1-2023

Designing payment models for the poor

Bhavani Shanker UPPARI
Singapore Management University, bhavaniu@smu.edu.sg

Sasa ZORC

Designing Payment Models for the Poor

Bhavani Shanker Uppari
Singapore Management University, 50 Stamford Road, 178899 Singapore, bhavaniu@smu.edu.sg

Saša Zorc
Darden School of Business, University of Virginia, Charlottesville, Virginia 22903, zorcs@darden.virginia.edu

Several basic services, such as energy, clean water and cooking gas, are currently out of reach for millions of people living in poverty. There has been an emergence of private firms that offer these services by, for example, selling solar home systems or clean cooking packages with remote lockout capabilities. These firms deploy a pay-as-you-go (PAYGo) model in which consumers are given the flexibility to manage the amount and frequency of their payments based on their own erratic cash flows. However, because a firm under this model cannot observe how much money the consumers have, they can pay less to the firm and turn their income to other needs. We employ an optimal contracting approach to investigate the incentives that can mitigate such misuse of payment flexibility.

The optimal contract summarizes a consumer's payment history with a single score, which we call the v-score. The contract allows the consumer to flexibly pay small (resp., large) amount when her disposable income is low (resp., high), and accordingly adjusts the v-score to incentivize payments. The v-score over time rises with relatively high payments and drops with relatively low payments. The v-score also determines the level of technology access granted to the consumer and whether the contract is terminated or continued. The contract proposes an inclusive downpayment scheme that allows consumers to purchase their initial v-score. We discuss how the optimal contract can be implemented in the field and how it can solve some of the practical problems currently faced by the PAYGo firms.

1. Introduction

A significant proportion of the world's population lives on less than \$2 (US) per day, with no access to facilities such as energy, clean water, and cooking gas, services the rest of us take for granted. This part of the world's population is often referred to as the bottom of the pyramid (BoP). Providing access to basic services is necessary to lift people out of poverty. Government-driven supply of such services (e.g., through connections to the national electrical grid and water supply network) requires substantial capital investment, and in many cases, it may neither be technically feasible nor economically viable to extend the network to remote impoverished regions. Consequently, we are seeing private organizations stepping in, for example by selling solar home systems and operating local water purification and distribution systems.

Often, purchasing technologies such as a solar home system or a prepaid meter for a piped water connection is costly for BoP consumers because their incomes are low, irregular, and unpredictable, and a near-total lack of access to efficient mechanisms for saving and borrowing makes matters even worse. But a new payment model called pay-as-you-go (PAYGo) is becoming prominent in the BoP markets. It aims to alleviate the liquidity constraints of these consumers by allowing them to make a downpayment to gain access to a technology, and thereafter make payments to maintain access. At the end of the process, a consumer may come to own the technology.

PAYGo models are extant in the off-grid solar industry. Players include Azuri, BBOXX, d.Light, M-Kopa, Simpa Networks, Sun King, and Zola, operating in sub-Saharan Africa and South Asia. They use the PAYGo model to sell a variety of solar products, ranging from small solar lanterns with a mobile-recharging facility that cost tens of dollars, to large solar home systems with multiple bulbs, fans, and televisions that cost hundreds to even a thousand dollars. Under PAYGo a low-income consumer can sign up for a solar loan without a bank account or a credit history. She does not have to pledge collateral, and can repay flexibly at a speed that matches her repayment capability (within some limits). The PAYGo firms leverage Internet of Things (IoT) technology to remotely lock or unlock their solar devices based on consumers' payment histories, thereby controlling consumer access to devices. The PAYGo firms also adopt a flexible product ownership model, whereby payments are spread out over a longer period of time, reducing the upfront cost burden on the borrower. The freedom to decide how much and when to pay is more attractive than a straight-jacket fixed installment plan for most households (Winiecki 2019). The underlying assets sold to consumers represent collateral that can be repossessed or returned if the borrower is not able to continue with the payments. The firms enable the use of mobile phones for repayments through mobile money or scratch cards, thereby reducing travel costs and time spent in making repayments. The combination of digital payments and remote lockout has given PAYGo solar firms the confidence to extend financing to millions of people who would otherwise go unserved.

Following the trend in the off-grid solar sector, several other industries have also adopted the PAYGo model. Companies such as M-Gas and PayGo Energy provide clean cooking packages, consisting of a cylinder filled with liquefied petroleum gas (LPG), a gas cooker, and a smart meter that tracks usage and locks when the consumers' purchased credits expire. Consumers pay flexibly for gas over time, and the companies change cylinders in a timely manner. Safe Water Network builds and operates small water enterprises which consist of pumping and purification stations that distribute water to private households equipped with prepaid meters that lock the water supply automatically when the purchased credits expire. PayJoy enables the purchase of costly smartphones through their patented lock technology that limits app and call functionalities when consumers are late on their payments. With similar locking technologies, Angaza (in collaboration with Endless Solutions) and Kenyan Green Supply (with PaygOps) have deployed PAYGo computers.

Payment flexibility is at the core of any PAYGo business model, but as we argue next, it is a double-edged sword. Our objective in this paper is to design mechanisms to manage it efficiently.

On the one hand, payment flexibility helps the poor manage payments based on their erratic cash flows, thereby giving them access to a life-improving product or service which would otherwise have been out of reach. Moreover, the ability to flexibly pay allows consumers to invest in other projects (e.g., purchase livestock, furniture, and motorbikes), where flexible payments might not be an option (Zollmann et al. 2017). Thus, given the uncertain and constrained environments that these consumers live in, offering flexibility is almost an essential feature of any payment contract designed to serve the poor. Indeed, consumers make use of this feature: they sometimes opt to make no payments at all in a given payment period, and at other times bundle their payments and pay in advance for future access (Guajardo 2021, Bonan et al. 2023). Research has shown that flexibility in payment schemes reduces financial stress and enables consumers to manage their income better (Field et al. 2012, Barboni 2017). They can also skip payments without feeling ashamed about their short-term cash flow problems (Zollmann et al. 2017). It therefore comes as no surprise that PAYGo technologies have spread widely in developing countries (Lighting Global 2022).

On the other hand, payment flexibility comes with its own operational challenges. Given that a consumer has the flexibility to pay but the firm has no way of observing exactly how much money she has, the consumer can choose to redirect her income to other needs and pay less to the firm. In fact, surveys conducted on low-income consumers reveal that when they have payment flexibility, consumers can deprioritize making payments to the firm (Zollmann et al. 2017). This can lead to payment delays and defaults that reduce the firm's revenue from the contract, and hurt its sustenance in the long run. Even worse, bad payment histories can result in inefficient contract terminations and device confiscations, which can be difficult because consumers are usually located in remote rural areas, complicating the logistics of repossessing the devices. Irregular payments and defaults remain an issue in the PAYGo industry, and even led to the bankruptcy of off-grid solar firms Mobisol and Solarkiosk (Wiemann and Lecoque 2019).

Moreover, in practice, the offer of flexibility is oftentimes implicit. The consumers are given a reference payment schedule (e.g., a fixed installment plan) from which they are allowed to deviate as they wish, and they are generally not aware of the consequences of their deviations. On top of it, the flexibility—in terms of both the amount that the consumers pay and the time at which they pay—makes the communication of payment performance difficult. This can create a disconnect between consumers' and firm's perceptions of good versus bad repayment behavior (Zollmann et al. 2017, Bonan et al. 2023). Such issues exacerbate payment problems and necessitate the design of a formal framework that effectively communicates to a consumer her current standing in terms of payments, and based on her current standing, gives her optimal payment incentives to prevent the misuse of flexibility, grants her access to the device, and calls for either the confiscation of the device or the transfer of its ownership in a timely manner. Our efforts in this paper are targeted towards creating such a framework—using the optimal contracting approach—and providing actionable insights into efficiently designing and managing a pay-as-you-go model.

Methodological approach. In our analytical model, the timeline is split into multiple payment periods (e.g., weeks or months) over a repayment term. If the consumer takes up the contract, then in each period, her ability to pay the firm is determined by the disposable income that is available to her after accounting for all her essential needs. We assume that the disposable income is a random variable whose distribution is known to the firm, while its realizations are privately observed by the consumer. In every period, the disposable income is used both to make the payment to the firm for access to the device (e.g., a solar home system) and to satisfy any alternative needs. The firm's objective is to incentivize the consumer to split her money between consumption and payment in an optimal manner such that its own payoff is maximized, even though the firm does *not* observe consumer's disposable income. This is a moral hazard problem, and we examine the contractual mechanisms that can solve it. Rather than imposing any particular structure on the contract between the firm (principal) and the consumer (agent), we solve for an optimal mechanism given certain incentive constraints.

In addition to the revenue made from payments, the firm might gain some payoff upon contract termination. If the contract terminates with device confiscation, the firm could sell the device in a secondary market or refurbish it and transfer it to another consumer. Conversely, if the device ownership is transferred to the consumer, the firm might earn a government subsidy or more funding from investors.

In any period, the firm can influence the consumer's behavior using three nonmonetary levers: (i) restrict the consumer's access to the device using the remote lockout technology, (ii) confiscate the device, and

(iii) transfer the ownership of the device to the consumer. In addition, the firm recommends a payment scheme to the consumer that is incentive-compatible given her preferences. Based on the incentive structure laid out by the firm, the consumer makes her payment decisions by maximizing her own long-term value derived from accessing the device and satisfying her alternative needs.

Thus, an optimal mechanism consists of a sequence of contractual policies that determine for each period the amount of access that must be granted, whether the device must be confiscated or the ownership transferred, and the payment scheme recommended to the consumer, which together maximize firm's long-term payoff. The firm's sequential decision problem involves solving for these policies as functions of all the possible payment histories; hence, the state space of the problem in each period is the entire payment history until that point. There are no trivial reductions of the problem's state space, and we naturally expect the policies and recommendations to depend on a consumer's entire history of payments. The problem is therefore complex. We simplify it by using the work of Green (1987) and Spear and Srivastava (1987), who show that there exists a recursive formulation that is *equivalent* to the firm's sequential maximization problem, in which after any history of payments by the consumer, the value promised to the consumer in the future effectively summarizes all the information provided by the history itself. Therefore, the promised future value can be used as a state variable, instead of the entire payment history, in the recursive version of the problem.

