+ \times [0, w^*_i]$.*

It follows that v^*_i is indeed the value function of the collection problem (P) as long as the account balance does not exceed w^*_i.

Theorem 5. *For any $(\lambda, w) \in \mathbb{R}_+ \times [0, w^*_i]$, $i \geq 0$, the value function $v^*_i(\lambda, w)$ is the value function of the collection problem (P).*

Analogous to the value function, the state-space partition into continuation, holding, and action regions can also be extended so as to include successively larger balances. For this, consider the continuation, holding, and action regions $\mathcal{C}^*_0$, $\mathcal{A}^*_0$, $\mathcal{H}^*_0$ defined in Corollary 4, and for $i \geq 1$ set

$$\mathcal{C}^*_i \triangleq (\lambda^*_i(w), \infty) \times (w^*_{i-1}, w^*_i], \quad \mathcal{H}^*_i \triangleq \{\lambda^*_i(w)\} \times (w^*_{i-1}, w^*_i],$$
$$\mathcal{A}^*_i \triangleq [0, \lambda^*_i(w)) \times (w^*_{i-1}, w^*_i].$$

Then the unique optimal collection strategy follows by iteration from Theorem 5.

Corollary 5. *For any $(\lambda, w) \in \mathbb{R}_+ \times (w^*_{i-1}, w^*_i]$, define $a^*_i(\lambda, w) \triangleq \max\{a(\lambda, \lambda^*_i(w)), 0\}$. An optimal collection strategy for an account with $(\lambda(0), W(0)) = (\lambda, w) \geq 0$ is*

$$A^*(t) = a^*(\lambda, w) + \int_{[0,t]} E^*(s)\, ds, \quad t > 0, \qquad (32)$$

with $A^*(0) = 0$. For $s \geq 0$, the optimal discrete collection effort is $a^*(\lambda, w) = a^*_{i(w)}(\lambda, w)$ while the optimal continuous collection effort is given by

$$E^*(s) = \sum_{k=1}^{i(w)} \mathbf{1}_{\{(\lambda(s), W(s))\in\mathcal{H}^*_k\}}(\kappa/\delta_2)(\lambda(s) - \lambda_\infty).$$

The optimal collection strategy in Equation (32) is admissible (as an element of $\mathscr{A}$) and solves the collection problem (P) for any given account state $(\lambda, w) \geq 0$. It includes a single discrete collection effort a^* induced by the first optimal account-treatment action at $t = 0$ to move the repayment intensity instantaneously to the action-region boundary. The optimal collection strategy also includes a continuous collection effort E^* induced by the first and subsequent optimal account-treatment actions so as to hold the repayment intensity at an optimal level when the waiting time between repayment events becomes too large.[16]

4. Implementation

The empirical identification of the repayment process is discussed by Chehrazi and Weber (2015) using both maximum-likelihood estimation and the generalized method of moments. Here we introduce optimization and compare the performance of optimized collections with the yield from untreated accounts. For this, we consider the case where the relative repayment distribution is uniform on $[\underline{r}, 1]$ with a minimal relative repayment of $\underline{r} = 0.1 > 0$, so MRA is satisfied.[17] In this implementation, a time period represents a quarter (three months), the mean reversion constant κ is 0.7, and the long-run steady state is $\lambda_\infty = 0.1$. Therefore, in the absence of account-treatment actions, the repayment intensity of an untreated account without repayments reverts by about 50% ($\approx e^{-0.7}$) each quarter toward λ_∞. The sensitivity of the repayment process with respect to willingness to repay is $\delta_{10} = 0.02$ and with respect to ability to repay is $\delta_{11} = 0.5$. The sensitivity of the repayment process with respect to the collection effort is $\delta_2 = 1$, effectively normalizing the magnitude of effort to be commensurable with the repayment intensity. The bank's quarterly effective discount rate is $\rho = 6\%$, and its marginal cost of effort is $c = \$6$ per additional unit of repayment intensity.

4.1. Account Value and Collection Strategy

Consider an account with outstanding balance $w = \$75$ and repayment intensity $\lambda = \lambda_\infty$ at the beginning of the collection process (i.e., at $t = 0$). Figure 3(a) depicts the partition of the state space that characterizes the optimal collection strategy: the shaded area indicates the action region $\mathcal{A}^*$; the portion of its boundary that lies above λ_∞ represents the holding region $\mathcal{H}^*$; the complement of $\mathcal{A}^* \cup \mathcal{H}^*$ is the continuation region $\mathcal{C}^*$; see Equations (9)–(11). As discussed in Section 3.2.4, the value function $v^*(\lambda, w)$ of the collection problem (P) can be obtained in $i(w) = 15$ iterations, where, starting with the terminal collection problem for account balances in $[0, w_0^*]$ (with $w_0^* = \$16.30$), at each step $i \in \{1, \ldots, i(w)\}$ an exact solution on $[0, w_{i-1}^*]$ is extended to $[0, w_i^*]$ (with $w_i^* = w_0^*/(1 - \underline{r})^i$) until $w \leq w_i^*$. Figure 3(b) compares the optimal account value $v^*(\lambda, w)$ for the optimally treated account with the autonomous account value $u(\lambda, w)$ for the untreated account given any account state $(\lambda, w) \in [0, 2] \times [0, \$100]$.

Figure 3(c) shows the optimal and autonomous account values for $\lambda = 1$, as a function of w. The first vertical line indicates the minimal actionable balance $\underline{w} = \$15$, below which the two value functions coincide. The second vertical line marks the outstanding balance at the optimal holding level $\lambda^*(w) = 1$; for larger levels of w, the account state $(1, w)$ is in the action region $\mathcal{A}^*$. As shown in Section 3, the value function $v^*(\lambda, w)$ is concave in w (i.e., it satisfies the value property (P3)). Moreover, as noted in Theorem 2, $u(\lambda, w)$ is linear in w. Figure 3(d) depicts the optimal and autonomous account values at $w = \$75$ as a function of the repayment intensity λ. The vertical lines indicate the long-run steady state λ_∞ and the optimal holding intensity $\lambda^*(\$75)$. Note that for $\lambda \leq \lambda^*(w)$, $v^*(\lambda, w)$ is affine in λ while for $\lambda > \lambda^*(w)$ it becomes convex in λ. By virtue of Theorem 2 the autonomous account value is strictly convex in λ. As the initial intensity grows beyond all bounds, the full outstanding balance can be recovered essentially without account treatment as $\lim_{\hat{\lambda}\to\infty} v^*(\hat{\lambda}, w) = \lim_{\hat{\lambda}\to\infty} u(\hat{\lambda}, w) = -w$, consistent with value property (P1).

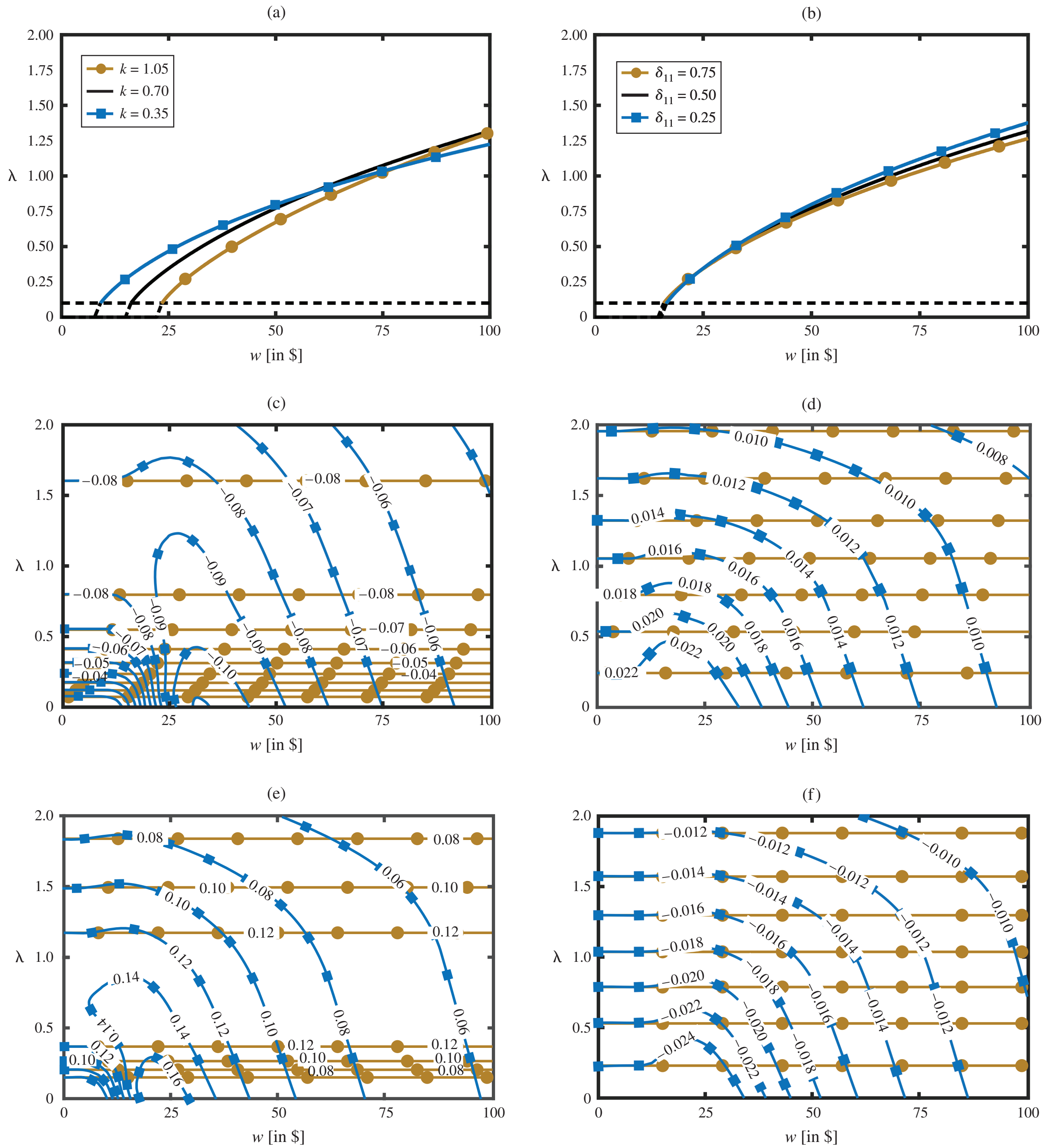
Figure 2 shows the particular realization of an account-state trajectory, starting at $(W(0), \lambda(0)) = (\$75, 0.1)$.[18] As the initial account state is an element of $\mathcal{A}^*$, an immediate discrete collection effort introduces an intensity jump, so that, at time $t = 0^+$, the account-state trajectory continues from $(\$75, \lambda(0^+))$ in the holding region; see Figure 2(a). Subsequently, the continuous collection effort $E^*(t)$ is applied (as obtained in Corollary 5) to prevent the account state from drifting back into the action region until a repayment is received. The impulse control (at $t = 0$), together with the subsequent continuous control, captures the effect of the first optimal account-treatment action. This action is chosen such that the intensity level it attains (via jump) and maintains (via continuous control) coincides with the optimal level $\lambda^*(\$75)$ (or, in practice, is close to $\lambda^*(\$75)$). Note that subsequent optimal account-treatment actions do not induce any jump as they are taken at the optimal time, that is, before the repayment intensity enters the action region. Figures 2(a)–2(d) show the evolution of the account's repayment intensity, the outstanding balance, and the optimal collection effort as a function of time. As noted before, in the absence of account treatment, the intensity of the repayments reverts exponentially toward the long-run steady state $\lambda_\infty = 0.1$. This reversion tendency is indicated in Figure 2(b) by a family of decreasing dashed trajectories, which are followed when there are no intensity jumps. The horizontal dashed lines correspond to the optimal holding levels, which correspond to the values of $\lambda^*(w)$ in Figure 2(a) for the different balances attained by the realization of the jump process $W(t)$.

Receiving a repayment of relative magnitude R_i increases the repayment intensity by $\delta_{10} + \delta_{11} R_i$ in accordance with the holder's estimated willingness and ability to repay, and it decreases the remaining outstanding balance by $W(t) R_i$. This introduces a jump into the account state that moves it away from the action and holding regions. For the sample path shown in this picture, after four separate repayments (Figure 2(c)) the account balance drops below the economic balance threshold $w_0^* = \$16.30$. Before the fourth repayment, the repayment intensity declines to a level that requires a second collection action by the bank. The optimal effort required at this point is smaller than for the first collection action primarily because the outstanding balance is considerably smaller; see Figure 2(d). Yet the second action lasts until the fourth payment is received. As indicated by our model, it cannot be optimal to terminate an account-treatment action before a repayment is received. From a practical point of view, such suboptimal termination is destined to have a negative impact on the credibility of future actions. In this example, the account makes a fifth repayment in the autonomous (treatment-free) regime below the economic balance threshold w_0^*. Note that because the long-run steady state λ_∞ is strictly positive, all accounts will eventually pay back in full if the bank is willing to wait indefinitely. However, because of the time value of money ($\rho > 0$), the expected economic account value $\bar{V}^*(\lambda, w)$ is strictly less than the outstanding balance.

4.2. Model Mis-specification and Robustness Analysis

The parameter vector $(\kappa, \delta_{10}, \delta_{11}, \delta_2)$ determines an account's repayment behavior. It is critical to understand how each parameter affects the optimal collection strategy as well as the optimal account value. The effect of perturbing δ_2 is fairly clear: a larger value for

Figure 4. Sensitivity of the Boundary of the Optimal Action Region, the Autonomous Account Value, and the Optimized Account Value with Respect to (a), (c), and (e) ±50% Perturbations in κ; (b), (d), and (f) ±50% Perturbations in δ_{11}



δ_2 implies an account holder who is more responsive to collection actions, thus increasing the size of the optimal action region and the optimal account value. We focus below on the interesting effects of changing κ and δ_{11}. The effect of δ_{10} on the optimal action region and the optimal account value is similar to δ_{11}.

Regarding perturbations of κ, Figure 4(a) shows how the boundary of the optimal action region changes in response to a ±50% variation. A larger value of κ increases the rate at which the repayment intensity declines to λ_∞. Consequently, in comparison with the base case, it is costlier to maintain any intensity level.

The impact of this change on the boundary of the optimal action region depends on the magnitude of the outstanding balance. For large balances, this translates to earlier and more forceful account treatment: larger expected rewards justify higher collection cost. For small balances, however, treatment actions are delayed and less forceful: the size of expected repayments may not justify an immediate additional collection outlay so that treatment actions are delayed until absolutely necessary. Converse effects obtained when decreasing κ since a slower reversion of the intensity to its long-run steady state implies a longer lasting effect of the treatment action. Thus, for high balances it is optimal to delay treatment activities since, if the account state is in the inaction region, (i) the intensity is already large enough and (ii) it declines sufficiently slowly for a repayment to arrive. In the unlikely event that a repayment is not made, a weaker treatment action proves optimal since the intensity declines at a slower pace, so the impact of an action lasts longer. For small balances, it is optimal to expedite an action primarily because it is cheaper to implement and also because its effects last longer. Figure 4(c) and 4(e) illustrates the relative change in the value function for both optimally treated and untreated (autonomous) accounts when κ changed by 50% either way. As in Figure 3(b)–3(d), the blue lines (marked by squares) show the contours for the relative change of the optimal account value whereas the brown lines (marked by circles) represent the contours of the relative change in the autonomous account value. Since for low balances there is no treatment, the contour levels coincide for small w. A higher κ decreases the account value, so the change becomes negative (Figure 4(c)). Conversely, a smaller κ increases the account value, resulting in a positive change (Figure 4(e)). Disregarding the signs, the relative change for the optimized account value slightly exceeds that for the autonomous account value; however, both are significantly below 50%.

The impact of δ_{11} on the size of the optimal action region and on the optimal account value is monotone. A larger δ_{11} produces a larger intensity jump when a unit repayment arrives, thus increasing also the likelihood of further repayment events, therefore leading to a higher account value. In addition, an increase in future repayment likelihood reduces the need for collection actions, thus shrinking the size of the optimal action region (Figure 4(b)). Figure 4(d) and 4(f) illustrates the contour levels of the relative change in the autonomous and optimal account value caused by +50% and −50% change in δ_{11}, respectively. Similar to Figure 4(c) and 4(e), the absolute values of these changes remain significantly below 50%.

As mentioned in Section 2.1, the intensity process $\lambda(t)$ can be viewed as the best Bayesian estimate of the actual (unobservable) repayment intensity when the bank dynamically updates its belief according to the observed repayment history; see Appendix C for details. It is, therefore, instructive to examine how model mis-specification may impact account value and collection cost; see Table 1.[19] In particular, we assume that an account's true parameters are $(\kappa, \delta_{10}, \delta_{11}, \delta_2) = (0.7, 0.2, 0.5, 1)$ and examine the account value, collection cost, and relative error when, instead of the optimal collection strategy A^* for the true account parameters, the collection strategy A for mis-specified account parameters is used. For example, when the bank estimates κ with +50% error, that is, incorrectly assumes $\kappa = 1.05$ instead of $\kappa = 0.7$, the account value decreases from its maximum \$55.69, obtained from the optimal collection strategy A^*, to \$55.58, obtained by the suboptimal collection strategy A, and the cost of collection increases from \$10.69 to \$11.42; cf. first column of Table 1.

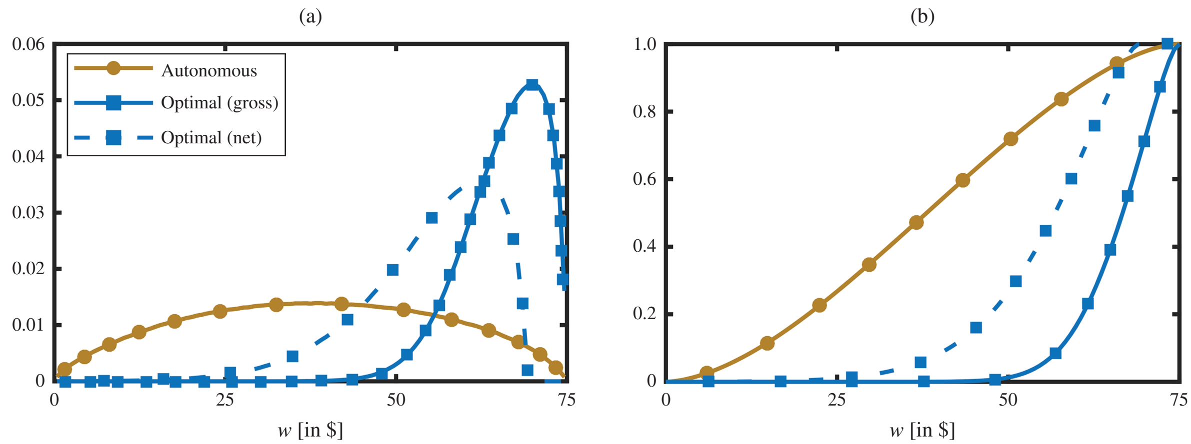
Despite a significant model mis-specification error, the relative error in the optimal account value remains usually small. There are two main reasons for this inherent robustness. First, as noted earlier, the action/inaction boundary is not sensitive to changes in the values of account parameters. Second, the error in the action/inaction boundary affects the outcome only when the account state hits this boundary. As can be observed in Figure 2, after an account is set on its recovery path, the account state tends to remain in the inaction region, limiting its exposure to the consequences of mis-specification. The impact of model mis-specification on the collection cost is more difficult to predict. For example, one may expect that

Table 1. Robustness with Respect to Model Misclassification

	κ		δ_{10}		δ_{11}		δ_2	
	+50%	−50%	+50%	−50%	+50%	−50%	+50%	−50%
$-v(0.1,\$75;A)$ (\$)	55.58	55.16	55.69	55.69	55.68	55.68	55.63	55.52
Cost of collection strategy A (\$)	11.42	8.84	10.69	10.69	10.39	11.01	10.46	10.00
$\left\lvert\frac{v(0.1,\$75;A)-v(0.1,\$75;A^*)}{v(0.1,\$75;A^*)}\right\rvert$ (%)	0.19	0.95	0[†]	0[†]	0.01	0.01	0.11	0.31

Note. The optimal account value $-v(0.1,\$75;A^*)$ and cost of optimal collection strategy A^* are \$55.69 and \$10.69, respectively.
[†]At $\delta_{10} = 0.02$, a relative change of magnitude ±50% does not amount to a significant change in the value function.