Results. The aforementioned problem simplification has an important practical implication. The better the consumer's payment history, the higher the value of the problem's state variable (the future value promised to the consumer). In other words, the optimal contract uses this state variable as a score to evaluate a consumer's performance; we refer to it as the v-score. This v-score acts as a simple and transparent way to communicate to the consumer how well she is doing in terms of payments, so we could also interpret it as a credit score. The v-score also plays a central role in shaping consumer incentives in the optimal contract.

To initiate the contract, the firm in our model proposes a menu of downpayments (i.e., a flexible *downpayment scheme*), through which consumers with different levels of liquid wealth can purchase their positions in the contract and can start the contract at different values of v-scores. The higher the downpayment, the higher the initial v-score will be. The menu could include the extreme of a large downpayment to obtain ownership of the device right away. In contrast to the rigid initiation methods with just two options that are adopted currently—wherein firms sell the devices to consumers who can afford to purchase, and to those who cannot, firms charge a single value of downpayment to all consumers—our menu is more *inclusive*, and beneficial for consumers with low levels of liquid wealth.

After the contract starts, the v-score evolves based on the consumer's payments through time. In every period, the optimal contract implements payment flexibility explicitly by offering a *payment scheme* that recommends a payment amount to the consumer as a function of the most she could possibly pay (i.e., her disposable income). The contract balances this flexibility with appropriate payment incentives by offering a *reward scheme* that informs the consumer how her v-score will be updated upon paying a certain amount. Thus, the contract incentivizes payments through updates to v-scores.

The suggested payment scheme recommends consumer pay a small (resp., large) amount when her disposable income is low (resp., high); this flexibility in payment respects consumer's liquidity constraints and accommodates uncertainty in her income. Interestingly, flexibility as a feature disappears—i.e., the contract asks for a fixed installment to be paid every period, which is commonplace in the top-of-the-pyramid (ToP) market—if consumers in the market have sufficiently high incomes and v-scores (analogous to credit scores), which are also the characteristics of the ToP market consumers. Thus, payment flexibility as a contractual feature is specific and essential to the BoP market.

The higher the payment made by the consumer, the higher the updated v-score will be, and hence the higher the ensuing reward will be (in the form of more access to the device, lower probability of confiscation, and higher probability of ownership transfer). But it is not the case that small payments will result in small increments to the v-score, because small payments could either indicate low disposable income realizations or large income diversions (which is the core of the agency issue in our setting). The contract moderates the latter issue by dropping the v-score for small payments. As a result, the v-score over time *rises* with relatively *high* payments and *drops* with relatively *low* payments. Furthermore, the optimal contract offers stronger payment incentives (i.e., asks for a higher amount and also rewards at a higher rate) to consumers with intermediate values of v-scores, and weaker payment incentives to those with relatively low and high v-scores. The consumers at the extremes are closer to termination and that itself is incentive enough for them, but those in the middle might require an extra push.

The optimal contract offers threshold-based guidance for termination and access provision. If the v-score in a period falls below a confiscation threshold, the firm's cost of continuing the contract overrides the benefit, making it optimal to initiate the confiscation process. Similarly, if the v-score goes above an ownership threshold, indicating that the consumer has paid sufficient to gain complete control of the device, it is optimal to initiate ownership transfer. If the score is somewhere in between, the contract continues to the next period. Moreover, the device is disabled (resp., enabled) for the entire duration of the period if the v-score is below (resp., above) a threshold, and partial access is granted for intermediate values of v-score. The contract rewards good payment histories with grace periods, during which the consumer can access the device even if she does not make any payment.

In addition to shedding light on the structure of the optimal contract, we discuss how it can be implemented in practice using simple (digital) lookup tables on devices. We also compare its key features with those of contracts that are usually seen in the ToP markets. The features suggested by the optimal contract are both insightful and implementable, and they can form the basis for experimentation in the field. In the appendix, we show that the three prominent business models in the BoP market—sales model, rent-to-own model, and rental model—emerge as special cases of our contracting framework. Overall, we believe that the results in our paper can help PAYGo firms offer efficient contracts that ensure their long-term survival, which is also key to delivering life-improving goods and services to low-income consumers.

2. Related Literature

We broadly relate to the growing body of research that studies the operational issues in the BoP markets. For example: Balasubramanian et al. (2023) study the inventory issues arising in the context of mobile money agents; Calmon et al. (2022) examine distribution strategies for durable goods in the BoP market;

Gernert et al. (2023) evaluate different business models for emergency transportation systems in low-income countries; de Zegher et al. (2018) propose payment strategies to curb illegal deforestation by smallholder farmers; Kundu and Ramdas (2022) discuss the importance of timely after-sales service in increasing the adoption of PAYGo solar home systems; and Ramdas and Sungu (2022) show that giving BoP consumers too much flexibility in mobile data usage can result in their information isolation.

Particularly relevant are the papers that discuss the importance of offering payment flexibility to BoP consumers. Uppari et al. (2019) study why some consumers may prefer using kerosene to rechargeable lamps even when the latter cost less. They find that the *quantity flexibility*—in terms of how much one pays to purchase—offered by kerosene plays an important role in shaping such a preference. Using the recharge data from Rwanda, Uppari et al. (2023) show that simple strategies that offer payment *timing flexibility* (e.g., allowing consumers to prepay for recharge or recharge on credit) help consumers manage their income uncertainties better, and can attain half the benefit arising from eliminating liquidity constraints altogether. Interestingly, a PAYGo model brings together *both* quantity and timing flexibility (i.e., it gives the freedom to decide how much and when to pay). As we argued earlier, offering flexibility without appropriate incentives to counter it can be detrimental, and how to balance the two is the topic of our paper. Additionally, Guajardo (2021) examines the repayment performance for PAYGo solar lamps in six different countries and finds that consumers make extensive use of payment flexibility: they tend to bundle their payments, and the bundles constitute both the payments that were skipped in the past weeks as well as the advance payments for the upcoming weeks. Guajardo (2021) also emphasizes the importance of PAYGo model design for the adoption of off-grid energy products, the focus of our paper.

Payment flexibility also features in the literature on Microfinance Institutions (MFIs), see e.g., Jain and Mansuri (2003), Tedeschi (2006), Field et al. (2012) and Barboni (2017). MFIs offer savings, credit, and insurance products, all of which require consumers to make payments regularly. Thus, in this case too, consumers can benefit from the flexibility to adapt financial transactions to their cash flows. A key difference between MFI and our setting is that device usage in a PAYGo model can be controlled remotely. If a consumer misses payments, the device can be automatically locked—an important mechanism to encourage more disciplined payment behavior. By contrast, if a consumer misses payments with an MFI, there is nothing much that an MFI can do remotely, which is why there is a strong emphasis on various alternative discipline mechanisms in MFI literature. Specifically, discipline mechanisms in this context are designed to encourage clients to fulfill their financial commitments. The common reasons cited for commitment breaches (see, e.g., Laureti 2012, Labie et al. 2017) are lack of self-control, inattention to planning, intra-household disagreements, and the payment-related moral hazard problem discussed in our paper. To counter these issues, MFIs deploy a variety of discipline mechanisms such as asking consumers to make (small) payments every week, sending them SMS reminders, requiring financial collateral, exploiting their social collateral, and sending deposit collectors frequently to their homes. Interestingly, the ability to control devices remotely in a PAYGo model mitigates the necessity for such strong discipline mechanisms. Our paper focuses closely on the moral hazard problem—with a particular type of discipline mechanism in place—and the incentives that solve this problem. Inattention and self-control problems are perhaps relevant in the PAYGo context as well (see e.g., Bonan et al. 2023), and we leave their theoretical examination to future research.

In terms of methodology, we make use of the techniques developed by Green (1987) and Spear and Srivastava (1987) to solve the dynamic principal-agent problems. In full generality, optimal dynamic mechanisms can be extremely complex in their structure. The primary cause of this complexity is that the mechanism can depend on the entire history of events up to a specific point. The techniques from the aforementioned papers help us reduce a sequential history-dependent contracting problem into a recursive contracting problem, in which the sole state variable is the agent's "promised future utility." Applications of this methodology in economics are vast; an overview of advances in this field is given by Acemoglu et al. (2013), pp. 89–178. This methodology has also been extensively applied in the finance literature to present theories on firm's investment dynamics (Clementi and Hopenhayn 2006, DeMarzo and Fishman 2007a), to design optimal securities (DeMarzo and Fishman 2007b), and to analyze mortgage contracts (Piskorski and Tchistyi 2010, 2011). Use of this methodology in the management science literature is more recent: it has been applied to study incentives for environmental disclosures (Wang et al. 2016), dynamic resource allocation problems (Balseiro et al. 2019), and repair and maintenance contracts (Tian et al. 2021).

The modeling of the agency problem in our paper is particularly inspired by the work of DeMarzo and Fishman (2007b), who design a financial contract that induces an agent managing a project to make payments to investors rather than diverting cash flows to herself. That said, there are some crucial differences that set our work apart. In most financing papers, both the payments (from the agent to the principal) and the rewards (from the principal to the agent) are *monetary*. This literature often deploys the revelation principle, under which it is optimal for the agent to pay all the cash flow from the project to the principal, and the principal thereafter pays cash rewards back to the agent (either now or later). The bidirectional flow of cash between the principal and the agent allows truthful revelation to be optimal, which also trivializes the agent's payment policy. By contrast, the principal in our setting uses *nonmonetary* levers (remote lockout, confiscation, and ownership transfer) to incentivize the agent, with also a limit on the maximum that the principal can promise the agent, which is ownership. Because of resultant constraints on what the principal can offer the agent and vice versa, the technique of the revelation principle does not extend to our setting. We solve for a payment policy that is incentive-compatible for the agent, in addition to a reward policy that is optimal for the principal—a harder problem.