Figure 5. Repayment Distributions: (a) pdf; (b) cdf



overestimating κ reduces the cost of collection as this shrinks the action region; see Figure 4(a). However, somewhat surprisingly the collection cost increases; see Table 1. Under mis-specified dynamics, the intensity hits the action region more frequently, so the bank unnecessarily takes more account-treatment actions, thus driving up collection cost in expectation.

4.3. Collectability Improvement

In terms of the present value of an account, any nontrivial account treatment produces a first-order stochastically dominant shift in the revenue distribution over any given time horizon. The solid lines in Figures 5(a) and 5(b) show the (gross) present-value distributions for untreated (autonomous) and optimally treated accounts, respectively.[20] For our example, the expected present value of the repayment increases from \$38.01 to \$66.38, corresponding to an almost 75% increase from 50.7% to 88.5% of the outstanding balance. Moreover, the coefficient of variation of gross collections decreases significantly from 46.9% to 8.7%. This illustrates that any nontrivial treatment strictly improves the asset quality of nonperforming loans by both increasing returns and decreasing risk. The dashed lines in Figure 5(a) and 5(b) indicate the distribution of the *net* present account value, that is, the optimal value of the account, including the collection costs. For any account with an initial outstanding balance above the economic balance threshold w_0^*, placement in collections translates to expending immediate positive effort to bring the account's repayment intensity from $\lambda(0) = \lambda_\infty$ to $\lambda(0^+) > \lambda_\infty$. This leads to a guaranteed lower bound for the overall collection cost, which includes investments related to establishing and continuing collections. As a result, a full net recovery of the outstanding amount is impossible for an optimally treated account (unless it starts with a repayment intensity outside $\mathcal{A}^*$). Hence, the autonomous repayment distribution cannot be first-order stochastically dominated by the distribution of the net present account value.

5. Conclusion

Based on the results presented in Section 3, a bank's optimal collection strategy maps any possible state of the account, expressed as a point (λ, w) in the (intensity, balance) space to an optimal action. The optimal account treatment aims at maintaining the repayment intensity at a minimum level $\lambda^*(w)$ until a repayment is received and the outstanding balance eventually falls below the economic balance threshold w_0^*, at which point it is best to suspend an active pursuit of the account. The optimal holding intensity $\lambda^*(w)$ is increasing in the outstanding balance, leading to stronger (i.e., more aggressive) collection actions for larger outstanding balances w. The economic balance threshold increases in the marginal *effective* collection cost $\hat{c}$ as the quotient of the marginal collection cost c and the collection effectiveness (i.e., sensitivity of the repayment intensity with respect to actions) δ_2. These critical performance indicators allow banks to sort collection agencies and, thus, to devise an optimal agency-assignment policy. Because of the increase in asset quality through collections, alluded to in Section 4, an optimal collection strategy has the potential to significantly reduce the bank's loss given default and, thus, also to lessen its required capital reserves, not only because of the decrease in outstanding balances, but because the pure risk in the collection of an outstanding balance is lowered by the bank's implementation of the optimal account-treatment strategy or an approximation thereof (see Appendix D); an in-depth treatment of approximately optimal collection strategies taking explicitly into account a bank's limited number of treatment measures is left as an interesting topic for future research.

Beyond the specific practical implications for a bank's optimal debt-collection strategy, the findings have implications for the control of (affine) *self-exciting* point processes. In this context, the construction of a semi-analytical solution for the value function together with a complete characterization of an optimal policy is new. Earlier results, including those by Costa and Davis (1989), focus on the successive approximation of solutions. A monotonicity in the state evolution (balance decreasing with each repayment event) is exploited to obtain a recursive extension of the domain of an exact value function, so as to include any finite initial state in a finite number of iterations. The proposed approach has the added value of providing a complete description of first- and second-order monotonicity as well as limiting behavior in the form of "value properties"; see Section 3.2.4. This, in turn, allows for important economic insights, such as, in our application, about the (dis-)economies of scale in collections as a function of the size of the outstanding consumer loan and the bank's account-treatment effort.

Appendix A. Proofs

The following auxiliary result is used in the proof of Theorem 1.

Proposition A.1. *Fix $(\lambda, w) \geq 0$, and let $v^*: \mathbb{R}_+ \times [0, w] \to \mathbb{R}$ be a bounded function. For any $t \geq 0$ and admissible strategy $A \in \mathcal{A}$, let*

$$H(t;A) \triangleq \int_{[0,t]} e^{-\rho s}\, dW(s) + c\int_{[0,t)} e^{-\rho s}\, dA(s) + e^{-\rho t} v^*(\lambda(t), W(t)),$$

where $\lambda(t)$ is the solution of Equation (4) for the initial condition $\lambda(0) = \lambda$, and $W(t) = w - \sum_i Z_i \mathbf{1}_{\{T_i \leq t\}}$. If v^* is such that (i) *$H(t;A)$ is a submartingale for any admissible collection strategy $A \in \mathcal{A}$ and* (ii) *there exists an admissible $A^* \in \mathcal{A}$ for which $H(t;A^*)$ is a martingale, then*

$$v^*(\lambda, w) = \inf_{A\in\mathcal{A}} \mathbb{E}\left[\int_{[0,\infty)} e^{-\rho s}\, dW(s) + c\int_{[0,\infty)} e^{-\rho s}\, dA^*(s) \,\middle|\, \lambda(0) = \lambda,\ W(0) = w\right],$$

and the collection strategy A^* attains the infimum.

Proof. Note first that at $t = 0$, it is $H(0;A) = v^*(\lambda, w)$. The submartingale property of $H(t;A)$ implies that $v^*(\lambda, w) = H(0;A) \leq \mathbb{E}[H(t;A) \mid \lambda(0) = \lambda, W(0) = w]$ for any $t \geq 0$. Since $\int_{[0,t]} e^{-\rho s}\, dW(s)$ is bounded and decreasing in t, $c\int_{[0,t)} e^{-\rho s}\, dA(s)$ is an increasing function of t, and v^* is bounded, using the definition of the expectation operator and the monotone-convergence theorem, one obtains

$$\lim_{t\to\infty} \mathbb{E}[H(t;A) \mid \lambda(0) = \lambda, W(0) = w] = \mathbb{E}\left[\lim_{t\to\infty} H(t;A) \,\middle|\, \lambda(0) = \lambda, W(0) = w\right].$$

Consequently,

$$v^*(\lambda, w) \leq \inf_{A\in\mathcal{A}} \mathbb{E}\left[\int_{[0,\infty)} e^{-\rho s}\, dW(s) + c\int_{[0,\infty)} e^{-\rho s}\, dA(s) \,\middle|\, \lambda(0) = \lambda, W(0) = w\right]. \quad \text{(A.1)}$$

By the martingale property the infimum is attained for the admissible collection strategy $A^* \in \mathcal{A}$, which completes the proof. □

Proof of Theorem 1. It is enough to show that v^* satisfies the conditions of Proposition A.1, which implies that it is the value function of the collection problem (P). Note that the boundary condition, $\lim_{\lambda\to\infty} v^*(\lambda, w) = -w$, yields that v^* is bounded on $\mathbb{R}_+ \times [0, w]$. To simplify the rest of the proof, we denote by

$$\nu^*(t) \triangleq v^*(\lambda(t), W(t)), \quad t \geq 0,$$

the value along a state trajectory, and by

$$\tau_k \triangleq \inf\{t > \tau_{k-1}: \lambda(t^+) \neq \lambda(t^-)\}, \quad k \geq 1,$$

with $\tau_0 \triangleq 0$, the (discrete) collection-event times. The latter leads to intensity jumps, be it via repayments (i.e., jumps in J) or via discrete collection efforts (i.e., jumps in A). Furthermore, let

$$\eta(t) \triangleq \sum_{k=0}^{\infty} \mathbf{1}_{\{\tau_k \leq t\}} \mathbf{1}_{\{\mathcal{N}(\tau_k^-) \neq \mathcal{N}(\tau_k)\}} + \sum_{k=0}^{\infty} \mathbf{1}_{\{\tau_k < t\}} \mathbf{1}_{\{\mathcal{N}(\tau_k^-) = \mathcal{N}(\tau_k)\}}, \quad t \geq 0,$$

be the number of (discrete) collection events, including repayments up to and including t and discrete collection efforts up to but not including t.[21] With the expected relative repayment $\bar{r} = \int_0^1 r\, dF_R(r)$ as in Equation (2), it is

$$\begin{aligned} H(t;A) - H(0;A) &= \int_{[0,t]} e^{-\rho s}[dW(s) + \bar{r}W(s^-)\lambda(s^-)ds] - \int_{[0,t]} e^{-\rho s}\bar{r}W(s^-)\lambda(s^-)\, ds \\ &+ c\int_{[0,t]} e^{-\rho s}E(s)\, ds + c\sum_{k:\,\tau_k < t} e^{-\rho\tau_k}\Delta A_k \\ &+ \sum_{k:\,\tau_k < t} e^{-\rho\tau_k}[\nu^*(\tau_k^+) - \nu^*(\tau_k)] + \sum_{k:\,\tau_k \leq t} e^{-\rho\tau_k}[\nu^*(\tau_k) - \nu^*(\tau_k^-)] \\ &+ \sum_{k=1}^{\eta(t)} [e^{-\rho\tau_k}\nu^*(\tau_k^-) - e^{-\rho\tau_{k-1}}\nu^*(\tau_{k-1}^+)] \\ &+ e^{-\rho t}\nu^*(t) - e^{-\rho\tau_{\eta(t)}}\nu^*(\tau_{\eta(t)}^+), \end{aligned} \quad \text{(A.2)}$$

where the first term on the right-hand side is a martingale and the last four terms decompose $e^{-\rho t}\nu^*(t) - \nu^*(0)$ into sums of increments at and between the jumps of $(\lambda(t), W(t))$, respectively. Since ν^* is continuous between any two consecutive jumps (and after the last jump), one obtains

$$e^{-\rho\tau_k}\nu^*(\tau_k^-) - e^{-\rho\tau_{k-1}}\nu^*(\tau_{k-1}^+) = \int_{\tau_{k-1}^+}^{\tau_k^-} e^{-\rho s}\left[\frac{d}{ds}\nu^*(s) - \rho\nu^*(s)\right] ds, \quad 1 \leq k \leq \eta(t),$$

and

$$e^{-\rho t}\nu^*(t) - e^{-\rho\tau_{\eta(t)}}\nu^*(\tau_{\eta(t)}^+) = \int_{\tau_{\eta(t)}^+}^{t} e^{-\rho s}\left[\frac{d}{ds}\nu^*(s) - \rho\nu^*(s)\right] ds, \quad t \geq 0.$$

Note that on each interval $(\tau_{k-1}, \min\{\tau_k, t\})$, $1 \le k \le \eta(t)+1$, the outstanding balance $W(t)$ and the discrete portion $A(t) - \int_{[0,t]} E(s)\,ds$ of the action process $A(t)$ are constant, so by Equation (4):

$$\frac{dv^*(s)}{ds} = \partial_1 v^*(\lambda(s), W(s))[\kappa(\lambda_\infty - \lambda(s)) + \delta_2 E(s)], \quad s \in (\tau_{k-1}, \min\{\tau_k, t\}).$$

Hence, the last two terms in Equation (A.2) simplifies to

$$\sum_{k=1}^{\eta(t)} [e^{-\rho\tau_k} v^*(\tau_k^-) - e^{-\rho\tau_{k-1}} v^*(\tau_{k-1}^+)] + e^{-\rho t} v^*(t) - e^{-\rho\tau_{\eta(t)}} v^*(\tau_{\eta(t)}^+)$$
$$= \int_0^t e^{-\rho s}[[\kappa(\lambda_\infty - \lambda(s)) + \delta_2 E(s)]\partial_1 v^*(\lambda(s), W(s)) - \rho v^*(\lambda(s), W(s))]\,ds, \quad t \ge 0.$$

For the repayment events, one obtains

$$\sum_{k:\tau_k \le t} e^{-\rho\tau_k}[v^*(\tau_k) - v^*(\tau_k^-)]$$
$$= \int_{[0,t]} e^{-\rho s}[v^*(\lambda(s), W(s)) - v^*(\lambda(s^-), W(s^-))]\,d\mathcal{N}(s)$$
$$= \int_{[0,t]\times[0,1]} e^{-\rho s}[v^*(\lambda(s^-) + \delta_{10} + \delta_{11} r, W(s^-)(1-r)) - v^*(\lambda(s^-), W(s^-))]\hat{\mathcal{N}}(ds, dr)$$
$$= \int_{[0,t]\times[0,1]} e^{-\rho s}[v^*(\lambda(s^-) + \delta_{10} + \delta_{11} r, W(s^-)(1-r)) - v^*(\lambda(s^-), W(s^-))](\hat{\mathcal{N}}(ds, dr) - \lambda(s)\,ds\,dF_R(r))$$
$$+ \int_{[0,t]\times[0,1]} e^{-\rho s}[v^*(\lambda(s^-) + \delta_{10} + \delta_{11} r, W(s^-)(1-r)) - v^*(\lambda(s^-), W(s^-))]\lambda(s)\,ds\,dF_R(r),$$

where for any Borel-subset $\mathscr{B} \in \mathcal{B}([0,1])$ the process $\hat{\mathcal{N}}(t, \mathscr{B})$ counts, on $[0,t]$, the number of all repayment events with relative repayments in $\mathscr{B}$.[22] Note that the first term on the right-hand side of the preceding equation is a martingale. Finally, at the times of discrete collection efforts it is

$$\sum_{k:\tau_k < t} e^{-\rho\tau_k}[v^*(\tau_k^+) - v^*(\tau_k)]$$
$$= \sum_{k:\tau_k < t, \Delta A_k > 0} e^{-\rho\tau_k}[v^*(\lambda(\tau_k) + \delta_2 \Delta A_k, W(\tau_k)) - v^*(\lambda(\tau_k), W(\tau_k))].$$

Combining these results, Equation (A.2) can be written in the form

$$H(t;A) - H(0;A) = \int_{[0,t]} e^{-\rho s}[dW(s) + \bar{r}W(s^-)\lambda(s^-)\,ds]$$
$$+ \int_{(0,t]\times[0,1]} e^{-\rho s}[v^*(\lambda(s^-) + \delta_{10} + \delta_{11} r, W(s^-)(1-r)) - v^*(\lambda(s^-), W(s^-))](\hat{\mathcal{N}}(ds, dr) - \lambda(s)\,ds\,dF_R(r))$$
$$+ \int_{[0,t]} e^{-\rho s}(\mathcal{D}_{E(s)} v^*)(\lambda(s), W(s))\,ds$$
$$+ \sum_{k:\tau_k < t, \Delta A_k > 0} e^{-\rho\tau_k}[v^*(\lambda(\tau_k) + \delta_2 \Delta A_k, W(\tau_k)) + c\Delta A_k - v^*(\lambda(\tau_k), W(\tau_k))]. \tag{A.3}$$

Hence, since by the Bellman equation (6) for any admissible $A \in \mathscr{A}$ the last two terms are nonnegative, $H(t;A)$ is a submartingale. It remains to be shown that there exists an admissible strategy $A^* \in \mathscr{A}$ for which $H(t;A)$ is a martingale. Defining $a^*(\lambda, w)$ as in Equation (8), the Bellman equation (6) naturally partitions the state space $\mathbb{R}_+^2$ into three regions: $\mathcal{A}^*$, $\mathcal{H}^*$, and $\mathcal{C}^*$ as given in Corollary 1. By the definition of $a^*(\lambda, w)$, it immediately follows that $\mathcal{A}^* \cap (\mathcal{H}^* \cup \mathcal{C}^*) = \varnothing$. In addition, since $\min_{\hat{a}\ge 0}\{v^*(\lambda + \delta_2\hat{a}, w) + c\hat{a}\} \le v^*(\lambda, w)$, and v^* satisfies the Bellman equation, it follows that if $(\lambda, w) \in \mathcal{A}^*$, then $v^*(\lambda + \delta_2 a^*(\lambda, w), w) + ca^*(\lambda, w) = v^*(\lambda, w)$. The definition of $a^*(\lambda, w)$ also allows us to conclude that when $(\lambda, w) \notin \mathcal{A}^*$, it is $\partial_1 v^*(\lambda, w) \ge -\hat{c}$ and, hence, $\min_{\hat{a}>0}\{v^*(\lambda + \delta_2\hat{a}, w) + c\hat{a}\}$ does not have any solution in this region. Therefore, for v^* to satisfy Equation (6), $\min_{\epsilon\ge 0}\{(\mathcal{D}_\epsilon v^*)(\lambda, w)\}$ has to be well-defined and equal to zero. However, this immediately follows, since $(\mathcal{D}_\epsilon v^*)(\lambda, w)$ is an affine function of ϵ whose slope is $\partial_1 v^*(\lambda, w) + \hat{c} \ge 0$. Then by the definitions in Equations (10)–(11), $(\lambda, w) \notin \mathcal{A}^*$ is either assigned to $\mathcal{H}^*$ (if $\partial_1 v^*(\lambda, w) + \hat{c} = 0$ and $\lambda > \lambda_\infty$) or $\mathcal{C}^*$ (if $\partial_1 v^*(\lambda, w) + \hat{c} > 0$ or $\lambda \le \lambda_\infty$). Hence, $\mathcal{H}^* \cap \mathcal{C}^* = \varnothing$ and $\mathcal{A}^* \cup \mathcal{H}^* \cup \mathcal{C}^* = \mathbb{R}_+^2$. Now let $A^*(0) \triangleq 0$, and define

$$A^*(t) \triangleq a^*(\lambda, w) + \int_{[0,t]} E^*(s)\,ds, \quad t > 0, \tag{A.4}$$

where $E^*(s) = \varepsilon^*(\lambda(s), W(s))$ (see Remark 3 for the definition of $\varepsilon^*(\lambda, w)$). Note that by definition, A^* is left-continuous, nondecreasing, and adapted to $\mathbb{F}$. At $t = 0$, A^* introduces a jump of size $\delta_2 a^*(\lambda, w)$ in the intensity process if and only if $(\lambda, w) \in \mathcal{A}^*$. For any $t > 0$, by the definition of $\varepsilon^*(\lambda, w)$, it is guaranteed that $(\lambda(t), W(t)) \notin \mathcal{A}^*$ over which $(\mathcal{D}_{E^*(t)} v^*)(\lambda(t), W(t)) = 0$. Consequently, for strategy A^*, the last two terms of Equation (A.3) are zero and $H(t;A^*)$ is a martingale. This concludes the proof. □

Proof of Corollary 1. See the proof of Theorem 1. □

Proof of Theorem 2. The arguments in the proofs of Proposition A.1 and Theorem 1 imply the result, provided that $(\mathcal{D}_0 u)(\lambda, w) = 0$. The latter can be verified by direct calculation as follows:

$$(\mathcal{D}_0 u)(\lambda, w)$$
$$= [\kappa(\lambda_\infty - \lambda)]\partial_1 u(\lambda, w) - \rho u(\lambda, w) + \lambda\mathbb{E}[u(\lambda + \delta_{10} + \delta_{11}R, w(1-R)) - u(\lambda, w) - wR]$$
$$= -[\kappa(\lambda_\infty - \lambda)]\rho w \int_0^\infty \alpha(t)\exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\,dt$$
$$+ \rho w - \rho^2 w \int_0^\infty \exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\,dt$$
$$+ \lambda\mathbb{E}\Big[-w(1-R) + \rho w(1-R) \cdot \int_0^\infty \exp(-\rho t - (\lambda + \delta_{10} + \delta_{11}R)\alpha(t) - \kappa\beta(t))\,dt - wR + w - \rho w \int_0^\infty \exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\,dt\Big]$$
$$= \rho w + \rho w \int_0^\infty \Big[-\kappa(\lambda_\infty - \lambda)\alpha(t) - \rho + \lambda\mathbb{E}[(1-R)\exp(-(\delta_{10} + \delta_{11}R)\alpha(t)) - 1]\Big] \cdot \exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\,dt.$$

Using Equations (15)–(16), we obtain

$$(\mathcal{D}_0 u)(\lambda, w) = \rho w + \rho w \int_0^\infty [-\rho - \lambda\dot{\alpha}(t) - \kappa\dot{\beta}(t)] \cdot \exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\, dt = 0,$$

as claimed. □

Proof of Corollary 2. (i) The first claim has already been established in the proof of Theorem 2. For the second claim, note that for any $t > 0$ such that $\alpha(t) = 0$, $\dot{\alpha}(t) = \bar{r} > 0$; therefore, $\alpha(t)$ and $\beta(t)$ are strictly positive for $t > 0$. Hence, as $\hat{\lambda} \to \infty$, the integrand in Equation (14) converges uniformly to zero, which implies the results. (ii) The claims follow by direct calculation:

$$\begin{aligned}
\partial_1 u(\lambda, w) &= -\rho w \int_0^\infty \alpha(t)\exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\, dt \le 0;\\
\partial_{11} u(\lambda, w) &= \rho w \int_0^\infty \alpha^2(t)\exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\, dt \ge 0;\\
\partial_{12} u(\lambda, w) &= -\rho \int_0^\infty \alpha(t)\exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\, dt < 0;\\
\partial_2 u(\lambda, w) &= -\left(1 - \rho\int_0^\infty \exp(-\rho t - \lambda\alpha(t) - \kappa\beta(t))\, dt\right) < 0\\
&= \partial_{22} u(\lambda, w).
\end{aligned}$$

Note that for $w > 0$, all inequalities become strict. (iii) The claims follow immediately from the calculations in parts (i) and (ii), also from the definition of the autonomous account value $u(\lambda, w)$. □

The following auxiliary result is used in the proof of Theorem 3.