Another source of complexity arises because the consumer in our model derives utility from two sources: from access to the device, and from satisfying her alternative needs (called consumption utility). The two need not be perfect substitutes, so we incorporate convex consumer preferences by making the consumption utility function *concave*. This feature disables many commonly-used solution techniques that rely on the linearity of utility functions (as in, e.g., DeMarzo and Fishman 2007b, Biais et al. 2010, Tian et al. 2021). Although a concave utility is not rare in dynamic moral hazard problems, the commonly-used information structure differs from ours. In a typical moral hazard problem, the agent takes an unobservable and costly action, and that action determines an observable output (e.g., Hopenhayn and Nicolini 1997, Balke and Lamadon 2022). The agent's preferences generally include a concave utility from an output-based reward and a convex cost incurred for action. The reward and action policies remain additively separable in such models. By contrast, the unobservable in our case is the disposable income, while the observable is the action, i.e., payment. The payment determines both the reward as well as the cost (reflected as a reduction in consumption utility). So, we lose the additive separability of the reward and payment policies in our model, which further complicates the analysis.

Specifically, the methodological complexities manifest in Section 4.3, and require novel solution techniques to solve the problem: we use a piecewise linear approximation of the consumption utility, we extend the standard stochastic dominance results so that they can be established via one-sided slopes (Lemma 1 in Appendix C), and we deploy a change-of-variable technique and smoothing approximations to prove Proposition 4—the key result for backward induction to work in the recursive contracting problem. We believe that the solution techniques that we have developed will generalize to models that share similar features.

3. The Contracting Problem

The consumer in our model lives on a low and uncertain income, with no access to efficient saving and borrowing mechanisms. She also lacks access to basic services such as energy, clean water, and cooking gas. The firm owns a device that provides one of these services when accessed. Depending on the technology that is offered, the access to a device could translate to accessing (i) light from the bulbs, (ii) a port that allows the recharge of other devices, (iii) information from the television, (iv) water from the piped connection, (v) gas from the cylinder, and (vi) apps on a smartphone or a computer. Owing to consumer's liquidity constraints, the firm creates a *long-term payment contract* for the consumer to give her access to the device.

Even if the consumer is located in a remote rural area, the following two important features of our setting make regular interactions between the consumer and firm possible. First, the consumer makes payments to the firm using mobile money. Such mobile transfers have become fairly common in low-income regions of Africa and Asia. This feature allows the consumer to make payments without incurring any inconvenience of travel. Second, the device can be remotely turned on and off by the firm using IoT technologies. This feature enables the firm to nudge the consumer when she is lagging on her payments.

It is important for any payment contract that is designed for the poor to incorporate payment flexibility so that it can accommodate uncertainties in their lives. We investigate the incentives that can mitigate the misuse of payment flexibility. To that end, we formulate the corresponding optimal contracting problem in this section. We focus primarily on a version of the PAYGo model that includes the option of device ownership (e.g., model for solar home systems). In Appendix B, we discuss a version of the model without an ownership option (e.g., for services such as water and gas).

3.1. Firm's Model

The firm interacts with the consumer for a maximum of θ payment periods. The length of a payment period is exogenously given (e.g., a week, a fortnight, or a month), and the total duration of θ periods represents the *repayment term* of the contract—the maximum duration for which the contract lasts. The device has a lifetime of T periods, beyond which it does not generate any value for either the firm or the consumer and must be scrapped. Therefore, the repayment term satisfies $\theta \le T$. For ease of exposition, we assume that $\theta = T$; in Appendix A, we discuss the case where the repayment term could be less than T. At $t = 0$, the firm offers a contract, which the consumer either accepts or rejects. The first payment period is $t = 1$ and the last payment period (also the end of the device's lifetime) is $t = T$.

In any period $t > 0$, the firm has the following three levers to influence consumer behavior: based on the consumer's payment history, the firm can (i) restrict access to the device, (ii) confiscate the device, or (iii) transfer the device ownership. Levers (ii) and (iii) result in the termination of the contract.

Given a contractual arrangement that is in place in period t, the consumer chooses a payment p_t to maximize her value subject to her liquidity constraints (see Section 3.2). In addition to consumer payments, the firm gains some payoff upon termination. On the one hand, if the firm confiscates the technology in period t, then either by selling the device in a secondary market, or by refurbishing and transferring it to another consumer, the firm gains payoff $\mathcal{C}_t$. ($\mathcal{C}_T$ is the scrap value of the device at the end of its lifetime.) On the other hand, if the firm transfers the ownership in period t, then it gains payoff $\mathcal{O}_t$, which could be a monetary incentive in the form of a government subsidy or more funding from investors and donors. (It is not necessary in our model for $\mathcal{O}_t$ to be strictly greater than zero.) We denote by $\mathcal{K}$ the cost that the firm incurs to place the technology in the consumer's hands (e.g., cost to build or source the technology, custom tariffs, marketing and installation costs).

The firm is forward-looking with a discount factor γ. It designs a contract to maximize its discounted payoff from the consumer payments $\{p_s\}_{s=1}^{\tau}$, from termination (either $\mathcal{C}_\tau$ or $\mathcal{O}_\tau$), and from a downpayment d that the firm might charge to initiate the contract, minus the cost $\mathcal{K}$. Here, τ is the period in which the contract terminates—endogenously determined by the contract terms.

3.2. Consumer's Model

In period t, y_t is the consumer's disposable income that characterizes her *ability to pay* the firm in that period. In other words, y_t imposes a constraint on the consumer's payment p_t in period t: $0 \le p_t \le y_t$. Her ability to pay may vary over time (e.g., low if one of the household members has lost employment and high if the consumer has harvested crop recently), could be seasonal (e.g., because of rain cycles and rotating saving accounts), and may not be easily observable. Therefore, we assume that y_t is a realization of a non-stationary continuous random variable Y_t. We denote the cumulative distribution function (cdf) of Y_t by G_t, its probability density function (pdf) by g_t, and support by $\mathcal{Y}_t$.

Although the consumer is able to pay y_t, she may not be *willing to pay* it all because some of that money could be spent on her alternative needs. For example, she may need to pay children's school fees, purchase farming equipment, deal with the death of a cow, accommodate the visits of relatives, and buy new clothes, alcohol, or cigarettes. (See Zollmann et al. 2017 for a survey-based discussion on trade-offs faced by consumers while making payments to a PAYGo firm.) Need-management can therefore be a quite complex undertaking, involving prioritizing different needs and allotting money across those needs. We refrain from modeling such specific details, and instead model the (aggregate) utility that the consumer derives from satisfying her needs. By consuming an amount x_t to satisfy alternative needs in period t, the consumer derives a *consumption utility* $u(x_t)$, where the function u is increasing and concave. To simplify exposition, we make the following functional-form assumption: $u(x) = \lambda_1 x + (\lambda_2 - \lambda_1)[x - \kappa]^+$, where $\lambda_1 > \lambda_2$ and $[k]^+ = \max\{k, 0\}$

(analogously, we denote $[k]^- = \min\{k, 0\}$ which is used later). In Section 4.3, we discuss the implications of more general formulations for $u(x)$.

The consumer also derives an *access utility* from the technology. If provided access for a fraction f_t of period t, the resultant utility to the consumer is $u_a(f_t)$. The value of f_t is determined by the contract depending on the consumer's payment history. We assume that $u_a(f_t) = f_t A$. (A more general formulation is discussed in Section 4.2.) We interpret A as the maximum utility that the consumer derives in a period from having access to the technology and f_t as the fraction of that maximum utility that is provided to the consumer. For example, A is the utility derived from having access to light in the nighttime of every day in a period for a sufficient number of hours (e.g., 6 hours) so that the consumer can finish all of her desired activities, and $f_t = 0$ indicates turning the device off for the entire period.

We assume that the consumer's preferences over the set of bundles $\{(x_t, f_t) \mid x_t \in \mathbb{R}_+,\, f_t \in [0,1]\}$ are additive and model the value from bundle (x_t, f_t) as $\mathcal{U}(x_t, f_t) = u(x_t) + u_a(f_t)$. The function $\mathcal{U}$ has convex indifference curves in the space of x_t and f_t. In other words, the consumer prefers a bundle that incorporates to some extent both consumption and device access over a bundle that solely incorporates one of these two.

If the contract terminates with confiscation in period t, then the consumer spends her disposable income as she wants (including any expenditures on alternative ways to satisfy the core need, e.g., by using flashlights or kerosene), thereby deriving a value $U_s = \mathbb{E}u(Y_s)$ perpetually in every period $s > t$. In contrast, if the ownership is transferred in period t, then on top of consuming all of her disposable income perpetually, the consumer can access the technology free of cost for its remaining lifetime and scrap it for a value of R thereafter. Denoting the former value by $\underline{v}_t$ and the latter by $\overline{v}_t$, we have

$$\underline{v}_t = \delta \sum_{s=t+1}^{\infty} \delta^{s-t-1} U_s \quad \text{and} \quad \overline{v}_t = \delta \left[\sum_{s=t+1}^{\infty} \delta^{s-t-1} U_s + \frac{A(1-\delta^{T-t})}{1-\delta} + \delta^{T-t} R \right], \tag{1}$$

where δ is the consumer's discount factor, such that $\delta < \gamma$. We assume that the consumer is forward-looking, and given a contract, she makes her payment decisions by maximizing her discounted value from the sequence of bundles $\{(x_s, f_s)\}_{s=1}^{\tau}$ and from termination (either $\underline{v}_\tau$ or $\overline{v}_\tau$), subject to her liquidity constraints. (Interestingly, although we assume the consumer to be forward-looking, as we will see later in Sections 4 and 5, the optimal contract requires the consumer to make only single-period decisions.)