Proposition A.2. *For any* $(\lambda, w) \in \mathbb{R}_+^2$, *it is*

$$\{a_0^*(\lambda, w)\} = \arg\min_{\hat{a}\ge 0}\{u(\lambda + \delta_2\hat{a}, w) + c\hat{a}\}.$$

Moreover, the (nonpositive) decrement $u(\lambda + \delta_2 a_0^*(\lambda, w), w) + c a_0^*(\lambda, w) - u(\lambda, w)$ *is* (i) *increasing in* λ, (ii) *decreasing in* w, *and* (iii) *concave with respect to* w.

Proof. Since $u(\cdot, w)$ is strictly convex (see part (ii) of Corollary 2), the minimizer is unique (i.e., the argument of the minimization problem is a singleton). Parts (i) and (ii) can be shown by direct calculation (using the envelope theorem):

$$\frac{d}{d\lambda}[u(\lambda + \delta_2 a_0^*(\lambda, w), w) + c a_0^*(\lambda, w) - u(\lambda, w)] = \max\{\partial_1 u(\lambda, w), -c/\delta_2\} - \partial_1 u(\lambda, w) \ge 0,$$

and

$$\frac{d}{dw}[u(\lambda + \delta_2 a_0^*(\lambda, w), w) + c a_0^*(\lambda, w) - u(\lambda, w)] = \partial_2 u(\lambda + \delta_2 a_0^*(\lambda, w), w) - \partial_2 u(\lambda, w) \le 0,$$

where the last inequality holds since, by Corollary 2(ii), $\partial_{12} u(\lambda, w) < 0$. Concavity follows, since $u(\lambda + \delta_2\hat{a}, w) + c\hat{a} - u(\lambda, w)$ is concave (affine) in w, and this property is preserved under pointwise minimization. □

Proof of Theorem 3. By construction, $v_0^*(\lambda, w)$ is continuous and has continuous first-order derivatives (see the proof of Corollary 3 for details). For $\lambda > \lambda_0^*(w)$, $(\mathcal{D}_0 v_0^*)(\lambda, w) = (\mathcal{D}_0 u)(\lambda, w) = 0$, and $\partial_1 v_0^*(\lambda, w) = \partial_1 u(\lambda, w) > -c/\delta_2$. Consequently, it follows that $\min_{a_0 > 0}\{u(\lambda + \delta_2 a_0, w) + c a_0\}$ has no solution, and $(\mathcal{D}_\epsilon v_0^*)(\lambda, w) > 0$ for any $\epsilon > 0$. For $\lambda = \lambda_0^*(w)$, it is $\partial_1 v_0^*(\lambda, w) = -c/\delta_2$. Thus, similarly, $\min_{a_0 > 0}\{u(\lambda + \delta_2 a_0, w) + c a_0\}$ has no solution; however, it is $(\mathcal{D}_\epsilon v_0^*)(\lambda, w) = 0$ for any $\epsilon \ge 0$. Note that since $\lambda_0^*(w) \le \lambda_\infty$, it is $\kappa(\lambda - \lambda_\infty)/\delta_2 \le 0$, and hence, $\mathbb{R}_+ \times [0, \underline{w}] \cup [\lambda_0^*(w), \infty) \times (\underline{w}, w_0^*] = \mathcal{C}^* \cap \mathbb{R}_+ \times [0, w_0^*]$ and $\mathcal{H}^* \cap \mathbb{R}_+ \times [0, w_0^*] = \varnothing$. In other words, if $(\lambda(t), W(t)) = (\lambda_0^*(w), w)$, then $(\lambda(t+s), W(t+s)) \in \mathcal{C}^*$ for any $s \ge 0$ and any choice of A. Finally, one needs to establish that for $\lambda < \lambda_0^*(w)$, $(\mathcal{D}_\epsilon v_0^*)(\lambda, w) \ge 0$, for $\epsilon \ge 0$. For this, consider the fact that

$$\begin{aligned}
&(\mathcal{D}_\epsilon v_0^*)(\lambda, w)\\
&= (\mathcal{D}_\epsilon v_0^*)(\lambda, w) - (\mathcal{D}_0 u)(\lambda, w)\\
&= -\rho(v_0^*(\lambda, w) - u(\lambda, w)) + \kappa(\lambda_\infty - \lambda)(\partial_1 v_0^*(\lambda, w) - \partial_1 u(\lambda, w))\\
&\quad + \delta_2\epsilon\,\partial_1 v_0^*(\lambda, w) + c\epsilon\\
&\quad + \lambda\mathbb{E}[v_0^*(\lambda + \delta_{10} + \delta_{11}R, w(1-R))\\
&\quad - u(\lambda + \delta_{10} + \delta_{11}R, w(1-R))] - \lambda(v_0^*(\lambda, w) - u(\lambda, w))\\
&= -\rho(v_0^*(\lambda, w) - u(\lambda, w)) + \kappa(\lambda_\infty - \lambda)(\partial_1 v_0^*(\lambda, w) - \partial_1 u(\lambda, w))\\
&\quad + \lambda\mathbb{E}[v_0^*(\lambda + \delta_{10} + \delta_{11}R, w(1-R))\\
&\quad - u(\lambda + \delta_{10} + \delta_{11}R, w(1-R))] - \lambda(v_0^*(\lambda, w) - u(\lambda, w))
\end{aligned}$$

Since $v_0^*(\lambda, w) = u(\lambda + \delta_2 a_0^*(\lambda, w), w) + c a_0^*(\lambda, w) = \min_{a_0 \ge 0}\{u(\lambda + \delta_2 a_0, w) + c a_0\} \le u(\lambda, w)$ and $\partial_1 u(\lambda, w) < -c/\delta_2$, all terms in the expression of $(\mathcal{D}_\epsilon v_0^*)(\lambda, w)$, except for the argument of the expectation operator, are positive. For any $r \in [0, 1]$, if $\lambda + \delta_{10} + \delta_{11} r \ge \lambda_0^*(w(1-r))$, then $v_0^*(\lambda + \delta_{10} + \delta_{11} r, w(1-r)) - u(\lambda + \delta_{10} + \delta_{11} r, w(1-r)) = 0$. Moreover, if $\lambda + \delta_{10} + \delta_{11} r < \lambda_0^*(w(1-r))$, then by Proposition A.2:

$$\begin{aligned}
&v_0^*(\lambda + \delta_{10} + \delta_{11} r, w(1-r)) - u(\lambda + \delta_{10} + \delta_{11} r, w(1-r))\\
&\quad \ge v_0^*(\lambda, w(1-r)) - u(\lambda, w(1-r))\\
&\quad \ge v_0^*(\lambda, w) - u(\lambda, w).
\end{aligned}$$

It follows that $(\mathcal{D}_\epsilon v_0^*)(\lambda, w) \ge 0$ for $\lambda - \lambda_0^*(w) < 0 \le \epsilon$. □

Proof of Corollary 3. It is enough to establish the claims for $w \in [\underline{w}, w_0^*]$ since otherwise $v_0^*(\lambda, w) = u(\lambda, w)$, and Corollary 3 becomes a special case of Corollary 2. Since $v_0^*(\lambda, w)$ is twice continuously differentiable for $\lambda \ne \lambda_0^*(w)$, let us focus on the interesting case where $\lambda = \lambda_0^*(w)$. Note that continuity of $v_0^*(\lambda, w)$ at $\lambda = \lambda_0^*(w)$ follows from its definition. For $\lambda < \lambda_0^*(w)$, by direct calculation, $\partial_1 v_0^*(\lambda, w) = -c/\delta_2$ and $\partial_2 v_0^*(\lambda, w) = \partial_2 u(\lambda + \delta_2 a_0^*(\lambda, w), w)$. Moreover, by the definition of $\lambda_0^*(w)$, it is $a_0^*(\lambda_0^*(w), w) = 0$ and $\partial_1 u(\lambda_0^*(w), w) = -c/\delta_2$. Thus, $\lim_{\lambda\to\lambda_0^*(w)^+} \partial_1 u(\lambda, w) = -c/\delta_2$, $\lim_{\lambda\to\lambda_0^*(w)^+} \partial_2 u(\lambda, w) = \lim_{\lambda\to\lambda_0^*(w)^-} \partial_2 u(\lambda + \delta_2 a^*(\lambda, w), w)$, $\lim_{\hat{w}\to w^-} \partial_1 u(\lambda_0^*(w), \hat{w}) = -c/\delta_2$, and finally

$$\begin{aligned}
\lim_{\hat{w}\to w^+} \partial_2 v_0^*(\lambda_0^*(w), \hat{w}) &= \lim_{\hat{w}\to w^+} \partial_2 u(\lambda_0^*(w) + \delta_2 a_0^*(\lambda_0^*(w), \hat{w}), \hat{w})\\
&= \partial_2 u(\lambda_0^*(w), w) = \lim_{\hat{w}\to w^-} \partial_2 v_0^*(\lambda_0^*(w), \hat{w}).
\end{aligned}$$

Hence, $v_0^*(\lambda, w)$ is C^1 and a.e. twice differentiable. (i) By Corollary 2, $-w \le u(\lambda, w)$, and hence, $-w \le v_0^*(\lambda, w) = \min_{a_0 \ge 0} u(\lambda + \delta_2 a_0, w) + c a_0 \le u(\lambda, w) \le 0$, which implies the first claim.

The second claim follows since, for $\hat{\lambda} \geq \lambda_0^*(w)$, $v_0^*(\hat{\lambda}, w) = u(\hat{\lambda}, w)$, and again by Corollary 2 it is $\lim_{\hat{\lambda}\to\infty} u(\hat{\lambda}, w) = -w$. (ii) By construction, for $\lambda \leq \lambda_0^*(w)$, it is $\partial_1 v_0^*(\lambda, w) = -\hat{c}$. Since for $\lambda \geq \lambda_0^*(w)$, $\partial_1 v_0^*(\lambda, w) = \partial_1 u(\lambda, w)$, the claim follows from Corollary 2. (iii) Similar to part (ii), by Corollary 2 the claims hold for $\lambda > \lambda_0^*(w)$. For $\lambda < \lambda_0^*(w)$, by the definition of $v_0^*(\lambda, w)$ we have $\partial_2 v_0^*(\lambda, w) = \partial_2 u(\lambda + \delta_2 a_0^*(\lambda, w), w) \leq 0$, $\partial_{12} v_0^*(\lambda, w) = \partial_{11} v_0^*(\lambda, w) = 0$, and $\partial_{22} v_0^*(\lambda, w) = \partial_{12} u(\lambda + \delta_2 a_0^*(\lambda, w), w)\delta_2 \partial_2 a_0^*(\lambda, w) \leq 0$. Note that at $\lambda = \lambda_0^*(w)$ the function $v_0^*(\lambda, w)$ does not have continuous second-order derivatives. Nevertheless, its one-sided second-order derivatives (left and right) satisfy the relevant inequalities. □

Proof of Corollary 4. The result can be obtained analogously to Corollary 1. □

Proof of Lemma 1. Assuming $\lambda \geq \hat{\lambda} \geq \lambda_\infty$, the present value of a unit repayment under sustained intensity process is

$$\begin{aligned}\mathbb{E}[e^{-\rho T} \mid \lambda(0) = \lambda] &= \int_0^\infty e^{-\rho s}\, dF_{\lambda,\hat{\lambda}}(s)\\ &= \int_0^{\theta(\lambda,\hat{\lambda})} e^{-\rho s} \exp(-\lambda_\infty s - (\lambda - \varphi(s,\lambda))/\kappa)\varphi(s,\lambda)\, ds\\ &\quad + \int_{\theta(\lambda,\hat{\lambda})}^\infty e^{-\rho s} \exp(-\lambda_\infty \theta(\lambda,\hat{\lambda}) - (\lambda - \hat{\lambda})/\kappa\\ &\qquad - \hat{\lambda}(s - \theta(\lambda,\hat{\lambda})))\hat{\lambda}\, ds.\end{aligned}$$

Using a change of variable, the first integral becomes

$$\begin{aligned}&\int_0^{\theta(\lambda,\hat{\lambda})} e^{-\rho s} \exp(-\lambda_\infty s - (\lambda - \varphi(s,\lambda))/\kappa)\varphi(s,\lambda)\, ds\\ &\quad = \int_\lambda^{\hat{\lambda}} e^{-\rho\theta(\lambda,l)} \exp(-\lambda_\infty \theta(\lambda,l) - (\lambda - l)/\kappa)\frac{-l}{\kappa(l - \lambda_\infty)}\, dl\\ &\quad = \int_{\hat{\lambda}}^\lambda \left(\frac{l - \lambda_\infty}{\lambda - \lambda_\infty}\right)^{(\rho+\lambda_\infty)/\kappa} \exp\left(-\frac{\lambda - l}{\kappa}\right)\frac{l}{\kappa(l - \lambda_\infty)}\, dl.\end{aligned}$$

Substituting the expression of $\theta(\lambda, \hat{\lambda})$, given in the main text, yields

$$\begin{aligned}&\int_{\theta(\lambda,\hat{\lambda})}^\infty e^{-\rho s} \exp(-\lambda_\infty \theta(\lambda,\hat{\lambda}) - (\lambda - \hat{\lambda})/\kappa - \hat{\lambda}(s - \theta(\lambda,\hat{\lambda})))\hat{\lambda}\, ds\\ &\quad = \exp(-(\rho + \lambda_\infty)\theta(\lambda,\hat{\lambda}))\exp\left(-\frac{\lambda - \hat{\lambda}}{\kappa}\right)\\ &\qquad \cdot \int_0^\infty \exp(-(\rho + \hat{\lambda})s)\hat{\lambda}\, ds\\ &\quad = \left(\frac{\hat{\lambda} - \lambda_\infty}{\lambda - \lambda_\infty}\right)^{(\rho+\lambda_\infty)/\kappa} \exp\left(-\frac{\lambda - \hat{\lambda}}{\kappa}\right)\frac{\hat{\lambda}}{(\rho + \hat{\lambda})}.\end{aligned}$$

Combining these two integrals gives the result. □

Proof of Lemma 2. (i) Similar to Corollary 3, it is enough to show that $\hat{v}(\lambda, w; \hat{\lambda})$ is continuously differentiable at $\lambda = \hat{\lambda}$. Note that the continuity of $\hat{v}(\lambda, w; \hat{\lambda})$ follows from the definition. For $\lambda > \hat{\lambda}$, it is

$$\begin{aligned}&\partial_1 \hat{v}(\lambda, w; \hat{\lambda})\\ &\quad = \frac{\lambda}{\kappa(\lambda - \lambda_\infty)}\bar{v}(\lambda, w) + \int_{\hat{\lambda}}^\lambda \bar{v}(l, w)\partial_1 Q(\lambda, l)\frac{l + \rho}{\kappa(l - \lambda_\infty)}\, dl\\ &\qquad + \left[\bar{v}(\hat{\lambda}, w) + \hat{c}\kappa\frac{\hat{\lambda} - \lambda_\infty}{\hat{\lambda}}\right]\partial_1 Q(\lambda, \hat{\lambda}),\end{aligned}$$

$$\begin{aligned}&\partial_2 \hat{v}(\lambda, w; \hat{\lambda})\\ &\quad = \int_{\hat{\lambda}}^\lambda \partial_2 \bar{v}(l, w)Q(\lambda, l)\frac{l + \rho}{\kappa(l - \lambda_\infty)}\, dl + \partial_2 \bar{v}(\hat{\lambda}, w)Q(\lambda, \hat{\lambda}).\end{aligned}$$

For $\lambda < \hat{\lambda}$, $\partial_1 \hat{v}(\lambda, w; \hat{\lambda}) = -\hat{c}$ and $\partial_2 \hat{v}(\lambda, w; \hat{\lambda}) = \partial_2 \bar{v}(\hat{\lambda}, w) \cdot Q(\hat{\lambda}, \hat{\lambda})$. The derivative of $Q(\lambda, \hat{\lambda})$ with respect to λ is

$$\begin{aligned}&\partial_1 Q(\lambda, \hat{\lambda})\\ &\quad = \frac{\hat{\lambda}}{(\rho + \hat{\lambda})}\left(\frac{\hat{\lambda} - \lambda_\infty}{\lambda - \lambda_\infty}\right)^{(\rho+\lambda_\infty)/\kappa} \exp\left(-\frac{\lambda - \hat{\lambda}}{\kappa}\right)\left(\frac{-(\rho + \lambda_\infty)}{\kappa(\lambda - \lambda_\infty)} - \frac{1}{\kappa}\right)\\ &\quad = \frac{\hat{\lambda}}{(\rho + \hat{\lambda})}\left(\frac{\hat{\lambda} - \lambda_\infty}{\lambda - \lambda_\infty}\right)^{(\rho+\lambda_\infty)/\kappa} \exp\left(-\frac{\lambda - \hat{\lambda}}{\kappa}\right)\left(\frac{-(\rho + \lambda)}{\kappa(\lambda - \lambda_\infty)}\right)\end{aligned}$$

so that $\lim_{\lambda\to\hat{\lambda}^+} \partial_1 \hat{v}(\lambda, w; \hat{\lambda}) = -\hat{c}$, and $\lim_{\lambda\to\hat{\lambda}^+} \partial_2 \hat{v}(\lambda, w; \hat{\lambda}) = \lim_{\lambda\to\hat{\lambda}^-} \partial_2 \hat{v}(\lambda, w; \hat{\lambda})$. The first- and second-order monotonicity of $\hat{v}$ with respect to w follow from the corresponding properties of v and the fact that these properties are preserved by (discounted) expectation. The monotonicity of $\hat{v}$ and $\partial_2 \hat{v}$ with respect to λ are obtained since for $\lambda \geq \hat{\lambda}$,

$$\begin{aligned}\hat{v}(\lambda, w; \hat{\lambda}) &= \int_0^\infty \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda, s)\}, w)e^{-\rho s}\, dF_{\lambda,\hat{\lambda}}(s)\\ &\quad + \hat{c}\kappa\frac{\hat{\lambda} - \lambda_\infty}{\hat{\lambda}}Q(\lambda, \hat{\lambda}),\\ \partial_2 \hat{v}(\lambda, w; \hat{\lambda}) &= \int_0^\infty \partial_2 \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda, s)\}, w)e^{-\rho s}\, dF_{\lambda,\hat{\lambda}}(s).\end{aligned}$$