The consumer can enter the contract only if her ability to make a downpayment, which we denote by w, is greater than the downpayment d charged by the firm. We interpret w as the liquid wealth that a consumer usually has access to, which she can use to take up the technology. The ability w could vary across consumers in the market, so we model this heterogeneity as a random variable W.

3.3. Information Structure, Contract, and Timing

We assume that the firm knows the distribution G_t of the consumer's disposable income Y_t but does not know its realization y_t. It is well known that the portfolios of the poor are quite complex, consisting of multiple monetary instruments that are deployed to manage money and satisfy daily needs (Collins et al. 2009). The consumer's essential spending and possibly her total income, and hence y_t, are therefore too costly or even impossible to monitor regularly over time. But it is a common practice among several PAYGo firms to vet consumers before onboarding them. The vetting process—usually conducted by firms' sales agents—includes surveys that help determine a consumer's type, represented here by G_t, by asking questions on the occupations of the working members and the usual incomes and expenditures in the household. We also assume that the parameters of the utility functions u and u_a (i.e., λ_1, λ_2, and A) are known to the firm. These could be elicited in the surveys using procedures described in the decision sciences literature (for a summary on the procedures to elicit multiattribute utility functions, see Fischer 1975, 1976).

Given this information structure, the source of the agency problem is that the consumer's payment p_t is constrained by her income y_t, which the firm cannot observe. Consequently, when the consumer makes a low payment, the firm does not know whether the consumer diverted more of her income toward consumption *or* her income realization itself was low. The firm must therefore *incentivize* the consumer to split her income optimally between payment and consumption. The firm must solve this *moral hazard* problem through a contract that does not condition its terms on the realizations of the consumer's income.

In solving for an optimal contract, we will not restrict the contract form but instead solve for an optimal mechanism that can depend on all the observable variables in our model. The firm in our model observes (a) the payments that it receives from the consumer, (b) the compensations (in the form of access to the device) given to the consumer, and (c) the outcomes of possible public randomizations. As a function of the history of these observables, a contract specifies (i) fractional access to the device that is granted to the consumer in each period; and (ii) circumstances under which the contract is terminated. Given a contract, the consumer chooses her actions (i.e., payments) optimally.

The timing is as follows. At time $t=0$, the firm offers a technology on contract to the consumer. The firm charges a downpayment d to initiate the contract and install the device. If the consumer finds the contract sufficiently valuable *and* can afford the downpayment ($w \geq d$), then she pays d, takes up the contract, and the contract dynamics begin. In a period $t>0$, the consumer's disposable income y_t realizes. The consumer makes a payment p_t, consuming $y_t - p_t$. Given the history h^t of observables up to t (including p_t), the contract specifies the fractional access granted to the consumer, denoted by the function $f_t(h^t)$. Before proceeding to the next period, the contract may call for termination. We let $c_t(h^t)$ denote the contractually-specified probability of confiscation in period t and $o_t(h^t)$ the probability of ownership transfer. (We discuss the analytical importance of *probabilistic* termination and related practicalities in Section 4.1.) The consumer may also terminate the contract by quitting (the firm repossesses the device in that case). If not terminated early, the contract reaches its natural termination in period T. If the consumer owns the technology at the end of its lifespan, she scraps it at value R, whereas if the firm still owns it, then the firm scraps it at $\mathcal{C}_T$. Our focus in the remaining part of this section will be on contract dynamics *after* the initiation; we will return to the problem of initiation in Section 4.5.

Although the history observed by the firm contains (a)–(c) mentioned above, it can be reduced to contain only the payments without any loss of generality. Because the randomizations in our model determine whether the contract continues or terminates, the history that is relevant for a *continuation* contract can only contain the randomization outcomes that did not result in a termination, thus making (c) redundant. Moreover, the access given in any previous period can be solved recursively in terms of past payments alone, thereby rendering (b) redundant. (For example, say the fractional access in period 2, as a function of past access f_1, past payment p_1, and current payment p_2, is $F_2(p_1, f_1(p_1), p_2)$. It can be equivalently written as another function $f_2(p_1, p_2)$.) Hence, the history h^t in our model hereafter is simply the *payment history*, so $h^t \in \mathbb{R}_+^t$. (The contract begins in period 0 with an empty history $h^0 = \{\emptyset\} \in \mathbb{R}_+^0 = \{\{\emptyset\}\}$.)

A contract in place at the end of period t is a triplet (f^t, c^t, o^t), consisting of sequences of functions $f^t = \{f_s\}_{s=t+1}^T$, $c^t = \{c_s\}_{s=t+1}^T$, and $o^t = \{o_s\}_{s=t+1}^T$, where $f_s(h^s)$, $c_s(h^s)$, and $o_s(h^s)$ are all functions that map payment histories ($\mathbb{R}_+^s$) to $[0,1]$. Given the contract, the consumer chooses payment p_s strategically—depending both on her privately-observed income realization y_s and the history h^{s-1}—to maximize her value, for $s > t$. We denote the resultant consumer's payment strategy by $p^t = \{p_s(y_s, h^{s-1})\}_{s=t+1}^T$. We say that a contract-strategy pair (f^t, c^t, o^t, p^t) is *feasible* if it satisfies the following two conditions: (i) $0 \le p_s(y_s, h^{s-1}) \le y_s$, and (ii) $c_s(h^{s-1}, \check{p}_s) + o_s(h^{s-1}, \check{p}_s) \le 1$, for all $s \in \{t+1, \ldots, T\}$, $\check{p}_s \in \mathbb{R}_+$, $y_s \in \mathcal{Y}_s$, and $h^{s-1} \in \mathbb{R}_+^{s-1}$. Condition (i) reflects the consumer's liquidity constraint, and condition (ii) ensures that the probabilities of ownership and confiscation sum up to no more than 1. The history is updated as $h^s = (h^{s-1}, \check{p}_s)$, where $\check{p}_s$ is the amount that the consumer actually pays in period s. We denote by Γ_t the set of all feasible contract-strategy pairs.

For a $(f^t, c^t, o^t, p^t) \in \Gamma_t$, the firm's payoff Π_t and the consumer's value V_t at the *end* of period t (i.e., in the subgame starting from h^t) are given by:

$$\Pi_t(f^t, c^t, o^t, p^t) = \mathbb{E}\left[\sum_{t<s\le\tau} \gamma^{s-t} p_s + \gamma^{\tau-t}(\rho\mathcal{C}_\tau + (1-\rho)\mathcal{O}_\tau) \,\middle|\, f^t, c^t, o^t, p^t\right], \tag{2}$$

$$V_t(f^t, c^t, o^t, p^t) = \mathbb{E}\left[\sum_{t<s\le\tau} \delta^{s-t} \mathcal{U}(Y_s - p_s, f_s) + \delta^{\tau-t}(\rho\underline{v}_\tau + (1-\rho)\overline{v}_\tau) \,\middle|\, f^t, c^t, o^t, p^t\right], \tag{3}$$

where τ is the termination time and ρ is the binary variable describing whether the termination results in confiscation. Both τ and ρ are endogenously determined by the contract and the consumer's actions.

For a contract-strategy pair (f^t, c^t, o^t, p^t) to be *incentive-compatible*, the consumer's strategy choice p^t must be optimal. Formally,

$$V_t(f^t, c^t, o^t, p^t) \ge V_t(f^t, c^t, o^t, \check{p}^t) \quad \text{for all } \check{p}^t \text{ such that } (f^t, c^t, o^t, \check{p}^t) \in \Gamma_t. \tag{4}$$

Finally, we say that a contract-strategy pair (f^t, c^t, o^t, p^t) is *optimal* if it is incentive-compatible and there is no other incentive-compatible pair that provides the same value to the consumer and a higher payoff to the firm. A contract is optimal if it is part of an optimal contract-strategy pair.

As we can see, a contract-strategy pair is a sequential game between the firm and consumer, in which the firm pre-specifies game rules for all possible histories of the consumer's actions, and given those rules, the consumer plays the game optimally. This sequential formulation of the problem is difficult to analyze: for all the possible payment histories, we must find the optimal f_t, c_t, o_t, and p_t, which are *functions of these histories*. Moreover, the problem is neither memoryless nor Markovian in payments because we naturally expect the consumer's rewards to depend on the entire payment history, and so there are no trivial reductions of state space. Fortunately, the works of Green (1987) and Spear and Srivastava (1987) show how we can reformulate this complex sequential decision problem as a simpler Markovian decision problem, but Markovian in the future value promised to the consumer. Before we proceed to this recursive formulation, we make two remarks regarding our modeling assumptions.

Remark 1. Equation (3) assumes that in any period, the part of the consumer's disposable income that remains after the payment is fully consumed within that period. This is not an unreasonable assumption in our context: the consumer usually has more than enough needs in her life to absorb the remaining amount. Even if the consumer saves a part of it, our model requires that those savings are redirected elsewhere and they do not characterize her ability to pay the firm (her disposable income) in the future.

Remark 2. We have assumed that the technology is not income-generating for the consumer: the disposable income Y_t does not depend on the usage of the technology. This reflects our focus in the paper on the technologies whose primary function is to improve the living standards of the poor, as opposed to the ones whose primary benefit is to generate income, such as irrigation systems, e-rickshaws, and tractors. More importantly, the remote lockout—a critical feature in our model—is less appropriate for income-generating technologies, because the consumer depends on the device for revenue to service her loan, and turning it off might be counter-productive if the goal is to recover capital.