In particular, $\hat{c}(\kappa(\hat{\lambda} - \lambda_\infty)/\hat{\lambda})Q(\lambda, \hat{\lambda})$ is decreasing in λ. Moreover, the functions $\bar{v}(\max\{\hat{\lambda}, \varphi(\lambda, s)\}, w)e^{-\rho s}$ and $\partial_2 \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda, s)\}, w)e^{-\rho s}$ are increasing in s. Hence, if $\lambda_1 > \lambda_2 > \hat{\lambda}$, then by the first-order stochastic dominance order in the family of distributions $F_{\lambda,\hat{\lambda}}$, as discussed in the main text, it is

$$\begin{aligned}&\int_0^\infty \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda_1, s)\}, w)e^{-\rho s}\, dF_{\lambda_1,\hat{\lambda}}(s)\\ &\quad \leq \int_0^\infty \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda_2, s)\}, w)e^{-\rho s}\, dF_{\lambda_1,\hat{\lambda}}(s)\\ &\quad \leq \int_0^\infty \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda_2, s)\}, w)e^{-\rho s}\, dF_{\lambda_2,\hat{\lambda}}(s),\end{aligned}$$

and similarly

$$\begin{aligned}&\int_0^\infty \partial_2 \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda_1, s)\}, w)e^{-\rho s}\, dF_{\lambda_1,\hat{\lambda}}(s)\\ &\quad \leq \int_0^\infty \partial_2 \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda_2, s)\}, w)e^{-\rho s}\, dF_{\lambda_1,\hat{\lambda}}(s)\\ &\quad \leq \int_0^\infty \partial_2 \bar{v}(\max\{\hat{\lambda}, \varphi(\lambda_2, s)\}, w)e^{-\rho s}\, dF_{\lambda_2,\hat{\lambda}}(s).\end{aligned}$$

(ii) Since $\hat{v}(\lambda, w; \hat{\lambda})$ is continuously differentiable with respect to $\hat{\lambda}$ as long as $\hat{\lambda} \neq \lambda$, one can restrict attention to checking that its right- and left-derivatives coincide for $\hat{\lambda} = \lambda$. For $\hat{\lambda} > \lambda$, substituting the expression for $Q(\hat{\lambda}, \hat{\lambda})$ using Equation (22) and taking the derivative of $\hat{v}(\lambda, w; \hat{\lambda})$ with respect to $\hat{\lambda}$ yields

$$\begin{aligned}\frac{d}{d\hat{\lambda}}\hat{v}(\lambda, w; \hat{\lambda}) &= \frac{\kappa(\rho + \lambda_\infty)}{(\hat{\lambda} + \rho)^2}\left[\hat{c}\frac{(\hat{\lambda} + \rho)^2}{\kappa(\rho + \lambda_\infty)} + \hat{c}\right.\\ &\quad \left. + \frac{\rho}{\kappa(\rho + \lambda_\infty)}\bar{v}(\hat{\lambda}, w) + \frac{\hat{\lambda}(\hat{\lambda} + \rho)}{\kappa(\rho + \lambda_\infty)}\partial_1 \bar{v}(\hat{\lambda}, w)\right].\end{aligned}$$

Similarly, for $\hat{\lambda} < \lambda$, by direct calculation,

$$\frac{d}{d\hat{\lambda}}\hat{v}(\lambda,w;\hat{\lambda}) = \left(\frac{\hat{\lambda}-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\lambda_\infty+\rho)/\kappa}\exp\left(-\frac{\lambda-\hat{\lambda}}{\kappa}\right)\frac{\kappa(\rho+\lambda_\infty)}{(\hat{\lambda}+\rho)^2} \cdot\left[\hat{c}\frac{(\hat{\lambda}+\rho)^2}{\kappa(\rho+\lambda_\infty)}+\hat{c}+\frac{\rho}{\kappa(\rho+\lambda_\infty)}\bar{v}(\hat{\lambda},w)+\frac{\hat{\lambda}(\hat{\lambda}+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_1\bar{v}(\hat{\lambda},w)\right].$$

Hence, one obtains

$$\lim_{\hat{\lambda}\to\lambda^+}\frac{d}{d\hat{\lambda}}\hat{v}(\lambda,w;\hat{\lambda}) = \lim_{\hat{\lambda}\to\lambda^-}\frac{d}{d\hat{\lambda}}\hat{v}(\lambda,w;\hat{\lambda}).$$

The second claim follows since, for $\lambda < \hat{\lambda}$, $Q(\hat{\lambda},\hat{\lambda}) = \hat{\lambda}/(\rho+\hat{\lambda})$ is positive and increasing while $\partial_2\bar{v}(\hat{\lambda},w)$ is negative and decreasing. For $\lambda \geq \hat{\lambda}$, similar to part (i), the result holds since $\partial_2\bar{v}(\max\{\hat{\lambda},\varphi(\lambda,s)\},w)e^{-\rho s}$ is an increasing function of s, and for $\hat{\lambda}_1 > \hat{\lambda}_2$, $F_{\lambda,\hat{\lambda}_1} <_{\mathrm{st}} F_{\lambda,\hat{\lambda}_2}$. Specifically,

$$\int_0^\infty \partial_2\bar{v}(\max\{\hat{\lambda}_1,\varphi(\lambda,s)\},w)e^{-\rho s}\,dF_{\lambda,\hat{\lambda}_1}(s) \leq \int_0^\infty \partial_2\bar{v}(\max\{\hat{\lambda}_2,\varphi(\lambda,s)\},w)e^{-\rho s}\,dF_{\lambda,\hat{\lambda}_1}(s) \leq \int_0^\infty \partial_2\bar{v}(\max\{\hat{\lambda}_2,\varphi(\lambda,s)\},w)e^{-\rho s}\,dF_{\lambda,\hat{\lambda}_2}(s),$$

which concludes the proof. □

Proof of Theorem 4. The derivative of $f_i(\hat{\lambda},w)$ with respect to $\hat{\lambda}$ is

$$\frac{d}{d\hat{\lambda}}f_i(\hat{\lambda},w) = \hat{c}\frac{2(\hat{\lambda}+\rho)}{\kappa(\rho+\lambda_\infty)}+\frac{2(\rho+\hat{\lambda})}{\kappa(\rho+\lambda_\infty)}\partial_1\bar{v}^*_{i-1}(\hat{\lambda},w) + \frac{\hat{\lambda}(\hat{\lambda}+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_{11}\bar{v}^*_{i-1}(\hat{\lambda},w) \geq 0;$$

the inequality holds since by Equation (24)

$$\bar{v}^*_{i-1}(\hat{\lambda},w) = \mathbb{E}[v^*_{i-1}(\hat{\lambda}+\delta_{10}+\delta_{11}R, w(1-R)) - wR],$$

and by the value properties (P2)–(P3) it is $\partial_1 v^*_{i-1}(\lambda,w) \geq -\hat{c}$ and $\partial_{11}v^*_{i-1}(\lambda,w) \geq 0$. Since by the value property (P2) it is $\lim_{\hat{\lambda}\to\infty}\partial_1 v^*_{i-1}(\hat{\lambda},w) = 0$, the gradient $f_i(\hat{\lambda},w)$ must be positive as long as $\hat{\lambda}$ is large enough. Therefore, it is possible to restrict attention to establishing that $f_i(\hat{\lambda},w) \leq 0$ for some finite $\hat{\lambda} \geq 0$. For $i = 1$, this is achieved at $\hat{\lambda} = \lambda_\infty$ because $\lambda^*_0(w^*_0) = \lambda_\infty$, $\partial_1 u(\lambda_\infty, w^*_0) = -\hat{c}$, $v^*_0(\hat{\lambda}+\delta_{10}+\delta_{11}R, w(1-R)) = u(\hat{\lambda}+\delta_{10}+\delta_{11}R, w(1-R))$ for $\hat{\lambda} \geq \lambda_\infty$, and by Corollary 2(i),

$$\frac{(\hat{\lambda}+\rho)^2}{\kappa(\rho+\lambda_\infty)}\partial_1 u(\hat{\lambda},w)+\partial_1 u(\hat{\lambda},w)+\frac{(\hat{\lambda}-\lambda_\infty)(\hat{\lambda}+\rho)}{(\rho+\lambda_\infty)}\partial_{11}u(\hat{\lambda},w) = \frac{\rho}{\kappa(\rho+\lambda_\infty)}\mathbb{E}[u(\hat{\lambda}+\delta_{10}+\delta_{11}R, w(1-R)) - wR] + \frac{\hat{\lambda}(\hat{\lambda}+\rho)}{\kappa(\rho+\lambda_\infty)}\mathbb{E}[\partial_1 u(\hat{\lambda}+\delta_{10}+\delta_{11}R, w(1-R))];$$

thus, $f_1(\lambda_\infty, w^*_0) = 0$.[23] Taking the derivative $f_1(\lambda_\infty, w)$ with respect to w and using the value property (P3) yields that

$$\frac{\rho}{\kappa(\rho+\lambda_\infty)}\partial_2\bar{v}^*_{i-1}(\hat{\lambda},w)+\frac{\hat{\lambda}(\hat{\lambda}+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_{12}\bar{v}^*_{i-1}(\hat{\lambda},w) < 0,$$

which implies the result. For $i > 1$, the claim follows via induction using a similar argument. Specifically, assume that $\lambda^*_{i-1}(w)$ satisfies the optimality condition (30) for all $w \in [w^*_{i-2}, w^*_{i-1}]$. Hence,

$$\hat{c}\frac{(\lambda^*_{i-1}(w)+\rho)^2}{\kappa(\rho+\lambda_\infty)}+\hat{c}+\frac{\rho}{\kappa(\rho+\lambda_\infty)}\bar{v}^*_{i-2}(\lambda^*_{i-1}(w),w) + \frac{\lambda^*_{i-1}(w)(\lambda^*_{i-1}(w)+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_1\bar{v}^*_{i-2}(\lambda^*_{i-1}(w),w) = 0.$$

On the other hand, by the definition of $v_i(\lambda,w;\hat{\lambda})$ and $v^*_{i-1}(\lambda,w)$ as well as by virtue of the value properties,

$$\begin{aligned}
&\hat{c}\frac{(\lambda^*_{i-1}(w^*_{i-1})+\rho)^2}{\kappa(\rho+\lambda_\infty)}+\hat{c}+\frac{\rho}{\kappa(\rho+\lambda_\infty)}\bar{v}^*_{i-2}(\lambda^*_{i-1}(w^*_{i-1}),w^*_{i-1}) + \frac{\lambda^*_{i-1}(w^*_{i-1})(\lambda^*_{i-1}(w^*_{i-1})+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_1\bar{v}^*_{i-2}(\lambda^*_{i-1}(w^*_{i-1}),w^*_{i-1})\\
&= \hat{c}\frac{(\lambda^*_{i-1}(w^*_{i-1})+\rho)^2}{\kappa(\rho+\lambda_\infty)}+\hat{c}+\frac{\rho}{\kappa(\rho+\lambda_\infty)}\bar{v}^*_{i-1}(\lambda^*_{i-1}(w^*_{i-1}),w^*_{i-1}) + \frac{\lambda^*_{i-1}(w^*_{i-1})(\lambda^*_{i-1}(w^*_{i-1})+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_1\bar{v}^*_{i-1}(\lambda^*_{i-1}(w^*_{i-1}),w^*_{i-1}) = 0.
\end{aligned}$$

Therefore, it is $f_i(\lambda^*_{i-1}(w^*_{i-1}),w^*_{i-1}) = f_{i-1}(\lambda^*_{i-1}(w^*_{i-1}),w^*_{i-1}) = 0$. Taking the derivative $f_i(\lambda^*_{i-1}(w^*_{i-1}),w)$ with respect to w yields

$$\frac{\rho}{\kappa(\rho+\lambda_\infty)}\partial_2\bar{v}^*_{i-1}(\lambda^*_{i-1}(w^*_{i-1}),w) + \frac{\lambda^*_{i-1}(w^*_{i-1})(\lambda^*_{i-1}(w^*_{i-1})+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_{12}\bar{v}^*_{i-1}(\lambda^*_{i-1}(w^*_{i-1}),w) < 0;$$

the last inequality is implied by the value property (P3). Hence, $f_i(\lambda^*_{i-1}(w^*_{i-1}),w)$ is negative for $w \in (w^*_{i-1}, w^*_i]$, which allows us to conclude that $\lambda^*_i(w)$ exists. The uniqueness of the solution follows from the fact that when the optimality condition $f_i(\hat{\lambda},w) = 0$ holds, then the second-order condition $\partial_1 f_i(\hat{\lambda},w) > 0$ is also satisfied. Indeed, if this is not true, then $\partial_1 f_i(\hat{\lambda},w) = 0$, which implies $\partial_1\bar{v}^*_{i-1}(\hat{\lambda},w) = -\hat{c}$ and $\partial_{11}\bar{v}^*_{i-1}(\hat{\lambda},w) = 0$. But this means that $\partial_1 v^*_{i-1}(\hat{\lambda}+\delta_{10}+\delta_{11}R, w(1-R)) = -\hat{c}$ and $\partial_{11}v^*_{i-1}(\hat{\lambda}+\delta_{10}+\delta_{11}R, w(1-R)) = 0$, F_R-almost surely. By the value properties, it is $\partial_{11}v^*_{i-1}(\lambda,w) \geq 0$ and $\partial_1 v^*_{i-1}(\lambda,w) \geq -\hat{c}$, so $\partial_1 v^*_{i-1}(\lambda+\delta_{10}+\delta_{11}R, w(1-R)) = -\hat{c}$ and $\partial_{11}v^*_{i-1}(\lambda+\delta_{10}+\delta_{11}R, w(1-R)) = 0$ for any $\lambda < \hat{\lambda}$. Hence, $f_i(\lambda,w) = 0$ for $\lambda \leq \hat{\lambda}$, that is, in particular for $\hat{\lambda} = \lambda_\infty$. This contradicts the fact that $f_i(\hat{\lambda},w)$ is increasing in $\hat{\lambda}$ and that there exists $\hat{\lambda} \geq \lambda_\infty$ at which $f_i(\hat{\lambda},w) < 0$. Thus, $\lambda^*_i(w)$ must be unique. Finally, the monotonicity obtains since $\partial_2 f_i(\hat{\lambda},w) < 0$. □

The following auxiliary result is used in the proof of Lemma 3.

Proposition A.3. *For any $\hat{\lambda} \geq \lambda_\infty$, $v_i(\lambda,w;\hat{\lambda})$ has the following properties:*

(i) *For any $(\lambda,w) \in [\hat{\lambda},\infty)\times[w^*_{i-1},w^*_i]$, it is*

$$\kappa(\lambda_\infty-\lambda)\partial_1 v_i(\lambda,w;\hat{\lambda}) - \rho v_i(\lambda,w;\hat{\lambda}) + \lambda\mathbb{E}[v^*_{i-1}(\lambda+\delta_{10}+\delta_{11}R, w(1-R)) - v_i(\lambda,w;\hat{\lambda}) - wR] = 0.$$

(ii) $\lim_{\lambda\to\hat{\lambda}^+}\partial_1 v_i(\lambda,w;\hat{\lambda}) = -\hat{c}$.

(iii) $\lim_{\lambda\to\hat{\lambda}^+}\partial_{11}v_i(\lambda,w;\hat{\lambda}) = ((\lambda_\infty+\rho)/((\hat{\lambda}-\lambda_\infty)(\hat{\lambda}+\rho)))\cdot f_i(\hat{\lambda},w)$ *where $f_i(\hat{\lambda},w)$ is defined in Equation (30).*

Proof. (i) Using the calculations in the proof of Lemma 2(i), it is

$$\begin{aligned}\partial_1 v_i(\lambda,w;\hat\lambda) &= \frac{\lambda}{\kappa(\lambda-\lambda_\infty)}\bar v^*_{i-1}(\lambda,w)\\ &\quad+\int_{\hat\lambda}^{\lambda}\bar v^*_{i-1}(l,w)\partial_1 Q(\lambda,l)\frac{l+\rho}{\kappa(l-\lambda_\infty)}\,dl\\ &\quad+\left[\bar v^*_{i-1}(\hat\lambda,w)+\hat c\kappa\frac{\hat\lambda-\lambda_\infty}{\hat\lambda}\right]\partial_1 Q(\lambda,\hat\lambda)\\ &=\frac{\lambda}{\kappa(\lambda-\lambda_\infty)}\bar v^*_{i-1}(\lambda,w)-\frac{(\rho+\lambda)}{\kappa(\lambda-\lambda_\infty)}v_i(\lambda,w;\hat\lambda),\end{aligned}$$

which implies the claim. (ii) The result is obtained analogously to the proof of Lemma 2(i). (iii) The second derivative of $v_i(\lambda,w;\hat\lambda)$ with respect to λ is

$$\begin{aligned}&\partial_{11} v_i(\lambda,w;\hat\lambda)\\ &=-\frac{\lambda+\rho}{\kappa(\lambda-\lambda_\infty)}\partial_1 v_i(\lambda,w;\hat\lambda)+\frac{\lambda_\infty+\rho}{\kappa(\lambda-\lambda_\infty)^2}v_i(\lambda,w;\hat\lambda)\\ &\quad+\frac{\lambda}{\kappa(\lambda-\lambda_\infty)}\partial_1\bar v^*_{i-1}(\lambda,w)-\frac{\lambda_\infty}{\kappa(\lambda-\lambda_\infty)^2}\bar v^*_{i-1}(\lambda,w),\end{aligned}$$

Taking the limit and substituting the corresponding expressions for $v_i(\hat\lambda,w;\hat\lambda)$ and $\partial_1 v_i(\hat\lambda,w;\hat\lambda)$ yield

$$\lim_{\lambda\to\hat\lambda^+}\partial_{11} v_i(\lambda,w;\hat\lambda)=\frac{\lambda_\infty+\rho}{(\hat\lambda-\lambda_\infty)(\hat\lambda+\rho)}f_i(\hat\lambda,w),$$

which completes the proof. □

Proof of Lemma 3. By definition (31) and Corollary 3, it suffices to prove the statements for $w\in[w^*_{i-1},w^*_i]$. Note that $v^*_i(\lambda,w)$ is twice continuously differentiable for $(\lambda,w)\in(\lambda^*_i(w),\infty)\times(w^*_{i-1},w^*_i]$ and $(\lambda,w)\in[0,\lambda^*_i(w))\times(w^*_{i-1},w^*_i]$. Therefore, we only need to show that its first-order derivatives are continuous at $(\lambda^*_i(w),w)$ when $w\in(w^*_{i-1},w^*_i]$ and (λ,w^*_{i-1}) when $\lambda\ge 0$. Note that continuity of $v^*_i(\lambda,w)$ at these two boundaries follows by the continuity of $\bar v^*_{i-1}$ and the definition of v^*_i. For any $(\lambda,w)\in(\lambda^*_i(w),\infty)\times(w^*_{i-1},w^*_i]$:

$$\begin{aligned}&\partial_1 v^*_i(\lambda,w)\\ &=\frac{\lambda}{\kappa(\lambda-\lambda_\infty)}\bar v^*_{i-1}(\lambda,w)\\ &\quad+\int_{\lambda^*_i(w)}^{\lambda}\bar v^*_{i-1}(l,w)\partial_1 Q(\lambda,l)\frac{l+\rho}{\kappa(l-\lambda_\infty)}\,dl\\ &\quad+\left[\bar v^*_{i-1}(\lambda^*_i(w),w)+\hat c\kappa\frac{\lambda^*_i(w)-\lambda_\infty}{\lambda^*_i(w)}\right]\partial_1 Q(\lambda,\lambda^*_i(w))\\ &=\frac{\lambda}{\kappa(\lambda-\lambda_\infty)}\bar v^*_{i-1}(\lambda,w)-\frac{(\rho+\lambda)}{\kappa(\lambda-\lambda_\infty)}v^*_i(\lambda,w),\end{aligned}\tag{A.5}$$

$$\begin{aligned}\partial_2 v^*_i(\lambda,w)&=\int_{\lambda^*_i(w)}^{\lambda}\partial_2\bar v^*_{i-1}(l,w)Q(\lambda,l)\frac{l+\rho}{\kappa(l-\lambda_\infty)}\,dl\\ &\quad+\partial_2\bar v^*_{i-1}(\lambda^*_i(w),w)Q(\lambda,\lambda^*_i(w)),\end{aligned}$$

where the derivative with respect to w is obtained using the envelope theorem. For $(\lambda,w)\in[0,\lambda^*_i(w))\times(w^*_{i-1},w^*_i]$:

$$\begin{aligned}&\partial_1 v^*_i(\lambda,w)=-\hat c,\\ &\partial_2 v^*_i(\lambda,w)=\partial_2\bar v^*_{i-1}(\lambda^*_i(w),w)Q(\lambda^*_i(w),\lambda^*_i(w)),\end{aligned}$$

where, as before, the optimality of $\lambda^*_i(w)$ (see Equation (30)) is taken into account when computing the derivative of v^*_i with respect to w. Therefore, by continuity of $\lambda^*_i(w)$, it is $\lim_{\lambda\to\lambda^*_i(w)^+}\partial_1 v^*_i(\lambda,w)=-c/\delta_2=-\hat c$, $\lim_{\lambda\to\lambda^*_i(w)^+}\partial_2 v^*_i(\lambda,w)=\partial_2\bar v^*_{i-1}(\lambda^*_i(w),w)Q(\lambda^*_i(w),\lambda^*_i(w))$, $\lim_{\hat w\to w^-}\partial_1 v^*_i(\lambda^*_i(w),\hat w)=-c/\delta_2=-\hat c$, and $\lim_{\hat w\to w^+}\partial_2 v^*_i(\lambda^*_i(w),\hat w)=\lim_{\hat w\to w^+}\partial_2\bar v^*_{i-1}(\lambda^*_i(\hat w),\hat w)Q(\lambda^*_i(\hat w),\lambda^*_i(\hat w))=\lim_{\hat w\to w^-}\partial_2 v^*_i(\lambda^*_i(w),\hat w)$. Hence, $v^*_i(\lambda,w)$ is C^1 in $\mathbb{R}_+\times(w^*_{i-1},w^*_i]$. It remains to be shown that $\lim_{w\to(w^*_{i-1})^-}\partial_1 v^*_{i-1}(\lambda,w)=\lim_{w\to(w^*_{i-1})^+}\partial_1 v^*_i(\lambda,w)$ and $\lim_{w\to(w^*_{i-1})^-}\partial_2 v^*_{i-1}(\lambda,w)=\lim_{w\to(w^*_{i-1})^+}\partial_2 v^*_i(\lambda,w)$. These properties follow since $\bar v^*_{i-1}$ is C^1.

***Value properties* (P1)–(P3).** Similar to the proof of Lemma 2(i), $v^*_i(\lambda,w)$ is decreasing and concave in w since $v^*_{i-1}(\lambda,w)$ is decreasing and concave in w, and these properties are preserved under both (discounted) expectation and pointwise minimization. Moreover, by the envelope theorem, $\partial_2 v^*_i(\lambda,w)=\partial_2 v_i(\lambda,w;\lambda^*_i(w))$; hence, by Lemma 2(i), $\partial_2 v^*_i(\lambda,w)$ is decreasing in λ. Note that by Equation (23) and the definition of $Q(\lambda,\hat\lambda)$ (see Equation (22)) one obtains that $\lim_{\lambda\to\infty}v^*_i(\lambda,w)=-w$. Using this and Equation (A.5), it immediately follows that $\lim_{\lambda\to\infty}\partial_1 v^*_i(\lambda,w)=0$. Moreover, as shown above, by the definition of $v^*_i(\lambda,w)$, it is $\partial_1 v^*_1(\lambda,w)=-\hat c$ for $\lambda\le\lambda^*_i(w)$. Hence, property (P2) is obtained by showing that $\partial_{11}v^*_i(\lambda,w)\ge 0$. To this end, we note that by Proposition A.3(i) for any $\lambda\ge\hat\lambda\ge\lambda^*_i(w)$:

$$\begin{aligned}v_i(\lambda,w;\hat\lambda)&=v^*_i(\lambda,w)+\left(\frac{\hat\lambda-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\lambda_\infty+\rho)/\kappa}\exp\left(-\frac{\lambda-\hat\lambda}{\kappa}\right)\\ &\quad\cdot\frac{\kappa(\hat\lambda-\lambda_\infty)}{\hat\lambda+\rho}\left[\partial_1 v^*_i(\hat\lambda,w)+\frac{c}{\delta_2}\right].\end{aligned}$$

Taking the derivative of $v_i(\lambda,w;\hat\lambda)$ with respect to $\hat\lambda$ yields

$$\begin{aligned}\frac{d}{d\hat\lambda}v_i(\lambda,w;\hat\lambda)&=\left(\frac{\hat\lambda-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\lambda_\infty+\rho)/\kappa}\exp\left(-\frac{\lambda-\hat\lambda}{\kappa}\right)\frac{\kappa(\lambda_\infty+\rho)}{(\hat\lambda+\rho)^2}\\ &\quad\cdot\left\{\left[\frac{(\hat\lambda+\rho)^2}{\kappa(\lambda_\infty+\rho)}+1\right]\left[\partial_1 v^*_i(\hat\lambda,w)+\frac{c}{\delta_2}\right]\right.\\ &\quad\left.+\frac{(\hat\lambda-\lambda_\infty)(\hat\lambda+\rho)}{(\rho+\lambda_\infty)}\partial_{11}v^*_i(\hat\lambda,w)\right\}.\end{aligned}\tag{A.6}$$

Equivalently, using the definition of $v_i(\lambda,w;\hat\lambda)$, definition (26), and Equation (22), one obtains

$$\begin{aligned}&\frac{d}{d\hat\lambda}v_i(\lambda,w;\hat\lambda)\\ &=-\left(\frac{\hat\lambda-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\lambda_\infty+\rho)/\kappa}\exp\left(-\frac{\lambda-\hat\lambda}{\kappa}\right)\frac{\hat\lambda}{\kappa(\hat\lambda-\lambda_\infty)}\bar v^*_{i-1}(\hat\lambda,w)\\ &\quad+\left(\frac{\hat\lambda-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\lambda_\infty+\rho)/\kappa}\exp\left(-\frac{\lambda-\hat\lambda}{\kappa}\right)\frac{\hat\lambda+\rho}{\kappa(\hat\lambda-\lambda_\infty)}\\ &\quad\cdot\left[\frac{\hat\lambda}{\hat\lambda+\rho}\bar v^*_{i-1}(\hat\lambda,w)+\frac{c}{\delta_2}\frac{\kappa(\hat\lambda-\lambda_\infty)}{\hat\lambda+\rho}\right]\\ &\quad+\left(\frac{\hat\lambda-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\lambda_\infty+\rho)/\kappa}\exp\left(-\frac{\lambda-\hat\lambda}{\kappa}\right)\\ &\quad\cdot\left[\frac{\rho}{(\hat\lambda+\rho)^2}\bar v^*_{i-1}(\hat\lambda,w)+\frac{\hat\lambda}{\hat\lambda+\rho}\partial_1\bar v^*_{i-1}(\hat\lambda,w)+\frac{c}{\delta_2}\frac{\kappa(\rho+\lambda_\infty)}{(\hat\lambda+\rho)^2}\right],\end{aligned}$$

which simplifies to

$$\begin{aligned}\frac{d}{d\hat\lambda}v_i(\lambda,w;\hat\lambda)&=\left(\frac{\hat\lambda-\lambda_\infty}{\lambda-\lambda_\infty}\right)^{(\lambda_\infty+\rho)/\kappa}\exp\left(-\frac{\lambda-\hat\lambda}{\kappa}\right)\\ &\quad\cdot\frac{\kappa(\rho+\lambda_\infty)}{(\hat\lambda+\rho)^2}f_i(\hat\lambda,w).\end{aligned}\tag{A.7}$$

Combining the preceding two equations implies that

$$\left[\frac{(\hat\lambda+\rho)^2}{\kappa(\lambda_\infty+\rho)}+1\right]\left[\partial_1 v_i^*(\hat\lambda,w)+\frac{c}{\delta_2}\right]+\frac{(\hat\lambda-\lambda_\infty)(\hat\lambda+\rho)}{(\rho+\lambda_\infty)}\partial_{11}v_i^*(\hat\lambda,w)$$
$$=\frac{c}{\delta_2}\left[\frac{(\hat\lambda+\rho)^2}{\kappa(\rho+\lambda_\infty)}+1\right]+\frac{\rho}{\kappa(\rho+\lambda_\infty)}\bar v_{i-1}^*(\hat\lambda,w)$$
$$+\frac{\hat\lambda(\hat\lambda+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_1\bar v_{i-1}^*(\hat\lambda,w).$$

Based on this, taking the derivative with respect to $\hat\lambda$ gives

$$\frac{2(\hat\lambda+\rho)}{\kappa(\lambda_\infty+\rho)}\left[\partial_1 v_i^*(\hat\lambda,w)+\frac{c}{\delta_2}\right]+\left[\frac{(\hat\lambda+\rho)^2}{\kappa(\lambda_\infty+\rho)}+1\right]\partial_{11}v_i^*(\hat\lambda,w)$$
$$+\frac{(\hat\lambda-\lambda_\infty)+(\hat\lambda+\rho)}{(\rho+\lambda_\infty)}\partial_{11}v_i^*(\hat\lambda,w)$$
$$+\frac{(\hat\lambda-\lambda_\infty)(\hat\lambda+\rho)}{(\rho+\lambda_\infty)}\partial_{111}v_i^*(\hat\lambda,w)$$
$$=\frac{2(\hat\lambda+\rho)}{\kappa(\rho+\lambda_\infty)}\left[\frac{c}{\delta_2}+\partial_1\bar v_{i-1}^*(\hat\lambda,w)\right]+\frac{\hat\lambda(\hat\lambda+\rho)}{\kappa(\rho+\lambda_\infty)}\partial_{11}\bar v_{i-1}^*(\hat\lambda,w). \tag{A.8}$$

Note that at $\hat\lambda=\lambda_i^*(w)$ it is $\partial_1 v_i^*(\lambda_i^*(w),w)=-c/\delta_2$, and by Proposition A.3(iii), $\partial_{11}v_i^*(\lambda_i^*(w),w)=0$. However, the right-hand side of the above equality is positive. Consequently, we can conclude that $\partial_{111}v_i^*(\lambda_i^*(w),w)>0$, which pushes $\partial_{11}v_i^*(\hat\lambda,w)$ into the positive domain and ensures that $\partial_1 v_i^*(\hat\lambda,w)+c/\delta_2>0$ for $\hat\lambda>\lambda_i^*(w)$. Now, assume there is $\hat\lambda>\lambda_i^*(w)$ for which $\partial_{11}v_i^*(\hat\lambda,w)=0$. Since $\hat\lambda\ge\lambda_i^*(w_{i-1}^*)=\lambda_{i-1}^*(w_{i-1}^*)$, one obtains

$$0<\partial_1 v_i^*(\hat\lambda,w)+\frac{c}{\delta_2}<\partial_1 v_i^*(\hat\lambda,w(1-R))+\frac{c}{\delta_2}$$
$$=\partial_1 v_{i-1}^*(\hat\lambda,w(1-R))+\frac{c}{\delta_2}$$
$$<\partial_1 v_{i-1}^*(\hat\lambda+\delta_{10}+\delta_{11}R,w(1-R))+\frac{c}{\delta_2}$$

where the last inequality follows from the convexity of $v_{i-1}^*(\lambda,w)$. Hence,

$$0<\partial_1 v_i^*(\hat\lambda,w)+\frac{c}{\delta_2}<\partial_1\bar v_{i-1}^*(\hat\lambda,w)+\frac{c}{\delta_2},$$

so that by Equation (A.8), $\partial_{11}v_i^*(\hat\lambda,w)$ never vanishes in the first place. Finally, we need to prove that $0\ge v_i^*(\lambda,w)$. For $\lambda\ge\lambda_\infty$, this follows, since $0\ge v_i(\lambda,w;\lambda_\infty)\ge\min_{\hat\lambda\ge\lambda_\infty}v_i(\lambda,w;\hat\lambda)=v_i^*(\lambda,w)$, where the first inequality holds because $0\ge v_{i-1}^*(\lambda,w)$. For $\lambda\le\lambda_\infty$, the last claim holds since $0\ge\hat c\lambda_\infty+u(\lambda_\infty,w)\ge\hat c\lambda_\infty+v_i(\lambda_\infty,w;\lambda_\infty)\ge\min_{\hat\lambda\ge\lambda_\infty}v_i(\lambda,w;\hat\lambda)=v_i^*(\lambda,w)$, where the first inequality follows since $w>w_0^*$ while the second follows from the definition of $v_i(\lambda,w;\lambda_\infty)$ and by virtue of the fact that $\bar u(\lambda,w)\ge\bar v_{i-1}^*(\lambda,w)$. □

Proof of Theorem 5. Similar to Lemma 3, it is sufficient to show that the statement is true when $w\in(w_{i-1}^*,w_i^*]$. By construction and by Proposition A.3(i), it is $(\mathcal{D}_0 v_i^*)(\lambda,w)=0$ for $\lambda>\lambda_i^*(w)$. Moreover, the fact that $\partial_1 v_i^*(\lambda,w)>-c/\delta_2$ implies that $(\mathcal{D}_\epsilon v_i^*)(\lambda,w)>0$ for $\epsilon>0$, and $\arg\min_{\hat a>0}\{v_i^*(\lambda+\delta_2\hat a,w)+c\hat a\}=\varnothing$. For $\lambda=\lambda_i^*(w)$, it is $\partial_1 v_i^*(\lambda_i^*(w),w)=-c/\delta_2$, and hence, $(\mathcal{D}_\epsilon v_i^*)(\lambda,w)=0$ for $\epsilon\ge0$. In addition, $\min_{\hat a>0}\{v_i^*(\lambda+\delta_2\hat a,w)+c\hat a\}$ does not have any solution. For $\lambda<\lambda_i^*(w)$,

$$a_i^*(\lambda,w)=(\lambda_i^*(w)-\lambda)/\delta_2$$
$$=\max\left\{a\ge0\colon a\in\arg\min_{\hat a>0}v_i^*(\lambda+\delta_2\hat a,w)+c\hat a\right\},$$

and the minimum value is equal to $v_i^*(\lambda,w)$. It only remains to be established that for $\lambda\in[0,\lambda_i^*(w)]$ it is

$$\frac{(\mathcal{D}_\epsilon v_i^*)(\lambda,w)}{\lambda+\rho}$$
$$=-[v_i^*(\lambda_i^*(w),w)+c/\delta_2(\lambda_i^*(w)-\lambda)]+\frac{\kappa(\lambda-\lambda_\infty)}{\lambda+\rho}\frac{c}{\delta_2}$$
$$+\frac{\lambda}{\lambda+\rho}\mathbb{E}[v_{i-1}^*(\lambda+\delta_{10}+\delta_{11}R,w(1-R))-wR]\ge0.$$

By Proposition A.3, equality is obtained at $\lambda_i^*(w)$; hence, it is sufficient to show that the above expression is decreasing with respect to λ. The latter holds because, by taking the derivative with respect to λ,

$$\frac{c}{\delta_2}+\frac{\kappa(\lambda_\infty+\rho)}{(\lambda+\rho)^2}\frac{c}{\delta_2}$$
$$+\frac{\rho}{(\lambda+\rho)^2}\mathbb{E}[v_{i-1}^*(\lambda+\delta_{10}+\delta_{11}R,w(1-R))-wR]$$
$$+\frac{\lambda}{\lambda+\rho}\mathbb{E}[\partial_1 v_{i-1}^*(\lambda+\delta_{10}+\delta_{11}R,w(1-R))]$$
$$=\frac{\kappa(\rho+\lambda_\infty)}{(\lambda+\rho)^2}f_i(\lambda,w)\le0,$$

which completes the proof. □

Proof of Corollary 5. The result can be found as in Corollary 1. □

Appendix B. Notation

Table B.1. Summary of Notation

Symbol	Description	Range
$\mathscr{A}$	Set of admissible collection strategies	$\mathbb{L}^\infty$
$\mathcal{A}^*$	Optimal action region	$\mathbb{R}_+^2$
$A(t)$	Bank's collection strategy	$\mathbb{R}_+$
$A^*(t)$	Optimal collection strategy	$\mathbb{R}_+$
$a^*(\lambda,w)$	Optimal discrete collection effort	$\mathbb{R}_+$
ΔA_k	kth discrete collection effort	$\mathbb{R}_+$
$\mathscr{B}$	Borel (sub-)set	$\mathcal{B}$