3.4. Recursive Formulation

We solve for an optimal contract using dynamic programming. Because the disposable income in our model is independent over time and we assumed that there are no private cash flows that affect the consumer's future disposable income, the functions Π_t and V_t are common knowledge in period t and independent of the prior history. (For a more detailed argument, see Spear and Srivastava 1987.) Consequently, an optimal contract must be optimal after any history. Otherwise, we could find an alternative contract that leaves the consumer's value and incentives unchanged but raises the firm's payoff. Therefore, for a given continuation value v_t^e promised to the consumer at the end of period t, we can write the continuation payoff π_t^e of the firm at the end of period t as follows:

$$\pi_t^e(v_t^e) = \max_{(f^t,c^t,o^t,p^t)\in\Gamma_t} \Pi_t(f^t,c^t,o^t,p^t)$$
$$\text{s.t.} \quad V_t(f^t,c^t,o^t,p^t) = v_t^e = \max_{\check{p}^t} V_t(f^t,c^t,o^t,\check{p}^t) \quad \text{for all } \check{p}^t \text{ such that } (f^t,c^t,o^t,\check{p}^t)\in\Gamma_t.$$

For any given v_t^e for the consumer, the *continuation function* π_t^e gives the highest possible payoff attainable by the firm. This continuation function fully characterizes the payoff-relevant attributes of the contractual relationship beyond period t. As a result, the payment history can be summarized by the current continuation value v_t^e promised to the consumer. An optimal contract provides incentives by effectively specifying how this continuation value varies with the consumer's payment history.

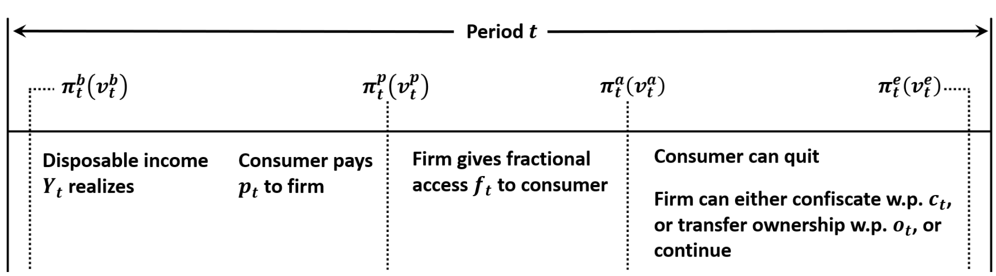

Figure 1 **Timeline of events within period t (w.p. stands for with probability).**

As illustrated in Figure 1, π_t^e is the continuation function at the *end* of period t. To set up the recursive version of the contracting problem, we define continuation functions at earlier stages in period t: π_t^a is the continuation function just after the firm gives fractional *access* f_t to the consumer, π_t^p is the continuation function just after the consumer makes *payment* p_t to the firm, and π_t^b is the continuation function at the *beginning* of period t before the consumer's income Y_t is realized. We also accordingly define the values that are promised to the consumer at different stages in a period, respectively, as v_t^e, v_t^a, v_t^p, and v_t^b. These continuation values satisfy the following accounting identity over time, which follows from (3):

$$\overbrace{\frac{V_{t-1}}{\delta}}^{v_t^b} = \mathbb{E}\Big[u(Y_t - p_t) + \underbrace{f_t A + \overbrace{c_t \underline{v}_t + o_t \overline{v}_t + (1 - c_t - o_t)\underbrace{V_t}_{v_t^e}}^{v_t^a}}_{v_t^p}\Big]. \tag{5}$$

The first four terms on the right-hand side of (5) correspond to the value that is given to the consumer in period t (in terms of consumption, access, and termination payoffs) and the last term to the expected future value that is promised at the end of period t. These current and future values must add up to the term on the left-hand side, the promise made to the consumer at the end of the previous period appreciated by the consumer's discount factor δ. Thus, equation (5), which we refer to hereafter as the *flow equation*, ensures that the promises made to the consumer over time are always met, either now or later. The flow equation also suggests an important trade-off that is critical to creating optimal dynamic incentives: the firm can give a higher value to the consumer in the current period and promise a lower value in the future, or vice versa.

We solve for the continuation functions by recursively assigning the current and future values to the consumer over time, working from the end of period t to its beginning. Going from period t to $t-1$, the intertemporal linkage is as follows:

$$\pi_{t-1}^e(v_{t-1}^e) = \gamma \pi_t^b(v_{t-1}^e/\delta). \tag{6}$$

From (5), by definition $v_t^b = v_{t-1}^e/\delta$. The promised future value at the end of period $t-1$, when postponed to the next period, appreciates by δ, and the firm discounts this future payoff by γ.

4. Solving for the Optimal Contract

In the following subsections, we characterize the firm's payoff functions π_t^a, π_t^p, and π_t^b using backward induction. A critical step in the analysis is to show that all these functions are concave. The analysis proceeds as follows. We begin with the inductive assumption that π_t^e is a concave function and then show in Sections 4.1, 4.2, and 4.3 that π_t^a, π_t^p, and π_t^b are also concave functions, implying that π_{t-1}^e is also concave because of (6). Section 4.4 solves the contracting problem in the last period and shows that π_T^b is also concave, which forms the basis of our inductive argument. It then follows by mathematical induction that all the payoff functions are concave. Thereafter, in Section 4.5 we discuss the contract initiation problem.

In the analysis, we denote by $\underline{v}_t^j$ and $\overline{v}_t^j$, for $j \in \{e, a, p, b\}$, the maximum and minimum values of the variable v_t^j. At the end of any period, the minimum value is delivered to the consumer by confiscating the device, and the maximum by transferring the ownership; therefore, $\underline{v}_t^e = \underline{v}_t$ and $\overline{v}_t^e = \overline{v}_t$, where $\underline{v}_t$ and $\overline{v}_t$ are the termination payoffs specified in (1). Figure 5(a) in the next section summarizes the findings from this section. An interested reader could refer to that figure while going through the results in this section.

4.1. Optimal Termination and Continuation

We first compute $\pi_t^a(v_t^a)$ from $\pi_t^e(v_t^e)$. The firm's problem is to decide whether to terminate or continue the contract while maximizing its own payoff, subject to the constraint that the consumer is delivered a value of v_t^a (in expectation) that is promised to her at the beginning of the problem. This involves determining the optimal probabilities of termination (c_t and o_t), and the optimal value promised to the consumer at the end of the period (v_t^e) if the contract continues,[1] by solving the following problem:

$$\pi_t^a(v_t^a) = \max_{c_t\in[0,1],\, o_t\in[0,1-c_t],\, v_t^e\in\left[\underline{v}_t^e,\overline{v}_t^e\right]} \quad c_t\mathcal{C}_t + o_t\mathcal{O}_t + (1-c_t-o_t)\pi_t^e(v_t^e) \tag{7}$$

$$\text{s.t.} \quad v_t^a = c_t\underline{v}_t^e + o_t\overline{v}_t^e + (1-c_t-o_t)v_t^e. \tag{8}$$

The first two terms in (7) capture the firm's payoffs from termination and the last term captures its payoff from continuation. This objective is maximized subject to (8), the *promise-keeping* constraint, evident from the flow equation (5). From (8), $\underline{v}_t^a = \underline{v}_t^e$ and $\overline{v}_t^a = \overline{v}_t^e$. The following result characterizes the optimal termination and continuation decisions. Throughout the paper, for a concave function π, we interpret the statement "$\pi'(v) = k$" as "k is a super-derivative of π at v." This allows us to maintain the derivative notation for the functions in the model which possess kinks. For more details on this notation, see Appendix C.

Proposition 1. *If the following inequality is satisfied:*

$$\pi_t^e(v_t^a) \le (\mathcal{C}_t(\overline{v}_t^a - v_t^a) + \mathcal{O}_t(v_t^a - \underline{v}_t^a))/(\overline{v}_t^a - \underline{v}_t^a), \quad \forall\, v_t^a \in [\underline{v}_t^a, \overline{v}_t^a], \tag{9}$$

then it is optimal to terminate in period t by transferring ownership with probability $o_t^*(v_t^a) = (v_t^a - \underline{v}_t^a)/(\overline{v}_t^a - \underline{v}_t^a)$, and confiscating the device with probability $c_t^*(v_t^a) = 1 - o_t^*(v_t^a)$. If the condition in (9) is not satisfied, then the optimal confiscation and ownership probabilities and the optimal continuation value are given by

$$c_t^*(v_t^a) = \left[\hat{v}_t^{\mathcal{C}} - v_t^a\right]^+ / \left(\hat{v}_t^{\mathcal{C}} - \underline{v}_t^a\right) \quad \text{if } \mathcal{C}_t > \pi_t^e(\underline{v}_t^a), \quad \text{and } 0 \text{ otherwise}, \tag{10}$$

$$o_t^*(v_t^a) = \left[v_t^a - \hat{v}_t^{\mathcal{O}}\right]^+ / \left(\overline{v}_t^a - \hat{v}_t^{\mathcal{O}}\right) \quad \text{if } \mathcal{O}_t > \pi_t^e(\overline{v}_t^a), \quad \text{and } 0 \text{ otherwise}, \tag{11}$$

$$v_t^{e*}(v_t^a) = v_t^a + \left[\hat{v}_t^{\mathcal{C}} - v_t^a\right]^+ + \left[\hat{v}_t^{\mathcal{O}} - v_t^a\right]^-, \tag{12}$$

and the contract continues with probability $1 - c_t^*(v_t^a) - o_t^*(v_t^a)$. Here, the thresholds $\hat{v}_t^{\mathcal{C}}$ and $\hat{v}_t^{\mathcal{O}}$ are

$$\hat{v}_t^{\mathcal{C}} = \inf\left\{ v \in (\underline{v}_t^e, \overline{v}_t^e) \,\middle|\, \pi_t^{e'}(v) \le (\pi_t^e(v) - \mathcal{C}_t)/(v - \underline{v}_t^e) \right\}, \quad \text{and} \tag{13}$$

$$\hat{v}_t^{\mathcal{O}} = \sup\left\{ v \in (\underline{v}_t^e, \overline{v}_t^e) \,\middle|\, \pi_t^{e'}(v) \ge (\mathcal{O}_t - \pi_t^e(v))/(\overline{v}_t^e - v) \right\}. \tag{14}$$

Moreover, the payoff function $\pi_t^a(v_t^a)$ is concave in v_t^a.

We first discuss the geometric construction of the above result and interpret it later. The firm in our model has the ability to terminate probabilistically. By randomizing between confiscation (resp., transferring ownership) and continuation, the firm can obtain all payoffs that lie on any line segment between $(\underline{v}_t^a, \mathcal{C}_t)$ (resp., $(\overline{v}_t^a, \mathcal{O}_t)$) and any point on the graph of $\pi_t^e(\cdot)$, and by randomizing between confiscation and ownership

[1] The consumer has the option to quit and receive $\underline{v}_t^e$. It is easy to show that any contract that induces the consumer to quit in period t could be replaced with a contract that terminates in period t. Thus, we restrict attention to contracts where the consumer does not quit before termination.

transfer, the payoffs on the line between $(\underline{v}_t^a, \mathcal{C}_t)$ and $(\overline{v}_t^a, \mathcal{O}_t)$ can be obtained. To illustrate how such a payoff function can be constructed, pick a point $\hat{v}_t \in [\underline{v}_t^a, \overline{v}_t^a]$, see Figure 2(a). Then, for any $v_t^a \in (\underline{v}_t^a, \hat{v}_t)$, confiscating the device with probability $\tilde{c}_t(v_t^a) = (\hat{v}_t - v_t^a)/(\hat{v}_t - \underline{v}_t^a)$ and promising the consumer a continuation value of $\tilde{v}_t^e(v_t^a) = \hat{v}_t$ ensures that the promise-keeping constraint $v_t^a = \tilde{c}_t(v_t^a)\, \underline{v}_t^a + (1 - \tilde{c}_t(v_t^a))\, \hat{v}_t$ is satisfied, while giving the firm a payoff of $\tilde{\pi}_t^a(v_t^a) = \tilde{c}_t(v_t^a)\, \mathcal{C}_t + (1 - \tilde{c}_t(v_t^a))\, \pi_t^e(\hat{v}_t)$. Thus, given a $\hat{v}_t$, the firm can obtain any payoff on the line segment between $(\underline{v}_t^a, \mathcal{C}_t)$ and $(\hat{v}_t, \pi_t^e(\hat{v}_t))$. The question then becomes, which of the lines constructed in this way gives the highest payoff to the firm. In the case where the line between $(\underline{v}_t^a, \mathcal{C}_t)$ and $(\overline{v}_t^a, \mathcal{O}_t)$ is entirely above the graph of π_t^e (as reflected by condition (9), and illustrated in Figure 2(b)), the firm sees no value in continuing the contract, and it is optimal to simply randomize between confiscation and ownership transfer. Otherwise, the highest such lines are the tangents to the graph of π_t^e which pass through $\mathcal{C}_t$ and $\mathcal{O}_t$. The intersection points of such tangents and the graph are the $\hat{v}_t^{\mathcal{C}}$ and $\hat{v}_t^{\mathcal{O}}$ from Proposition 1, as illustrated in Figure 2(a). In both cases, the optimal payoff function $\pi_t^a(\cdot)$ constructed in this manner forms the upper hull of $(\underline{v}_t^a, \mathcal{C}_t)$, $(\overline{v}_t^a, \mathcal{O}_t)$ and the graph of $\pi_t^e(\cdot)$. Simple inspection of Figure 2(a) also reveals that this payoff for the firm, by construction, is higher than $\max\{\mathcal{C}_t, \pi_t^e(\cdot)\}$ and $\max\{\mathcal{O}_t, \pi_t^e(\cdot)\}$, the payoffs implied by the optimal deterministic confiscation and ownership transfer rules respectively.

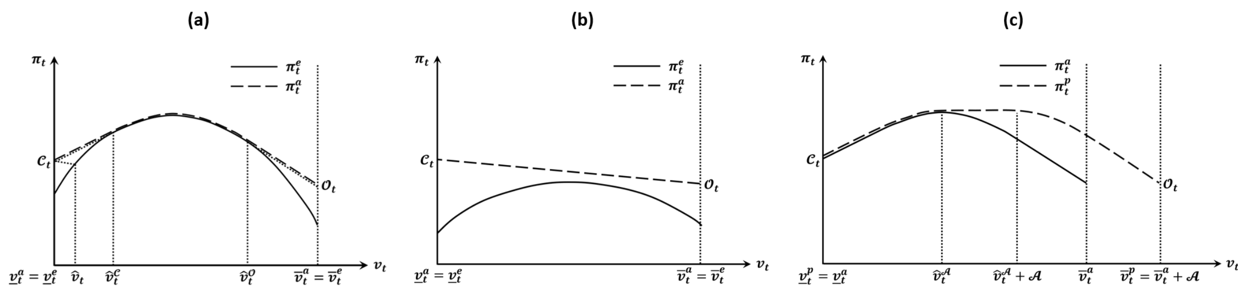

Figure 2 **Construction of π_t^a is shown in (a) when (9) is not satisfied, and in (b) when (9) is satisfied; and construction of π_t^p is shown in (c).**

It is noteworthy that in (10)–(12), we have changed the notation from Section 3.3 by writing c_t^* and o_t^*, along with v_t^{e*}, as functions of the state variable v_t^a alone, rather than the entire payment history h^t. This is because the consumer's continuation value captures the payoff-relevant information regarding the history. The contract offers threshold-based guidance for termination in Proposition 1. If the promised future value v_t^a falls below the *confiscation threshold* $\hat{v}_t^{\mathcal{C}}$, it is optimal for the firm to confiscate the device with probability c_t^*. Analogously, if v_t^a is above the *ownership threshold* $\hat{v}_t^{\mathcal{O}}$, it is optimal to grant ownership with probability o_t^*. Otherwise, the contractual relationship should continue. Moreover, c_t^* is decreasing in v_t^a, and o_t^* and v_t^{e*} are increasing in v_t^a: the higher the value promised to the consumer, the lower (resp., higher) the chance of confiscation (resp., ownership transfer), and the higher the value promised upon continuation. The concavity of the payoff function π_t^a in v_t^a follows by construction.

Finally, we emphasize that the probabilistic termination deployed in our model is more of an analytical convenience, and it need not be implemented exactly. If c_t^* is below (resp., above) a threshold, e.g., 0.5, then the firm can discretize c_t^* as 0 (resp., 1) and continue the contract with a promised value $\hat{v}_t^{\mathcal{C}}$ (resp., confiscate the device), which would be a deterministic approximation of the truly optimal policy. The discretization threshold can depend on any context-specific information that is not accounted for in the model. We can similarly interpret randomization between confiscation and ownership transfer in Figure 2(b).

4.2. Optimal Technology Access

We now compute $\pi_t^p(v_t^p)$ from $\pi_t^a(v_t^a)$. In this problem, the firm decides the fractional access f_t that must be granted to the consumer in period t, along with the continuation value v_t^a after granting access, subject to the constraint to deliver a total value of v_t^p:

$$\pi_t^p(v_t^p) = \max_{f_t \in [0,1],\, v_t^a \in \left[\underline{v}_t^a, \overline{v}_t^a\right]} \quad \pi_t^a(v_t^a) \tag{15}$$

$$\text{s.t.} \quad v_t^p = f_t A + v_t^a. \tag{16}$$

The firm's objective in (15) is to maximize its payoff by providing optimal access. The promise-keeping constraint (16) follows from (5) and reflects the trade-off faced by the firm: either give too much access now and promise too little later, or vice versa. From (16), $\underline{v}_t^p = \underline{v}_t^a$ and $\overline{v}_t^p = A + \overline{v}_t^a$. We have the following result.

Proposition 2. *The following fractional access and the continuation value are optimal:*

$$f_t^*(v_t^p) = \min\left\{ \left[v_t^p - \hat{v}_t^{\mathcal{A}}\right]^+ /A, 1\right\}, \quad \textit{and} \tag{17}$$

$$v_t^{a*}(v_t^p) = v_t^p - \min\left\{ \left[v_t^p - \hat{v}_t^{\mathcal{A}}\right]^+, A\right\}, \tag{18}$$

where $\hat{v}_t^{\mathcal{A}} \in \arg\max_{v \in \left[\underline{v}_t^a, \overline{v}_t^a\right]} \pi_t^a(v)$. *Moreover,* $\pi_t^p(v_t^p)$ *is concave in* v_t^p.

As shown in Figure 2(c), we obtain π_t^p by laterally shifting π_t^a by A units starting from its maximum value (achieved at $\hat{v}_t^{\mathcal{A}}$). The concavity of $\pi_t^p(v_t^p)$ follows by construction. Proposition 2 states that if the consumer's promised future value v_t^p is below the threshold $\hat{v}_t^{\mathcal{A}}$ (reflecting a bad payment history), then it is optimal for the firm to turn the device off ($f_t^* = 0$) in period t. If v_t^p is above the threshold $\hat{v}_t^{\mathcal{A}} + A$, then the consumer's payment history is good enough and the device is turned on ($f_t^* = 1$). Otherwise, providing *partial* access is optimal ($0 < f_t^* < 1$). In practice, partial access could translate to, for example, providing light only for a certain number of hours in a day, or for a certain number of days in a week. Both f_t^* and v_t^{a*}, in (17) and (18), are increasing in v_t^p: the higher the value promised to the consumer, the higher the access granted to her, and her value-to-go thereafter is also higher. Extending Proposition 2 to accommodate more general formulations of device-access utility is straightforward. For example, if $u_a(f) = \tilde{u}_a(f)A$, where $\tilde{u}_a$ is an increasing and concave function, then (17) gives us $\tilde{u}_a(f_t^*)$, instead of f_t^*.

Two additional features are suggested by the optimal contract. First, after providing access to the device, the consumer's promised future value must be lowered (by $f_t^* A$) to incentivize her to pay in the future; it will go up again after the consumer makes her next payment (see next subsection). Second, notice that if v_t^a is sufficiently high, f_t^* will be equal to 1, not only in the current period, but perhaps also in the next few periods. These are the *grace periods* in which the consumer is provided full access to the device even if she does not make any payment. Thus, the contract rewards a good payment history with grace periods.