Table B.1. Continued

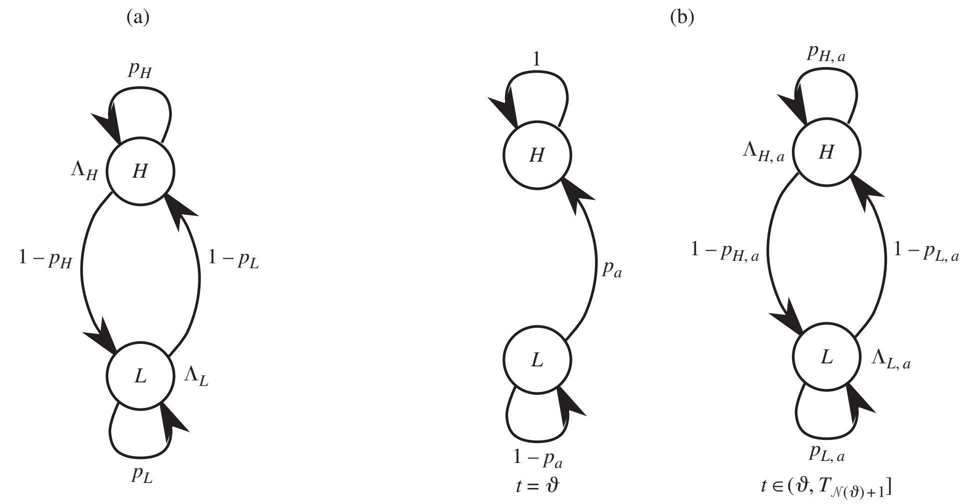
Symbol	Description	Range
c	Marginal cost of effort	$\mathbb{R}_{++}$
$\hat{c}$	Effective marginal cost of effort ($= c/\delta_2$)	$\mathbb{R}_+$
$(\mathcal{D}_\epsilon v)(\lambda, w)$	Integro-differential operator (at $\epsilon \geq 0$)	$\mathbb{R}$
$\mathbb{F} \triangleq \{\mathcal{F}_t : t \in \mathbb{R}_+\}$	Information filtration	—
$\mathcal{F}_t$	Available information at t	$\mathbb{F}$
F_R	Distribution of relative repayment R_i	$\mathbb{L}^\infty$
i	Index of repayment event	$\mathbb{N}$
$J(t) \triangleq [\mathcal{N}(t), \mathcal{R}(t)]^\top$	Repayment process	$\mathbb{N} \times \mathbb{R}_+$
$\mathcal{N}(t) \triangleq \sum_i \mathbf{1}_{\{T_i \leq t\}}$	Repayment counting process (willingness to repay)	$\mathbb{N}$
$\bar{r}$	Expected relative repayment ($\bar{r} \triangleq \mathbb{E}[R_i]$)	$[0,1]$
$R_i \triangleq Z_i / W(T_{i-1})$	Relative repayment at $t = T_i$	$[0,1]$
$\mathcal{R}(t) \triangleq \sum_i R_i \mathbf{1}_{\{T_i \leq t\}}$	Cumulative relative repayment process (ability to repay)	$\mathbb{R}_+$
s	Generic time index	$\mathbb{R}_+$
$(\mathcal{S}_{\hat{\lambda}} v)(\lambda, w)$	Sustained-extension operator (with holding intensity $\hat{\lambda}$)	$\mathbb{R}_+$
t	Time in the collection process	$\mathbb{R}_+$
T_i	Arrival time of ith repayment	$\mathbb{R}_+$
$u(\lambda, w)$	Autonomous account value	$\mathbb{R}_-$
$v(\lambda, w)$	Generic value function	$\mathbb{R}$
$v^*(\lambda, w)$	Optimal account value	$\mathbb{R}_-$
$\bar{V}^*(\lambda, w)$	Account's expected economic value	$\mathbb{R}_+$
w	Outstanding balance	$\mathbb{R}$
$\underline{w}$	Minimal actionable balance	$\mathbb{R}_+$
w_0^*	Economic balance threshold	$[\underline{w}, \infty)$
$W(t)$	Outstanding balance (at t)	$\mathbb{R}_+$
Z_i	Amount of ith repayment	$\mathbb{R}_+$
$(\alpha(t), \beta(t))$	Solution to the initial-value problem (15)–(16)	$\mathbb{R}_+^2$
$\delta_1 \triangleq [\delta_{11}, \delta_{12}]^\top$	Sensitivity of $\lambda(t)$ with respect to $J(t)$	$\mathbb{R}_+^2$
δ_2	Sensitivity of $\lambda(t)$ with respect to $A(t)$	$\mathbb{R}_+$
$\varepsilon^*(\lambda, w)$	Optimal infinitesimal collection effort	$\mathbb{R}_+$
$E^*(t) \triangleq \varepsilon^*(\lambda(t), W(t))$	Infinitesimal collection-effort trajectory	$\mathbb{R}_+$
ϑ_k	Time of kth discrete collection effort	$\mathbb{R}_+$
κ	Mean reversion rate of the intensity process	$\mathbb{R}_+$
λ	Repayment intensity	$\mathbb{R}_+$
λ_∞	Long-run steady-state intensity	$\mathbb{R}_+$
$\lambda(t)$	Intensity process	$\mathbb{R}_+$
$\nu^*(t) \triangleq v^*(\lambda(t), W(t))$	Value trajectory	$\mathbb{R}$
ρ	Discount factor	$\mathbb{R}_{++}$
$\varphi(s, \lambda(t))$	Flow of intensity (between jumps)	$\mathbb{R}_+$

Appendix C. Comparison with Hidden-Markov Model Variant

As noted in Section 2, we provide here a simple model for repayments with an unobservable arrival rate and show that in a Bayesian setting the dynamics of the repayment intensity, viewed as the expected value of the repayment arrival rate conditional on available information, is qualitatively equivalent to the intensity dynamics implied by the self-exciting point process with law of motion specified in Equation (3) of the main text.

Preliminaries. Consider a Markov jump process $M(t)$, defined for $t \geq 0$, whose states belong to the set $\mathcal{S} = \{L, H\}$. Given that $M(t) = H$ at the current time t, the next state-transition event (at a random transition time τ_H) arrives at the rate Λ_H. Upon arrival of a state transition at time $\tau_H \geq t$, the process M will *either* remain in state H (with probability p_H) *or* move to state L (with probability $1 - p_H$); see Figure C.1(a). Similarly, if $M(t) = L$, the next transition arrives at the rate Λ_L, and upon arrival of such a transition at τ_L, M will *either* remain in state L (with probability p_L) *or* move to state H (with probability $1 - p_L$). The Markov jump process M is not observable, yet its value affects the arrival rate of an observable Poisson process $(T_i, i \geq 1)$. If M is in state H (resp., L), the unobservable arrival rate $\tilde{\lambda}(t)$ of the observable Poisson process is equal to λ_H (resp., $\tilde{\lambda}(t) = \lambda_L$), where—without any loss of generality—we assume throughout that $\lambda_L < \lambda_H$. Let $\mathbb{F} = (\mathcal{F}_t, t \geq 0)$ denote the filtration generated by $(T_i, i \geq 1)$ and $\mathcal{N}(t) = \sum_i \mathbf{1}_{\{T_i \leq t\}}$ the corresponding counting process. Our goal is to characterize the dynamics of $\Pi_H(t) \triangleq \mathbb{P}(M(t) = H \mid \mathcal{F}_t)$, that is, the probability of M being in state H, which, in turn, allows us to determine the expected arrival intensity $\lambda(t) \triangleq \mathbb{E}[\tilde{\lambda}(t) \mid \mathcal{F}_t]$. Bayraktar and Ludkovski (2009) carry out this task for a general Markov process with the aid of infinitesimal generators. Here, we derive the corresponding results in a more elementary way by taking advantage of the fact that the state space $\mathcal{S}$ is binary.

Figure C.1. (a) Binary Markov Jump Process; (b) Controlled Markov Jump Process



Proposition C.1. *The probability $\Pi_H(t)$ satisfies the ordinary differential equation*

$$\dot{\Pi}_H(t) = (1-p_L)\Lambda_L(1-\Pi_H(t)) - (1-p_H)\Lambda_H\Pi_H(t) - (\lambda_H-\lambda_L)(1-\Pi_H(t))\Pi_H(t) \tag{C.1}$$

for all $t \geq 0$ with $t \neq T_{\mathcal{N}(t)}$, that is, between any two consecutive observable events $(T_i, i \geq 1)$.

Proof. Between any two observable events, that is, for $t \neq T_{\mathcal{N}(t)}$, the probability

$$\Pi_H(t) = \mathbb{P}(M(t)=H \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1} > t)$$

can be decomposed using the law of total probability and is equal to

$$\begin{aligned}
&\mathbb{P}(M(t)=H, M(t-\epsilon)=H \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\quad+\mathbb{P}(M(t)=H, M(t-\epsilon)=L \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&=\mathbb{P}(M(t)=H \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\quad\cdot\mathbb{P}(M(t-\epsilon)=H \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\quad+\mathbb{P}(M(t)=H \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\quad\cdot\mathbb{P}(M(t-\epsilon)=L \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t).
\end{aligned}$$

Hence,

$$\begin{aligned}
\dot{\Pi}_H(t) &= \lim_{\epsilon\to 0^+}\frac{\Pi_H(t)-\Pi_H(t-\epsilon)}{\epsilon}\\
&=\lim_{\epsilon\to 0^+}\Big\{\frac{1}{\epsilon}\big[\mathbb{P}(M(t)=H \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\qquad-\Pi_H(t-\epsilon)\big]\mathbb{P}(M(t-\epsilon)=H \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\quad+\frac{1}{\epsilon}\big[\mathbb{P}(M(t)=H \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\qquad-\Pi_H(t-\epsilon)\big]\mathbb{P}(M(t-\epsilon)=L \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\Big\}.
\end{aligned} \tag{C.2}$$

Using Bayes' rule, it is

$$\begin{aligned}
&\mathbb{P}(M(t-\epsilon)=H \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&=\mathbb{P}(M(t-\epsilon)=H, T_{\mathcal{N}(t)+1}>t \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad\Big/\big(\mathbb{P}(M(t-\epsilon)=H, T_{\mathcal{N}(t)+1}>t \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad+\mathbb{P}(M(t-\epsilon)=L, T_{\mathcal{N}(t)+1}>t \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\big)\\
&=\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\Pi_H(t-\epsilon)\\
&\quad\Big/\big(\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\Pi_H(t-\epsilon)\\
&\quad+\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad\cdot(1-\Pi_H(t-\epsilon))\big),
\end{aligned}$$

where

$$\begin{aligned}
&\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad=1-\lambda_H\epsilon+o(\epsilon^2),\\
&\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad=1-\lambda_L\epsilon+o(\epsilon^2);
\end{aligned}$$

see Equations (C.3)–(C.6) for details. Consequently,

$$\begin{aligned}
&\mathbb{P}(M(t-\epsilon)=H \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&=\frac{(1-\lambda_H\epsilon+o(\epsilon^2))\Pi_H(t-\epsilon)}{(1-\lambda_H\epsilon+o(\epsilon^2))\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon+o(\epsilon^2))(1-\Pi_H(t-\epsilon))}\\
&\approx\frac{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)}{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M(t-\epsilon)=L \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&=1-\mathbb{P}(M(t-\epsilon)=H \mid \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&=\frac{(1-\lambda_L\epsilon+o(\epsilon^2))(1-\Pi_H(t-\epsilon))}{(1-\lambda_H\epsilon+o(\epsilon^2))\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon+o(\epsilon^2))(1-\Pi_H(t-\epsilon))}\\
&\approx\frac{(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}.
\end{aligned}$$

Similarly, it is

$$\begin{aligned}
&\mathbb{P}(M(t)=H \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&=\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad\Big/\big(\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad+\mathbb{P}(M(t)=L, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\big),
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M(t)=H \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&=\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad \big/ \big(\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad +\mathbb{P}(M(t)=L, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\big).
\end{aligned}$$

Note that

$$\begin{aligned}
&\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t)=H, M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad \cdot \mathbb{P}(M(t)=H \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=(1-\lambda_H\epsilon+o(\epsilon^2))(1-\Lambda_H\epsilon+p_H\Lambda_H\epsilon+o(\epsilon^2))\\
&\approx 1-\lambda_H\epsilon-(1-p_H)\Lambda_H\epsilon, \qquad \text{(C.3)}
\end{aligned}$$

since the chance of M switching its state within $(t-\epsilon, t)$ more than once is of the order of ϵ^2. Similarly,

$$\begin{aligned}
&\mathbb{P}(M(t)=L, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t)=L, M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad \cdot \mathbb{P}(M(t)=L \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=(1+o(\epsilon))((1-p_H)\Lambda_H\epsilon+o(\epsilon^2))\approx(1-p_H)\Lambda_H\epsilon, \qquad \text{(C.4)}
\end{aligned}$$

where the chance of having more than one event for M, irrespective of whether it leads to a change in the state, is of the order of ϵ^2. Repeating the above argument produces

$$\begin{aligned}
&\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t)=H, M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad \cdot \mathbb{P}(M(t)=H \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=(1+o(\epsilon))((1-p_L)\Lambda_L\epsilon+o(\epsilon^2))\approx(1-p_L)\Lambda_L\epsilon. \qquad \text{(C.5)}
\end{aligned}$$

Finally,

$$\begin{aligned}
&\mathbb{P}(M(t)=L, T_{\mathcal{N}(t)+1}>t \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=\mathbb{P}(T_{\mathcal{N}(t)+1}>t \mid M(t)=L, M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&\quad \cdot \mathbb{P}(M(t)=L \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t-\epsilon)\\
&=(1-\lambda_L\epsilon+o(\epsilon^2))(1-\Lambda_L\epsilon+p_L\Lambda_L\epsilon+o(\epsilon^2))\\
&\approx 1-\lambda_L\epsilon-(1-p_L)\Lambda_L\epsilon, \qquad \text{(C.6)}
\end{aligned}$$

where again, the likelihood of having M switching its state within $(t-\epsilon, t)$ more than once is of the order of ϵ^2. Combining Equations (C.3)–(C.6), one obtains

$$\begin{aligned}
&\mathbb{P}(M(t)=H \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\approx \frac{1-\lambda_H\epsilon-(1-p_H)\Lambda_H\epsilon}{1-\lambda_H\epsilon-(1-p_H)\Lambda_H\epsilon+(1-p_H)\Lambda_H\epsilon}\\
&=\frac{1-\lambda_H\epsilon-(1-p_H)\Lambda_H\epsilon}{1-\lambda_H\epsilon},
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M(t)=H \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)}}, T_{\mathcal{N}(t)+1}>t)\\
&\approx \frac{(1-p_L)\Lambda_L\epsilon}{(1-p_L)\Lambda_L\epsilon+1-\lambda_L\epsilon-(1-p_L)\Lambda_L\epsilon}\\
&=\frac{(1-p_L)\Lambda_L\epsilon}{1-\lambda_L\epsilon}.
\end{aligned}$$

Finally, substituting the above in Equation (C.2) gives

$$\begin{aligned}
&\lim_{\epsilon\to 0^+}\Bigg\{\frac{1}{\epsilon}\left[\frac{1-\lambda_H\epsilon-(1-p_H)\Lambda_H\epsilon}{1-\lambda_H\epsilon}-\Pi_H(t-\epsilon)\right]\\
&\qquad\cdot\frac{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)}{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}\\
&\qquad+\frac{1}{\epsilon}\left[\frac{(1-p_L)\Lambda_L\epsilon}{1-\lambda_L\epsilon}-\Pi_H(t-\epsilon)\right]\\
&\qquad\cdot\frac{(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}\Bigg\}\\
&=\lim_{\epsilon\to 0^+}\Bigg\{\frac{1}{\epsilon}\Big[(1-\lambda_H\epsilon)(1-\Pi_H(t-\epsilon))-(1-p_H)\Lambda_H\epsilon\Big]\\
&\qquad\cdot\frac{\Pi_H(t-\epsilon)}{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}\\
&\qquad+\frac{1}{\epsilon}\Big[(1-p_L)\Lambda_L\epsilon-(1-\lambda_L\epsilon)\Pi_H(t-\epsilon)\Big]\\
&\qquad\cdot\frac{(1-\Pi_H(t-\epsilon))}{(1-\lambda_H\epsilon)\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))}\Bigg\}\\
&=\lim_{\epsilon\to 0^+}\big((\lambda_L-\lambda_H)(1-\Pi_H(t-\epsilon))\Pi_H(t-\epsilon)\\
&\qquad+(1-p_L)\Lambda_L(1-\Pi_H(t-\epsilon))-(1-p_H)\Lambda_H\Pi_H(t-\epsilon)\big)\\
&\qquad\big/\big((1-\lambda_H\epsilon)\Pi_H(t-\epsilon)+(1-\lambda_L\epsilon)(1-\Pi_H(t-\epsilon))\big),
\end{aligned}$$

which leads to the ordinary differential equation (C.1) as claimed. □

Proposition C.2. *Between any two consecutive observable events T_i and T_{i+1}, the probability $\Pi_H(t)$ is given by*

$$\begin{aligned}
\Pi_H(t) = {}& \pi_{H,2}-(\pi_{H,2}-\pi_{H,1})\\
&\Bigg/\left(1+\frac{\Pi_H(T_i)-\pi_{H,1}}{\pi_{H,2}-\Pi_H(T_i)}\exp[-(\pi_{H,2}-\pi_{H,1})(t-T_i)]\right),\\
&\qquad t\in[T_i, T_{i+1}), \qquad \text{(C.7)}
\end{aligned}$$

where $\Pi_H(T_i)$ is the initial value for $t=T_i$ at the beginning of the interarrival interval; furthermore,

$$0\le\pi_{H,1}\triangleq\frac{\gamma_1-\gamma_2}{2(\lambda_H-\lambda_L)}\le 1\le\pi_{H,2}\triangleq\frac{\gamma_1+\gamma_2}{2(\lambda_H-\lambda_L)}, \qquad \text{(C.8)}$$

with

$$\begin{aligned}
\gamma_1 &\triangleq(\lambda_H-\lambda_L)+(1-p_H)\Lambda_H+(1-p_L)\Lambda_L,\\
\gamma_2 &\triangleq\sqrt{\gamma_1^2-4(\lambda_H-\lambda_L)\Lambda_L(1-p_L)}.
\end{aligned}$$

Proof. By Proposition C.1, the probability Π_H satisfies Equation (C.1) for all $t\in(T_i, T_{i+1})$. The equilibrium points $\pi_{H,1}$ and $\pi_{H,2}$ are obtained by setting the right-hand side to zero. It is $0\le\pi_{H,1}\le 1\le\pi_{H,2}$ since the function $f(x)=(\lambda_H-\lambda_L)x^2-[(\lambda_H-\lambda_L)+(1-p_H)\Lambda_H+(1-p_L)\Lambda_L]x+(1-p_L)\Lambda_L$ is convex, positive at $x=0$, and negative at $x=1$. Equation (C.1) is a Riccati differential equation and can be solved using standard methods (see, e.g., Weber 2011, Chapter 2), which—by taking into account the initial value $\Pi_H(T_i)$—yields Equation (C.7). □

Remark C.1. If $\Pi_H(T_i)>\pi_{H,1}$, then Π_H is strictly decreasing whereas for $\Pi_H(T_i)<\pi_{H,1}$ it is strictly increasing. Finally, if $\Pi_H(0)=\pi_{H,1}$, then $\Pi_H(t)\equiv\pi_{H,1}$ on (T_i, T_{i+1}) since then the system is in equilibrium (as $\dot{\Pi}_H(t)\equiv 0$).

Proposition C.3. *At any arrival time $t = T_{\mathcal{N}(t)}$, it is $\Pi_H(t) > \Pi_H(t^-)$, that is, the probability Π_H experiences a positive jump regardless of the sign of $\Pi_H(t^-) - \pi_{H,1}$.*

Proof. At any $t = T_{\mathcal{N}}(t)$, we have

$$\begin{aligned}\Pi_H(t) &= \mathbb{P}(M(t) = H \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, T_{\mathcal{N}(t)} = t)\\ &= \lim_{\epsilon\to 0^+} \mathbb{P}(M(t) = H \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)} \le t)\\ &= \lim_{\epsilon\to 0^+} \frac{\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})}{\mathbb{P}(T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})}.\end{aligned}$$

The numerator in the preceding expression can be rewritten as

$$\begin{aligned}&\mathbb{P}(M(t)=H, M(t-\epsilon)=H, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad + \mathbb{P}(M(t)=H, M(t-\epsilon)=L, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\end{aligned}$$

while the denominator takes the form

$$\begin{aligned}&\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad + \mathbb{P}(M(t)=L, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)}),\end{aligned}$$

where

$$\begin{aligned}&\mathbb{P}(M(t)=L, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &= \mathbb{P}(M(t)=L, M(t-\epsilon)=H, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad + \mathbb{P}(M(t)=L, M(t-\epsilon)=L, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)}).\end{aligned}$$

Similar to Equations (C.3)–(C.6), it is

$$\begin{aligned}&\mathbb{P}(M(t)=H, M(t-\epsilon)=H, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &= \mathbb{P}(T_{\mathcal{N}(t)} \le t \mid M(t)=H, M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad \cdot \mathbb{P}(M(t)=H \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\Pi_H(t-\epsilon)\\ &= (\lambda_H\epsilon + o(\epsilon^2))(1-\Lambda_H\epsilon + p_H\Lambda_H\epsilon + o(\epsilon^2))\Pi_H(t-\epsilon)\\ &\approx \lambda_H\epsilon\Pi_H(t-\epsilon),\\ &\mathbb{P}(M(t)=H, M(t-\epsilon)=L, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &= \mathbb{P}(T_{\mathcal{N}(t)} \le t \mid M(t)=H, M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad \cdot \mathbb{P}(M(t)=H \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad \cdot (1-\Pi_H(t-\epsilon))\\ &= o(\epsilon)((1-p_L)\Lambda_L\epsilon + o(\epsilon^2))(1-\Pi_H(t-\epsilon)) \approx 0,\\ &\mathbb{P}(M(t)=L, M(t-\epsilon)=H, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &= \mathbb{P}(T_{\mathcal{N}(t)} \le t \mid M(t)=L, M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad \cdot \mathbb{P}(M(t)=L \mid M(t-\epsilon)=H, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\Pi_H(t-\epsilon)\\ &= o(\epsilon)((1-p_H)\Lambda_H\epsilon + o(\epsilon^2))\Pi_H(t-\epsilon) \approx 0,\end{aligned}$$

and

$$\begin{aligned}&\mathbb{P}(M(t)=L, M(t-\epsilon)=L, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &= \mathbb{P}(T_{\mathcal{N}(t)} \le t \mid M(t)=L, M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad \cdot \mathbb{P}(M(t)=L \mid M(t-\epsilon)=L, \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})\\ &\quad \cdot (1-\Pi_H(t-\epsilon))\\ &= (\lambda_L\epsilon + o(\epsilon^2))(1-\Lambda_L\epsilon + p_L\Lambda_L\epsilon + o(\epsilon^2))(1-\Pi_H(t-\epsilon))\\ &\approx \lambda_L\epsilon(1-\Pi_H(t-\epsilon)).\end{aligned}$$

Consequently, one obtains

$$\begin{aligned}\Pi_H(t) &= \lim_{\epsilon\to 0^+} \frac{\mathbb{P}(M(t)=H, T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})}{\mathbb{P}(T_{\mathcal{N}(t)} \le t \mid \mathcal{F}_{T_{\mathcal{N}(t)-1}}, t-\epsilon < T_{\mathcal{N}(t)})}\\ &= \lim_{\epsilon\to 0^+} \frac{\lambda_H\epsilon\Pi_H(t-\epsilon)}{\lambda_H\epsilon\Pi_H(t-\epsilon) + \lambda_L\epsilon(1-\Pi_H(t-\epsilon))}\\ &= \frac{\lambda_H\Pi_H(t^-)}{\lambda_H\Pi_H(t^-) + \lambda_L(1-\Pi_H(t^-))} > \Pi_H(t^-),\end{aligned}$$

as claimed. □

Autonomous Intensity Dynamics. We are now ready to consider the law of motion for the conditional expectation $\lambda(t) = \mathbb{E}[\tilde{\lambda}(t) \mid \mathcal{F}_t]$ introduced earlier in the absence of any account-treatment action. Given a nondegenerate parameter vector $(p_L, p_H, \Lambda_L, \Lambda_H, \lambda_L, \lambda_H) \gg 0$, we interpret the sequence of observable random stopping times $(T_i, i \ge 1)$ as the account holder's repayment events, whereby each repayment is drawn from a certain empirically identified repayment distribution, similar to the assumptions in the main text. In this context, the hidden Markov jump process M describes the account holder's unobservable repayment priority (with H for "high priority" and L for "low priority"). The resulting law of motion,

$$\lambda(t) = \lambda_H\Pi_H(t) + \lambda_L(1-\Pi_H(t)) = \lambda_L + (\lambda_H - \lambda_L)\Pi_H(t),$$

closely resembles the autonomous intensity dynamics obtained in Section 2.1 in the sense that it features positive jumps at repayment events and a smooth decrease toward a long-run stationary value ($\lambda_\infty = \lambda_L + (\lambda_H - \lambda_L)\pi_{H,1}$) in the prolonged absence of repayment events; see Figures 2(b) and C.2(a).

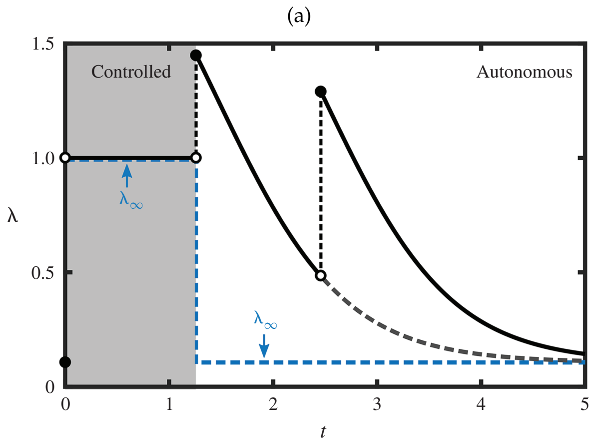
Controlled Intensity Dynamics. As discussed in the main text, an account-treatment strategy consists of a sequence of actions. We denote the starting time of a collection action by ϑ and assume that the collection action continues until a repayment is received, that is, until $T_{\mathcal{N}(\vartheta)+1}$. The first measure taken to implement a suitable collection action often takes the account holder by surprise and can immediately change the account holder's priorities, so that M moves from "low" to "high." Consequently, we assume that if $M(\vartheta) = L$ at time $t = \vartheta$, the state of the hidden Markov jump process can change to H with probability p_a. If this transition does not take place at $t = \vartheta$, it will take place at a random time τ_a with arrival intensity $(1-p_{L,a})\Lambda_{L,a}$ as long as the collection action remains in effect (i.e., over $[\vartheta, T_{\mathcal{N}(\vartheta)+1}]$). Upon moving to state H, the Markov jump process M evolution is governed by $\Lambda_{H,a}$ and $p_{H,a}$, which have the same interpretation as Λ_H and p_H before. Figure C.1(b) illustrates these dynamics. Note that at $t = \vartheta$,

$$\begin{aligned}\Pi_H(\vartheta^+) &= \lim_{\epsilon\to 0^+} \mathbb{P}(M(\vartheta+\epsilon)=H \mid \mathcal{F}_\vartheta, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon)\\ &= \lim_{\epsilon\to 0^+} \mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=H \mid \mathcal{F}_\vartheta, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon)\\ &\quad + \lim_{\epsilon\to 0^+} \mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=L \mid \mathcal{F}_\vartheta, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon).\end{aligned}$$

In particular,

$$\begin{aligned}&\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=H \mid \mathcal{F}_\vartheta, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon)\\ &\quad = \frac{\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=H, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)}{\mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)},\end{aligned}$$

Figure C.2. Numerical Illustration: (a) Bayesian Intensity Dynamics; (b) Autonomous and Controlled Parameters

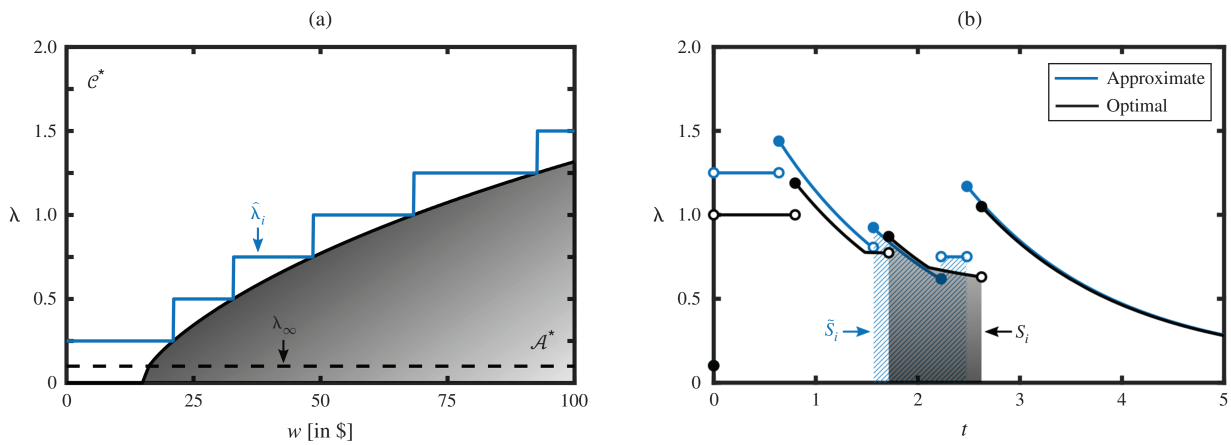


(b)

	p_L	p_H	Λ_L	Λ_H	p_a	λ_L	λ_H
Autonomous	0.95	0.10	0.20	1.00	—	0.10	1.50
Controlled	0.10	0.50	3.00	2.00	0.65		

	$\pi_{H,1}$	$\pi_{H,2}$	λ_∞
Autonomous	0.004	1.646	0.106
Controlled	0.643	3.000	1.000

Figure D.1. Approximation of the Optimal Discrete Collection Strategy $A^*_{\mathscr{M}}$ by an $\mathscr{M}$-Envelope Strategy $\tilde{A}^*_{\mathscr{M}}$: (a) Upper $\mathscr{M}$-Envelope Approximation Scheme Based on A^* in the (Intensity, Balance) Space; (b) Intensity Dynamics Obtained from $A^*_{\mathscr{M}}$ and $\tilde{A}^*_{\mathscr{M}}$, respectively (Where $\tilde{S}_i = S_i$ Are i.i.d. Exponential)



and

$$\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=L \mid \mathcal{F}_\vartheta, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon) = \frac{\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=L, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)}{\mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)},$$

where

$$\begin{aligned}
&\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=H, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)\\
&\quad= \mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid M(\vartheta+\epsilon)=H, M(\vartheta)=H, \mathcal{F}_\vartheta)\\
&\qquad \cdot \mathbb{P}(M(\vartheta+\epsilon)=H \mid M(\vartheta)=H, \mathcal{F}_\vartheta)\Pi_H(\vartheta)\\
&\quad= (1+o(\epsilon))(1+o(\epsilon))\Pi_H(\vartheta) \approx \Pi_H(\vartheta), \qquad \text{(C.9)}\\
&\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=L, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)\\
&\quad= \mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid M(\vartheta+\epsilon)=H, M(\vartheta)=L, \mathcal{F}_\vartheta)\\
&\qquad \cdot \mathbb{P}(M(\vartheta+\epsilon)=H \mid M(\vartheta)=L, \mathcal{F}_\vartheta)(1-\Pi_H(\vartheta))\\
&\quad= (1+o(\epsilon))(p_a+o(\epsilon))(1-\Pi_H(\vartheta)) \approx p_a(1-\Pi_H(\vartheta)), \qquad \text{(C.10)}\\
&\mathbb{P}(M(\vartheta+\epsilon)=L, M(\vartheta)=L, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)\\
&\quad= \mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid M(\vartheta+\epsilon)=L, M(\vartheta)=L, \mathcal{F}_\vartheta)\\
&\qquad \cdot \mathbb{P}(M(\vartheta+\epsilon)=L \mid M(\vartheta)=L, \mathcal{F}_\vartheta)(1-\Pi_H(\vartheta))\\
&\quad= (1+o(\epsilon))(1-p_a+o(\epsilon))(1-\Pi_H(\vartheta))\\
&\quad\approx (1-p_a)(1-\Pi_H(\vartheta)), \qquad \text{(C.11)}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{P}(M(\vartheta+\epsilon)=L, M(\vartheta)=H, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)\\
&= \mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid M(\vartheta+\epsilon)=L, M(\vartheta)=H, \mathcal{F}_\vartheta)\\
&\quad \cdot \mathbb{P}(M(\vartheta+\epsilon)=L \mid M(\vartheta)=H, \mathcal{F}_\vartheta)\Pi_H(\vartheta)\\
&= (1+o(\epsilon))o(\epsilon)\Pi_H(\vartheta) \approx 0. \qquad \text{(C.12)}
\end{aligned}$$

Combining Equations (C.9)–(C.12) gives

$$\begin{aligned}
&\lim_{\epsilon\to 0^+} \mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=H \mid \mathcal{F}_\vartheta, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon)\\
&\quad= \lim_{\epsilon\to 0^+} \frac{\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=H, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)}{\mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)}\\
&\quad= \Pi_H(\vartheta),
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{\epsilon\to 0^+} \mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=L \mid \mathcal{F}_\vartheta, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon)\\
&\quad= \lim_{\epsilon\to 0^+} \frac{\mathbb{P}(M(\vartheta+\epsilon)=H, M(\vartheta)=L, T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)}{\mathbb{P}(T_{\mathcal{N}(\vartheta)+1} > \vartheta+\epsilon \mid \mathcal{F}_\vartheta)}\\
&\quad= p_a(1-\Pi_H(\vartheta)).
\end{aligned}$$

Hence, for any $p_a > 0$, $\Pi_H(t)$ and, in turn, also $\lambda(t) = \mathbb{E}[\tilde{\lambda}(t) \mid \mathcal{F}_t]$ experience a positive jump. Choosing $(1-p_{L,a})\Lambda_{L,a}$ and

$(1 - p_{H,a})\Lambda_{H,a}$ such that $\pi_{H,1} = \Pi_H(\vartheta) + p_a(1 - \Pi_H(\vartheta))$, with the parameter vector $(p_L, p_H, \Lambda_L, \Lambda_H)$ replaced by $(p_{H,a}, p_{L,a}, \Lambda_{L,a}, \Lambda_{H,a})$ in Equation (C.8), yields dynamics that are qualitatively equivalent to the intensity dynamics in the main text.

Figure C.2(a) shows the corresponding dynamics for an account whose autonomous intensity dynamics evolve according to the parameter vector in the table of Figure C.2(b); that is with $(p_L, p_H, \Lambda_L, \Lambda_H) = (0.95, 0.10, 0.20, 1.00)$. When the repayment priority is "high," that is, $M(t) = H$, on average 1.5 repayments are received per time unit (quarter). When this priority is "low," that is, $M(t) = L$, the repayment arrival rate is 0.1 per quarter. In the absence of account treatment, this leads to $\pi_{H,1} = 0.004$ and $\pi_{H,2} = 1.646$ determining the intensity's long-run stationary value, $\lambda_\infty = 0.106$. Figure C.2(a) depicts a case where the initial intensity level, at $t = 0$, is λ_∞. As in the main text, assuming that the outstanding balance is large enough so $(\lambda_\infty, w) \in \mathcal{A}^*$, one can increase and maintain the intensity by taking a suitable account-treatment action. An action, taken at $\vartheta = 0$ with parameter vector $(p_{L,a}, p_{H,a}, \Lambda_{L,a}, \Lambda_{H,a}, p_a) = (0.10, 0.50, 3.00, 2.00, 0.65)$, moves the repayment intensity (via a jump caused by p_a) to a new long-run stationary value, $\lambda_\infty = 1.000$, compatible with $\pi_{H,1} = 0.643$ and $\pi_{H,2} = 3.000$. The action remains in effect until a repayment is received. During this period, the intensity level remains constant since the system is held in equilibrium. Upon receiving a repayment, the intensity experiences a positive jump as described before. For the case illustrated in Figure C.2(a), there are no subsequent actions. Consequently, the intensity reverts back to its autonomous long-run stationary value, $\lambda_\infty = 0.106$, unless a further repayment is received (e.g., at $T_2 = 2.5$).

Appendix D. Implementation Details

To aid in the practical implementation of the optimal collection strategy discussed in the main text, we now provide a brief discussion of model identification and algorithm for a simple collection strategy constructed from the optimal solution to (P), provided there are only a finite number of (empirically identified) collection actions and attainable intensity levels.

Model Identification. Let $y \in \mathbb{R}^n$ be a vector of account-specific characteristics, such as the predefault FICO score and outstanding balance at the point of collections placement. To identify the account-specific parameter vector $(\kappa(y), \delta_{10}(y), \delta_{11}(y), \delta_2(y))$, one can use the estimation methods developed by Chehrazi and Weber (2015) given observations of a portfolio of delinquent accounts with respect to collection activity and repayment events over the time interval $[0, h]$ for some $h > 0$ (usually in the order of 6–24 months). For each account (the index of which is omitted), the set of observed data usually includes the account-specific attribute vector y, the repayment events $(T_i, R_i)_{i=1}^{\mathcal{N}(h)}$, and the collection activity $(\vartheta_k, m_k)_{k=1}^{\mathcal{K}(h)}$ in terms of timing and type of action on $[0, h]$. At time ϑ_k the kth collection action of type m_k (from an available menu of actions $\mathcal{M} \triangleq \{1, \ldots, \bar{m}\}$) is taken, for example, in the form of a legal action or the establishment of a repayment plan with a certain number of installments. The number of collection actions up to but not including time $t \in [0, h]$ is denoted by $\mathcal{K}(t) = \sum_{k\ge1} \mathbf{1}_{\{\vartheta_k < t\}}$. The bank observes the starting time of an action, its termination time (which may coincide with a repayment or a starting time of a subsequent action), and also all specific measures taken to implement the action (which together determine the action type). In addition to the account-specific parameter vector $(\kappa(y), \delta_{10}(y), \delta_{11}(y), \delta_2(y))$, it is necessary to estimate the common long-run steady-state intensity λ_∞ (which can be set to zero for conservative/robust estimates; see Chehrazi and Weber 2015), and the vector of action-specific intensities $\hat{\lambda}_{\mathcal{M}} \triangleq (\hat{\lambda}_m)_{m\in\mathcal{M}}$. Each action-specific intensity is scaled by $\delta_2(y)$ to pin down its account-specific impact, that is, the account's responsiveness to the relevant type of collection action. The log-likelihood of observing the repayment times $(T_i)_{i=1}^{\mathcal{N}(h)}$ given $(R_i)_{i=1}^{\mathcal{N}(h)}$, and $(\vartheta_k, m_k)_{k=1}^{\mathcal{K}(h)}$ as a function of $(\kappa(y), \delta_{10}(y), \delta_{11}(y), \delta_2(y), \lambda_\infty, \hat{\lambda}_{\mathcal{M}})$ is

$$L((T_i)_{i=1}^{\mathcal{N}(h)} \mid (R_i)_{i=1}^{\mathcal{N}(h)}, (\vartheta_k, m_k)_{k=1}^{\mathcal{K}(h)}) = \sum_{i=1}^{\mathcal{N}(h)} \ln(\lambda(T_i^-)) - \int_0^h \lambda(s)\, ds,$$

where

$$\lambda(t) = \begin{cases} \lambda_\infty, \quad t = 0, \\ \lambda(T_i^-) + \delta_{10}(y) + \delta_{11}(y) R_i, \quad t = T_i, \\ \lambda(\vartheta_k^-), \quad t = \vartheta_k, \\ \lambda_\infty + (\lambda(T_i) - \lambda_\infty)\exp(-\kappa(y)(t - T_i)), \\ \quad t \in (T_i, T_{i+1} \wedge \vartheta_{\mathcal{K}(T_i)+1}), \\ \hat{\lambda}_{m_k}\delta_2(y), \\ \quad t \in (\vartheta_k, T_{\mathcal{N}(\vartheta_k)+1} \wedge \vartheta_{k+1}) \text{ and } \lambda(\vartheta_k) \le \hat{\lambda}_{m_k}\delta_2(y), \\ \hat{\lambda}_{m_k}\delta_2(y) + (\lambda(T_i) - \hat{\lambda}_{m_k}\delta_2(y))\exp(-\kappa(y)(t - T_i)), \\ \quad t \in (\vartheta_k, T_{\mathcal{N}(\vartheta_k)+1} \wedge \vartheta_{k+1}) \text{ and } \lambda(\vartheta_k) > \hat{\lambda}_{m_k}\delta_2(y). \end{cases}$$

As in lattice theory, the binary $\wedge$-operator denotes the minimum. Assuming an affine dependence of the parameters κ, δ_{10}, δ_{11}, δ_2 on the vector of account-specific characteristics y (as in Chehrazi and Weber 2015), maximum-likelihood estimates can be found by maximizing the sum of log-likelihoods over the observed account portfolio subject to feasibility.

Approximately Optimal Collections. Upon identifying $\hat{\lambda}_{\mathcal{M}}$ together with the rest of the model parameters, the solution A^* to the collection problem (P) in the main text, without restrictions of intensity levels, can be used to obtain an approximately optimal solution to the discrete version of the collection problem where controlled intensity levels are restricted to components of the vector $\hat{\lambda}_{\mathcal{M}}$. It turns out that a direct solution $A^*_{\mathcal{M}}$ of the discrete collection problem leads to a nonconvex action region $\mathcal{A}^*_{\mathcal{M}}$ and, thus, to analytical complexities that are beyond the scope of this paper. Instead of dealing with the additional details, we provide here a simple method for enveloping A^* (i.e., the solution to (P)) by a collection strategy $\tilde{A}^*_{\mathcal{M}}$ and show numerically, by example, that the resulting collection performance tends to be close to the collection performance of the optimal discrete collection strategy $A^*_{\mathcal{M}}$ despite a significant optimality gap between A^* and $A^*_{\mathcal{M}}$ because of the discretization. For this, we first define the (upper) $\mathcal{M}$-envelope of the optimal action set $\mathcal{A}^*$ as

$$\tilde{\lambda}^*_{\mathcal{M}}(w) \triangleq \hat{\lambda}_{m_{\mathcal{M}}(w)},$$

where $m_{\mathcal{M}}(w) \triangleq \min\{m \in \mathcal{M}\colon \lambda^*(w) \le \hat{\lambda}_m \text{ or } m = \bar{m}\}$.[24] The $\mathcal{M}$-envelope collection strategy $\tilde{A}^*_{\mathcal{M}}$ is then defined as a collection strategy that results from taking an action if and only if, at the current time $t \geq 0$, the account state lies in the closure of the action set $\mathcal{A}^*$, that is, if and only if $(\lambda(t), W(t)) \in \mathrm{cl}\,\mathcal{A}^*$. Any action is such that it pushes the repayment intensity from $\lambda(t)$ to the upper $\mathcal{M}$-envelope $\tilde{\lambda}^*_{\mathcal{M}}(w)$ using a discrete effort and then maintains that intensity level until the next repayment using a continuous effort.