4.3. Optimal Payment Incentives and Flexibility

This brings us to the most challenging part of solving for the optimal contract: determining $\pi_t^b(v_t^b)$ from $\pi_t^p(v_t^p)$. The firm in this problem decides the *optimal payment policy* that it would like to suggest to the consumer, and the *optimal reward policy*, in the form of promised future value, that it would like to offer in return for the consumer's payment. The technical challenge in this problem is that the payment policy p_t is contingent on the consumer's realized income y_t, and the reward policy v_t^p is contingent on the consumer's payment p_t. Unlike in Sections 4.1 and 4.2, where the firm solves for the *parameters* that are optimal, the firm here solves for the *functions* $p_t(y)$ and $v_t^p(p)$ that are optimal. We restrict the space of these functions to $\mathcal{P}_t \times \mathcal{V}_t$, where $\mathcal{P}_t$ is the set of all functions $p_t : \mathcal{Y}_t \to \mathbb{R}_+$ that are absolutely continuous, left- and right-differentiable, and satisfy $p_t(y) \le y$; and $\mathcal{V}_t$ is the set of all functions $v_t^p : \mathbb{R}_+ \to [\underline{v}_t^p, \overline{v}_t^p]$ that are absolutely continuous and left- and right-differentiable. Formally, the firm solves:

$$\pi_t^b(v_t^b) = \max_{p_t(\cdot)\in\mathcal{P}_t,\, v_t^p(\cdot)\in\mathcal{V}_t} \quad \mathbb{E}\big[p_t(Y_t) + \pi_t^p(v_t^p(p_t(Y_t)))\big] \tag{19}$$

$$\text{s.t.} \quad p_t(y_t) \in \operatorname*{argmax}_{0\le p\le y_t} u(y_t - p) + v_t^p(p), \quad \forall\, y_t \in \mathcal{Y}_t, \quad \text{and} \tag{20}$$

$$v_t^b = \mathbb{E}\big[u(Y_t - p_t(Y_t)) + v_t^p(p_t(Y_t))\big]. \tag{21}$$

The firm's objective in (19) consists of two components: the payment p_t that it gets now, and the continuation payoff π_t^p that it gets later from promising reward v_t^p. The incentive compatibility constraint in (20) states that the suggested payment function $p_t(\cdot)$ is consistent with the consumer's preference maximization. Notice in (20) that the firm *shapes* the consumer's preferences by determining an appropriate $v_t^p(\cdot)$ function that aligns her preferences with its own. In other words, the payment and reward policies together incentivize the consumer to optimally split her income between consumption and payment. This payment problem, from the consumer's viewpoint in (20), is always a single-period problem—an intrinsic feature of the promised future utility framework. We will see later that there could be multiple payment functions that maximize the consumer's preferences; we therefore follow the convention in the contract theory literature that out of all these payment functions, the firm picks the one that maximizes its own payoff. The promise-keeping constraint (21) follows from (5), and ensures that the consumer is delivered v_t^b promised at the beginning. From (21), $\underline{v}_t^b = \underline{v}_t^p$ and $\overline{v}_t^b = U_t + \overline{v}_t^p$.

We solve the above problem in two steps. First, we determine the shapes of the optimal payment and reward functions, which convert the optimization problem over a space of functions into one over a space of parameters. Second, with this reformulation, we show that $\pi_t^b(v_t^b)$ is concave in v_t^b. In the analysis, we assume that $\mathcal{Y}_t = \mathbb{R}_+$. Admittedly, assuming that the domain of income y_t spans till ∞ is unrealistic. However, this assumption simplifies exposition, and more importantly, imposing an arbitrary lower bound and a sufficiently high upper bound on $\mathcal{Y}_t$ does not qualitatively affect the features of the optimal contract.

Shapes of optimal incentive functions. The following result states that the optimal payment and reward functions can be written simply as functions of two parameters α_t and κ_t^v.

Proposition 3. *It is optimal to choose the following reward function:*

$$v_t^{p*}(p\,;\alpha_t,\kappa_t^v) = \begin{cases} \alpha_t + \lambda_1 p + (\lambda_2 - \lambda_1)[p - \kappa_t^v]^+ & \textit{for} \quad 0 \le p < \hat{p}_{0,t}(\alpha_t,\kappa_t^v), \\ \overline{v}_t^p & \textit{for} \quad p \ge \hat{p}_{0,t}(\alpha_t,\kappa_t^v), \end{cases} \tag{22}$$

where (a) $\alpha_t \geq \underline{v}_t^p$*, and (b)* $\hat{p}_{0,t}(\alpha_t, \kappa_t^v)$ *solves the equation* $\lambda_1 p + (\lambda_2 - \lambda_1)[p - \kappa_t^v]^+ = \overline{v}_t^p - \alpha_t$*. Given the reward function* $v_t^{p*}(p\,;\alpha_t, \kappa_t^v)$*, the following payment function* $p_t^*(\cdot)$ *is optimal:*

$$p_t^*(y\,;\alpha_t,\kappa_t^v) = \begin{cases} y & \text{for} \quad 0 \leq y < \hat{p}_{1,t}(\alpha_t,\kappa_t^v), \\ \hat{p}_{1,t}(\alpha_t,\kappa_t^v) & \text{for} \quad \hat{p}_{1,t}(\alpha_t,\kappa_t^v) \leq y < \hat{p}_{1,t}(\alpha_t,\kappa_t^v) + \kappa, \\ y - \kappa & \text{for} \quad \hat{p}_{1,t}(\alpha_t,\kappa_t^v) + \kappa \leq y < \hat{p}_{2,t}(\alpha_t,\kappa_t^v) + \kappa, \\ \hat{p}_{2,t}(\alpha_t,\kappa_t^v) & \text{for} \quad y \geq \hat{p}_{2,t}(\alpha_t,\kappa_t^v) + \kappa, \end{cases} \tag{23}$$

where the thresholds are

$$\hat{p}_{1,t}(\alpha_t,\kappa_t^v) = \min\left\{[\Lambda_{1,t} - \alpha_t]^+/\lambda_1, \kappa_t^v, \hat{p}_{0,t}(\alpha_t,\kappa_t^v)\right\}, \tag{24}$$

$$\hat{p}_{2,t}(\alpha_t,\kappa_t^v) = \max\left\{(\Lambda_{2,t} - \alpha_t - (\lambda_1 - \lambda_2)\kappa_t^v)/\lambda_2, \min\left\{\kappa_t^v, \hat{p}_{0,t}(\alpha_t,\kappa_t^v)\right\}\right\}, \tag{25}$$

and $\Lambda_{j,t} = \max\left\{\underline{v}_t^p, \sup\left\{v \in [\underline{v}_t^p, \overline{v}_t^p] \mid \pi_t^{p'}(v) \geq -1/\lambda_j\right\}\right\}$*, for* $j \in \{1,2\}$*.*

We now pictorially demonstrate why the shapes of the optimal reward and payment functions are as stated in Proposition 3. We can see from (20) that different v_t^p functions will induce different consumer preferences over the bundles (x,p), where x is the amount consumed and p is the amount paid, which are subject to the budget constraint $x + p = y$. For now, fix $v_t^p = v_t^{p*}$ given in (22). Note that this reward function mimics the shape of the consumption utility function (piecewise linear and concave with only two possible slopes: λ_1 and λ_2), and is sandwiched between $\underline{v}_t^p$ and $\overline{v}_t^p$. Figure 3(a) plots the indifference curves $u(x) + v_t^{p*}(p) = k_j$, for $k_1 < \cdots < k_4$, where k_j is the value derived from the bundle (x,p) that is constant on a given indifference curve. We know from basic microeconomics that an optimal bundle lies on the highest indifference curve that is tangential to the budget line (Mas-Colell et al. 1995). Figure 3(a) also plots the tangential budget lines $x + p = y_j$ for $y_1 < \ldots < y_4$. These lines have a slope of -1 in the (x,p) plane. Thus, for a given y_j, every bundle that is on the segment with slope -1 of the corresponding indifference curve is optimal for the consumer. On such segments, we denote the payment in the leftmost bundle by $\overline{p}_t(y)$ and in the rightmost bundle by $\underline{p}_t(y)$. We plot $\underline{p}_t(y)$ and $\overline{p}_t(y)$ as functions of y on the right-hand side of Figure 3(b).[2] The space between these two functions contains all the incentive-compatible payment functions, and the firm must choose the optimal one (i.e., $p_t^*(y)$) from this space.

To understand the shape of $p_t^*(y)$, we focus on the region in Figure 3(b) where $0 \leq y \leq \kappa_t^v + \kappa$. To obtain the optimal payment function in this region, the firm maximizes its payoff pointwise for each y, i.e., $p_t^*(y) = \max_{\underline{p}_t(y) \leq p \leq \overline{p}_t(y)} p + \pi_t^p(v_t^{p*}(p))$. Given that π_t^p is concave and $v_t^{p*}(p) = \alpha_t + \lambda_1 p$ in this region, the derivative of the payoff function, $1 + \pi_t^{p'}(\alpha_t + \lambda_1 p)\lambda_1$, is monotonically decreasing in p (see left-hand side of Figure 3(b)). Therefore, the optimal $p_t^*(y)$, as demonstrated in Figure 3(b), is either at one of the boundaries $\overline{p}_t(y)$ and $\underline{p}_t(y)$, or is an interior solution $\hat{p}_{1,t}$ that solves $\pi_t^{p'}(\alpha_t + \lambda_1 p) = -1/\lambda_1$ or equivalently $\alpha_t + \lambda_1 p = \Lambda_{1,t}$. Proposition 3 defines $\hat{p}_{1,t}$ and $\Lambda_{1,t}$ more generally such that they do not violate the boundary constraints. We can similarly interpret the shape of $p_t^*(y)$ for $y \geq \kappa_t^v + \kappa$ in Figure 3(b).