As an example, consider a situation where $\bar{m} = 6$ discrete intensity levels are available and assume that $\hat{\lambda}_{\mathcal{M}} = (m/4)_{m\in\mathcal{M}} = (0.25, 0.5, \ldots, 1.5)$. To allow for comparison, we use the same set of parameters as in the main text, that is, $(\lambda_\infty, \kappa, \delta_{10}, \delta_{11}, \delta_2, \rho, \underline{r}, c) = (0.1, 0.7, 0.02, 0.5, 1, 6\%, 0.1, \$6)$. Figure D.1 illustrates the construction of the $\mathcal{M}$-envelope collection strategy $\tilde{A}^*_{\mathcal{M}}$, which effectively rounds up the optimal intensity level under the original strategy (which solves (P)) as long as it is feasible to do so. This figure also depicts the direct solution $A^*_{\mathcal{M}}$ of the discrete collection problem for the same sample path (with perfect coupling; Thorisson 2000). Denoting the value generated by this collection policy by $\tilde{v}_{\mathcal{M}}(\lambda, w)$ and the (numerical) value of the optimal discrete policy by $v^*_{\mathcal{M}}(\lambda, w)$, the relative error is $\tilde{e}_{\mathcal{M}}(\lambda, w) \triangleq |(\tilde{v}_{\mathcal{M}}(\lambda, w) - v^*_{\mathcal{M}}(\lambda, w))/v^*_{\mathcal{M}}(\lambda, w)|$. Given an initial intensity $\lambda = \lambda_\infty$ and initial outstanding balance $w = \$75$, the example yields $\tilde{e}_{\mathcal{M}}(\lambda, w) = 0.14\%$. Meanwhile, the optimality gap caused by discrete action space is $g_{\mathcal{M}}(\lambda, w) \triangleq |(v^*_{\mathcal{M}}(\lambda, w) - v^*(\lambda, w))/v^*(\lambda, w)|$, whence, in the example, $g_{\mathcal{M}}(\lambda, w) = 6.8\% \gg \tilde{e}_{\mathcal{M}}(\lambda, w)$.

Continuous vs. Discrete Collection Strategies. An account-treatment action is composed of a set of measures that can be taken in different order by different people (with different levels of authority and at different levels of proficiency). The perceived authority of the bank's representative (henceforth referred to as "agent") and the agent's experience in dealing with debtors play a significant role in determining the level of repayment intensity $\hat{\lambda}_m$ a collection action (of type $m \in \mathcal{M}$) can achieve and maintain. Collection records available in practice usually show measures taken at discrete instants in time, which, compared with an agent's actual activities (that the authors have witnessed in person, on recordings, and through interviews), capture only a very narrow aspect of the task. Indeed, an effective action typically cannot be implemented instantaneously; a particular measure is often combined with other measures over multiple days or even weeks. For example, when a pre-agreed repayment date is approaching, automated phone calls, text messages, and reminder emails are sent to the debtor. Once a repayment has been missed and a sufficiently long time has elapsed (so the account state hits the boundary of $\mathcal{A}^*$), attempts are made by the collection team to establish contact and cooperation; indeed, more than one agent may be "working the account" because different skill sets are required or for behavioral reinforcement, somewhat analogous to "good cop/bad cop" interrogation techniques. A significant fraction of the agents' daily routine is, therefore, spent on automated dialing. Once first-party contact has been established, the next action depends on the state of the account. For example, negotiating an updated repayment schedule requires an agent to review the file, request updated information about the debtor's financial standing, and possibly conducting a field visit. To finally determine a sequence of installments acceptable to all parties requires additional contact cycles. Thus, in reality, the action of restructuring a repayment plan is not as discrete (both in time and intensity level) as it may appear in a data set. A collection agency's menu of actions $\mathcal{M}$ is finite not because only a finite set of intensity levels can be achieved a priori, but because the implementation of actions within that agency is often standardized and carried out by the same set of agents in a more or less similar manner.[25] The $\mathcal{M}$-envelope strategy outlined earlier captures the discretization effects caused by standardizing the process. It also captures both discrete and continuous elements of collection efforts, driven by the generic mismatch between the theoretically optimal intensity to implement, as prescribed by the solution A^* to (P), and the best feasible intensity level in the optimal discrete collection strategy $A^*_{\mathcal{M}}$.

Endnotes

[1] Meier and Sprenger (2010) find that in the United States a larger present bias exhibited by consumers is associated with higher credit-card debt. Telyukova and Wright (2008) show that the credit-card debt puzzle, that is, the failure of consumers to pay off high-interest credit debt using balances in low-interest-bearing accounts or cash, can be viewed as a special case of the return-dominance puzzle concerning the coexistence of assets with different rates of return, such as cash and savings accounts.

[2] This approach relates to the use of the Snell envelope in optimal stopping problems; see, for example, Karatzas and Shreve (1998, Appendix D).

[3] Related to the control of pure jump processes (including PDPs) is the control of Markovian systems with Brownian motion (Davis et al. 2010). The additional diffusion term tends to augment the differentiability of the value function, which, in turn, simplifies optimality proofs for threshold-type policies.

[4] The possibility of early settlement offers is subject to future research; see Chehrazi and Weber (2010) for a (robust) static approach.

[5] The chance of recovering an outstanding balance in the long run increases in λ_∞. The empirical regularity of observing larger repayments during an economic boom is reflected by a probabilistic increase in the number of repayments (Chehrazi and Weber 2015).

[6] The standard information filtration $\mathbb{F}$ represents the internal history of the repayment process $(T_i, Z_i\colon i \geq 1)$. Specifically, $\mathcal{F}_t \in \mathbb{F}$ with $\mathcal{F}_t = \sigma(\mathbf{1}_{\{Z_i \in \mathcal{B}\}} \mathbf{1}_{\{T_i \leq s\}}; i \geq 1, s \in [0, t], \mathcal{B} \in \mathcal{B}([0, 1]))$, where $\mathcal{B}([0, 1])$ is the collection of all Borel-subsets of $[0, 1]$.

[7] While the notion of "effort" may suggest a continuous dependence of the change of the repayment intensity, our formulation does allow for instantaneous changes in the intensity dynamics (i.e., jumps); see Section 3.1.

[8] In the United States, the relevant statutes of limitations vary by state and type of debt and also depend on the specifics of the credit agreement. According to the Consumer Financial Protection Bureau:

> In some states, the statute of limitations period begins when you failed to make a required payment on a debt. In other states it is counted from when you made your most recent payment, even if that payment was made during collection. In some states, even a partial payment on the debt will restart the time period. In most states, debt collectors can still attempt to collect debts after the statute of limitations expires.

See https://www.consumerfinance.gov/askcfpb/1389/what-statute-limitations-on-debt.html for details. The possibility of early termination of collections by the account holder through filing for bankruptcy protection can be captured in our formulation by increasing the discount rate ρ.

[9] In our context, a collection strategy consists of a sequence of account-treatment actions, each of which is implemented by taking a sequence of treatment measures over a period of time. For example, establishing first-party contact is an action that is implemented by taking measures such as phone calls, letters, emails, messages, and field visits.

[10] This is reflected in our model by the coefficient δ_2.

[11] The complementarity of the terms in Equation (6), relating to discrete and continuous collection efforts, requires excluding the somewhat trivial situation where $a^*(\lambda, w) = 0$. Thus, if a is required to be strictly positive, the corresponding minimum may not exist, and in that case, its value is set to $+\infty$. Dropping this restriction, the largest optimal discrete effort in Equation (8) naturally always exists.

[12] Given $A = 0$ and $t \in [T_i, T_{i+1})$, the point $(\lambda(t), W(t)) \in \partial\mathcal{A}^*$ lies in $\mathcal{H}^*$ if and only if $\lambda(t) > \lambda_\infty$ and $(\lambda(\hat{t}), W(\hat{t})) \in \mathcal{A}^*$ for all $\hat{t} \in (t, t+\epsilon]$ and some $\epsilon > 0$. Since the repayment probability, $\mathbb{P}(T_{i+1} \in (t, t+\epsilon] \mid \mathcal{F}_t) = 1 - \exp(-\int_t^{t+\epsilon} \lambda(s)\, ds)$, goes to zero as $\epsilon \to 0^+$, with probability 1 there is no repayment in some right neighborhood of t, so $W(\hat{t}) = W(t)$ for all $\hat{t} \in (t, t+\epsilon]$.

[13] The claim follows since, by Corollary 2(iii), $\lim_{\hat{w}\to\infty} \partial_1 u(\lambda, \hat{w}) = -\infty$.

[14] MRA can be relaxed by taking the limit for $\underline{r} \to 0^+$. For any $\underline{r}$, the construction of the value function presented in this paper is exact relative to the repayment distribution F_R. The latter is empirically indistinguishable from a distribution that satisfies MRA by identifying $\underline{r}$ with the lowest observed nonzero relative repayment.

[15] See the proof of Lemma 2 for algebraic details.

[16] The optimal collection strategy in Corollary 5 is unique up to changes of the continuous collection effort on a set of measure zero. Moreover, the second and subsequent account-treatment actions do not induce any jump in the intensity process since they are taken at the optimal time, that is, before the intensity enters the action region.

[17] In any practical setting, the relative repayment distribution $F_R(\cdot)$ can be estimated by an empirical cdf $\hat{F}_R(\cdot)$, which, in our case, has been discretized in 0.005 intervals on $[\underline{r}, 1]$.

[18] It is assumed that at the time of placement $\lambda(0) = \lambda_\infty$. Since the primary reason for collection is that the holder has failed to make any repayment for a considerably long period (between 90 and 180 days, depending on the institution), given the intensity dynamics, it is reasonable to assume that $\lambda(t)$, $t \le 0$, has converged to λ_∞ before the collection process starts (at the time of placement, $t = 0$).

[19] The objective function of the collection problem (P) for a given collection strategy A starting at (λ, w) is denoted by $v(\lambda, w; A)$.

[20] These distributions are related to the dynamic collectability score (DCS) introduced by Chehrazi and Weber (2015); they are computed here for a simulation horizon that is large enough to cover the time required for full repayment for all sample paths.

[21] The intensity process is right-continuous at τ_k that are due to repayment events, as these events are not predictable, and it is left-continuous at τ_k that are due to (predictable) discrete collection efforts.

[22] As an alternative to the SDE (4), one can describe the evolution of the repayment intensity also in integral form: $\lambda(t) = \lambda(0) + \int_{[0,t]} \kappa(\lambda_\infty - \lambda(s^-))\, ds + \int_{[0,t]} \int_{[0,1]} (\delta_{10} + \delta_{11} r) \hat{\mathcal{N}}(ds, dr) + \delta_2 A(t)$, for all $t \ge 0$ (see Brémaud 1981).

[23] The last expression is obtained by solving for $u(\hat{\lambda}, w)$ using $(\mathcal{D}_0 u)(\hat{\lambda}, w) = 0$ and then differentiating with respect to $\hat{\lambda}$.

[24] As in the main text, $\lambda^*(w)$ is the boundary of the action region $\mathcal{A}^*$. Without loss of generality, $\lambda_\infty < \hat{\lambda}_1 < \cdots < \hat{\lambda}_M < \infty$.

[25] A set of actions in a given collection policy applied to statistically indistinguishable accounts by two different collection agencies may well induce statistically different intensity levels since the agencies are likely to implement the "same" nominal actions by taking different measures, in different orders, employing agents with heterogeneous levels of proficiency and debtor-perceived authority.

References

Aït-Sahalia Y, Cacho-Diaz J, Laeven RJA (2015) Modeling financial contagion using mutually exciting jump processes. *J. Financial Econom.* 117(3):585–606.

de Almeida Filho AT, Mues C, Thomas LC (2010) Optimizing the collections process in consumer credit. *Productions Oper. Management* 19(6):698–708.

Bar-Ilan A (1990) Overdrafts and the demand for money. *Amer. Econom. Rev.* 80(5):1201–1216.

Bar-Ilan A, Maimon O (1993) An impulse-control method for investment decisions in dynamic technology. *Managerial Decision Econom.* 14(1):65–70.

Basel Committee on Banking Supervision (2004) *International Convergence of Capital Measurement and Capital Standards: A Revised Framework* (Bank for International Settlements, Basel, Switzerland).

Bayraktar E, Ludkovski M (2009) Sequential tracking of a hidden Markov chain using point process observations. *Stochastic Processes Appl.* 119(6):1792–1822.

Bensoussan A, Lions J-L (1982) *Contrôle Impulsionnel et Inéquations Quasi-Variationnelles* (Dunod, Paris). [English translation: *Impulse Control and Quasi-Variational Inequalities*, Gauthier-Villars, Paris, 1984.]

Bensoussan A, Liu RH, Sethi SP (2005) Optimality of an (s, S) policy with compound poisson and diffusion demands: A quasi-variational inequalities approach. *SIAM J. Control Optim.* 44(5):1650–1676.

Berge C (1963) *Topological Spaces* (Oliver and Boyd, London, UK). [Originally published in French: *Espaces Topologiques—Fonctions Multivoques*, Dunod, Paris, 1959.]

Bertsekas DP (1995) *Nonlinear Programming* (Athena Scientific, Belmont, MA).

Bertaut CC, Haliassos M (2006) Credit cards: Facts and theories. Bertola G, Disney R, Grant C, eds. *The Economics of Consumer Credit* (MIT Press, Cambridge, MA), 181–237.

Bertaut CC, Haliassos M, Reiter M (2009) Credit card debt puzzles and debt revolvers for self control. *Rev. Finance* 13(4):657–692.

Bierman H, Hausman WH (1970) The credit granting decision. *Management Sci.* 16(8):B519–B532.

Brémaud P (1981) *Point Processes and Queues: Martingale Dynamics* (Springer, New York).

Chehrazi N, Weber TA (2010) Monotone approximation of decision problems. *Oper. Res.* 58(4):1158–1177.

Chehrazi N, Weber TA (2015) Dynamic valuation of delinquent credit-card accounts. *Management Sci.* 61(12):3077–3096.

Costa OLV, Davis MHA (1989) Impulse control of piecewise-deterministic processes. *Math. Control, Signals Systems* 2(3): 187–206.

Crook JN, Edelman DB, Thomas LC (2007) Recent developments in consumer credit risk assessment. *Eur. J. Oper. Res.* 183(3): 1447–1465.

Dai JG (1995) On positive Harris recurrence of multiclass queueing networks: A unified approach via fluid limit models. *Ann. Appl. Probab.* 5(1):49–77.

Davis MHA (1984) Piecewise-deterministic markov processes: A general class of non-diffusion stochastic models. *J. Roy. Statist. Soc.* (*Ser. B: Methodological*) 46(3):353–388.

Davis MHA, Zervos M (1998) A pair of explicitly solvable singular stochastic control problems. *Appl. Math. Optim.* 38(3):327–352.

Davis MHA, Guo X, Wu G (2010) Impulse control of multidimensional jump diffusions. *SIAM J. Control Optim.* 48(8):5276–5293.

Dixit AK (1991) A simplified treatment of the theory of optimal regulation of Brownian motion. *J. Econom. Dynam. Control* 15(4): 657–673.

Duffie D, Pan J, Singleton KJ (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica* 68(6):1343–1376.

Feng H, Muthuraman K (2010) A computational method for stochastic impulse control problems. *Math. Oper. Res.* 35(4): 830–850.

Fernandez R, Rosenthal RW (1990) Strategic models of sovereign-debt renegotiations. *Rev. Econom. Stud.* 57(3):331–349.

FRB (2018) Charge-Off and Delinquency Rates on Loans and Leases at Commercial Banks. Federal Reserve Bank, Washington, DC, http://www.federalreserve.gov/releases/chargeoff/.

Guo X, Miao H, Morellec E (2005) Irreversible investment with regime shifts. *J. Econom. Theory* 122(1):37–59.

Harrison JM, Sellke TM, Taylor AM (1983) Impulse control of Brownian motion. *Math. Oper. Res.* 8(3):454–466.

Hawkes AG (1971) Spectra of some self-exciting and mutually exciting point processes. *Biometrica* 58(1):83–90.

Henderson V, Hobson D (2008) An explicit solution for an optimal stopping/optimal control problem which models an asset sale. *Ann. Appl. Probab.* 18(5):1681–1705.

Jarrow R, Yu F (2001) Counterparty risk and the pricing of defaultable securities. *J. Finance* 56(5):1765–1799.

Karatzas I, Shreve SE (1998) *Methods of Mathematical Finance* (Springer, New York).

Laibson D (1997) Golden eggs and hyperbolic discounting. *Quart. J. Econom.* 112(2):443–477.

Liebman LH (1972) A Markov decision model for selecting optimal credit control policies. *Management Sci.* 10(10):B519–B525.

Meier S, Sprenger C (2010) Present-biased preferences and credit card borrowing. *Amer. Econom. J.: Appl. Econom.* 2(1): 193–210.

Mitchner M, Peterson RP (1957) An operations-research study of the collection of defaulted loans. *Oper. Res.* 5(4):522–545.

Pliska SR (1975) Controlled jump processes. *Stochastic Processes Appl.* 3(3):259–282.

Posner MJM, Zuckerman D (1990) Optimal R&D programs in a random environment. *J. Appl. Probab.* 27(2):343–350.

Porter MD, White G (2012) Self-exciting hurdle models for terrorist activity. *Ann. Appl. Statist.* 6(1):106–124.

Roberts MR, Sufi A (2009) Financial contracting: A survey of empirical research and future directions. *Annual Rev. Financial Econom.* 1:207–226.

Rosenberg E, Gleit A (1994) Quantitative methods in credit management: A survey. *Oper. Res.* 42(4):589–613.

Scarf HE (1960) The optimality of (s, S) policies in the dynamic inventory problem. Arrow KJ, Karlin S, Suppes P, eds. *Mathematical Methods in the Social Sciences* (Stanford University Press, Stanford, CA), 196–202.

Telyukova IA, Wright R (2008) A model of money and credit, with application to the credit card debt puzzle. *Rev. Econom. Stud.* 75(2):629–647.

Thomas LC (2000) A survey of credit and behavioral scoring: Forecasting financial risk of lending to consumers. *Internat. J. Forecasting* 16(2):149–172.

Thomas LC, Oliver RW, Hand DJ (2005) A survey of the issues in consumer credit modelling research. *J. Oper. Res. Soc.* 56(9): 1006–1015.

Thorisson H (2000) *Coupling, Stationarity, and Regeneration* (Springer, New York).

Trench MS, Pederson SP, Lau ET, Ma L, Wang H, Nair SK (2003) Managing credit lines and prices for bank one credit cards. *Interfaces* 33(5):4–21.

Weber TA (2011) *Optimal Control Theory with Applications in Economics* (MIT Press, Cambridge, MA).

Xu L, Duan JA, Whinston AB (2014) Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Sci.* 60(6):1392–1412.

Yang H, Zhang L (2005) Optimal investment for insurer with jump-diffusion risk process. *Insurance: Math. Econom.* 37(3): 615–634.