The above arguments show that $p_t^*(y)$ is optimal, given that the reward function is $v_t^{p*}(p)$. We now demonstrate the optimality of v_t^{p*} itself. As we know, this function has only two possible slopes. Let us

[2] The formal definitions of $\underline{p}_t(y)$ and $\overline{p}_t(y)$ are as follows: denote by $\hat{p}_t = \min\{\kappa_t^v, \hat{p}_{0,t}\}$, then (i) $\underline{p}_t(y)$ is equal to 0 for $0 \leq y < \kappa$; $y - \kappa$ for $\kappa \leq y < \hat{p}_t + \kappa$; and $\hat{p}_t$ for $y \geq \hat{p}_t + \kappa$; and (ii) $\overline{p}_t(y)$ is equal to y for $0 \leq y < \hat{p}_t$; $\hat{p}_t$ for $\hat{p}_t \leq y < \hat{p}_t + \kappa$; $y - \kappa$ for $\hat{p}_t + \kappa \leq y < \hat{p}_{0,t} + \kappa$; and $\hat{p}_{0,t}$ for $y \geq \hat{p}_{0,t} + \kappa$.

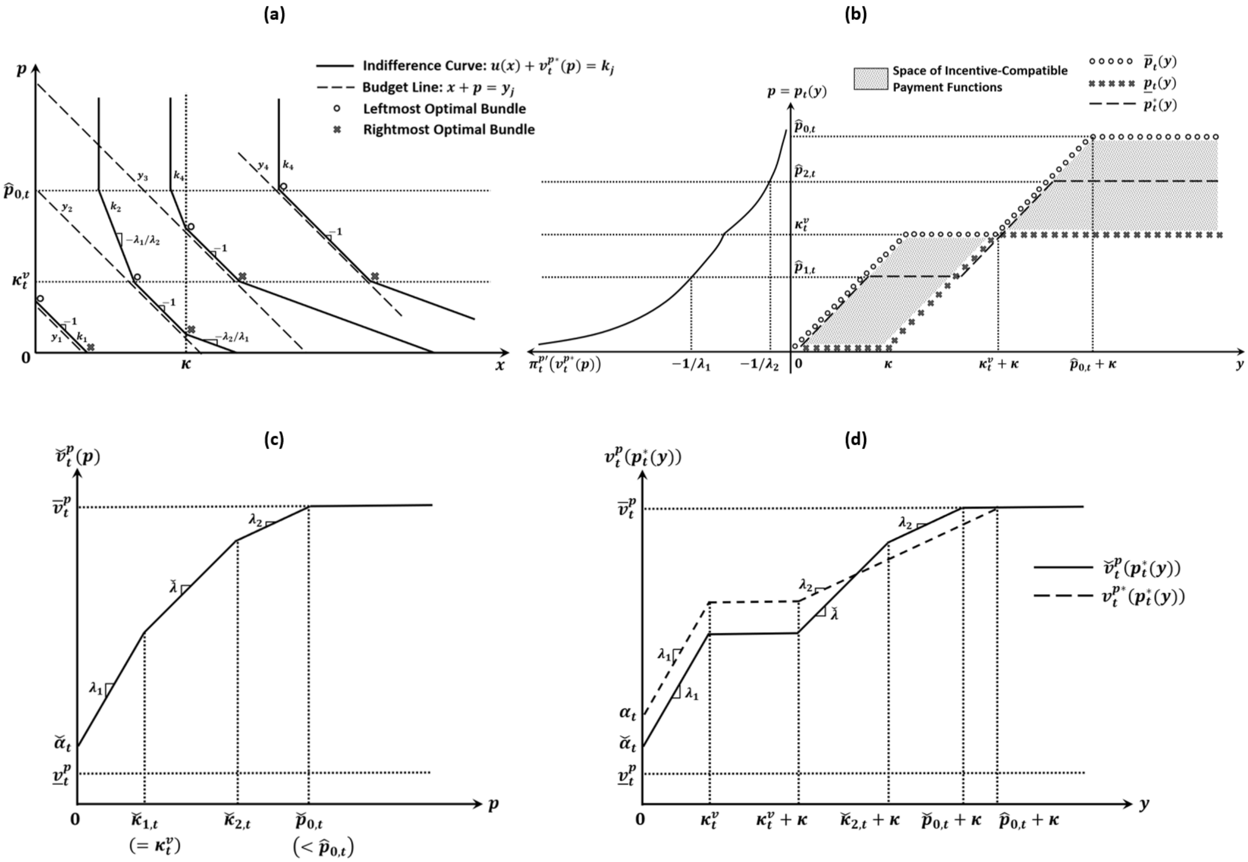

Figure 3 **Construction of optimal payment and reward functions: (a) Indifference curves $u(x)+v_t^{p*}(p)=k_j$, along with the tangential budget lines $x+p=y_j$, shown in the (x,p) plane for $j\in\{1,\dots,4\}$; (b) on the right-hand side is the space of incentive-compatible payment functions sandwiched between $\underline{p}_t(y)$ and $\overline{p}_t(y)$, and on the left-hand side is $\pi_t^{p'}(v_t^{p*}(y))$; (c) an alternative reward function $\check{v}_t^p$ that has an intermediate slope $\check{\lambda}$; and (d) $\check{v}_t^p(p_t^*(y))$ and $v_t^{p*}(p_t^*(y))$ plotted together.**

examine what happens under an alternative design of v_t^p that includes an additional intermediate slope $\check{\lambda}\in(\lambda_1,\lambda_2)$. We call it $\check{v}_t^p(p)$, plotted in Figure 3(c). Following the logic that we discussed above, it is easy to see that both $\check{v}_t^p$ and v_t^{p*} induce the same payment function $p_t^*(y)$; then by (21), we have $\mathbb{E}\check{v}_t^p(p_t^*(y))=\mathbb{E}v_t^{p*}(p_t^*(y))$. Figure 3(d) plots $\check{v}_t^p(p_t^*(y))$ and $v_t^{p*}(p_t^*(y))$, and it is clear that the former is a mean-preserving spread of the latter. Because π_t^p is concave, the firm prefers the reward function that results in a lower payoff variance (see (19)), which in this case is v_t^{p*}. This logic continues to hold more generally, see the proof in Appendix C. The reward function v_t^{p*} results in the lowest payoff variance when compared to alternative designs of v_t^p, and hence it becomes optimal.

Concavity of the firm's payoff. It is noteworthy that with Proposition 3 in hand, the optimization problem (19)–(21) reduces to the following with only two parameters α_t and κ_t^v:

$$\pi_t^b(v_t^b)=\max_{\alpha_t\in\left[\underline{v}_t^p,\overline{v}_t^p\right],\,\kappa_t^v\in\mathbb{R}_+}\ \mathbb{E}\left[p_t^*(Y_t\,;\alpha_t,\kappa_t^v)+\pi_t^p(v_t^{p*}(p_t^*(Y_t\,;\alpha_t,\kappa_t^v)\,;\alpha_t,\kappa_t^v))\right] \tag{26}$$

$$\text{s.t.}\quad v_t^b=\mathbb{E}\left[u(Y_t-p_t^*(Y_t\,;\alpha_t,\kappa_t^v))+v_t^{p*}(p_t^*(Y_t\,;\alpha_t,\kappa_t^v)\,;\alpha_t,\kappa_t^v)\right]. \tag{27}$$

It remains to show that $\pi_t^b(v_t^b)$ is concave in v_t^b. This continues to be a challenging exercise even with the reduced problem (26)–(27), because the maximand need not always be jointly concave in $(v_t^b, \alpha_t, \kappa_t^v)$. To make the problem well-behaved, we take the following three steps. First, we assume that the consumer's income distribution has a weakly decreasing hazard rate. In our context, this is not an unreasonable assumption, because decreasing hazard rate implies that (i) the pdf g_t is decreasing, and (ii) the distribution has a fatter tail (compared to exponential distribution). Given that the consumer is cash-constrained, she is more likely to have low disposable income (as reflected by (i)); but given the consumer's complex portfolios, she might have high realizations of income now and then (as reflected by (ii)).

Second, we use a change of variable technique and reformulate (27) without loss of optimality as follows:

$$v_t^b = \alpha_t + \omega_t, \quad \text{where} \tag{28}$$

$$\omega_t = \int_0^{\kappa_t^v + \kappa} \lambda_1 y \, dG_t(y) + \int_{\kappa_t^v + \kappa}^{\infty} \left[(\lambda_1 - \lambda_2)(\kappa_t^v + \kappa) + \lambda_2 y\right] dG_t(y). \tag{29}$$

Equation (28) splits the promised pie of size v_t^b into two components: a fixed component α_t that is independent of the consumer's payment, and an average ω_t of the variable component that is contingent on the consumer's payment. The minimum value of parameter ω_t is $\underline{\omega}_t = U_t$ (when $\kappa_t^v = 0$) and its maximum is $\overline{\omega}_t = \lambda_1 \mathbb{E} Y_t$ (as $\kappa_t^v \to \infty$). Using (28), we can replace α_t with $v_t^b - \omega_t$. Moreover, the right hand side of (29), as a function of κ_t^v, is monotonically increasing in κ_t^v. Therefore, one can invert that function to obtain the value of κ_t^v that satisfies (29) for any given value of ω_t. With a slight abuse of notation, we represent this inverted function as $\kappa_t^v(\omega_t)$, which is increasing in ω_t. With these changes, the problem (26)–(27) can be written as:

$$\pi_t^b(v_t^b) = \max_{\omega_t \in [\underline{\omega}_t, \overline{\omega}_t]} \tilde{\pi}_t^b(v_t^b, \omega_t), \quad \text{where} \tag{30}$$

$$\tilde{\pi}_t^b(v_t^b, \omega_t) = \mathbb{E}\left[p_t^*(Y_t\,;\, v_t^b - \omega_t,\, \kappa_t^v(\omega_t)) + \pi_t^p(v_t^{p*}(p_t^*(Y_t\,;\, v_t^b - \omega_t,\, \kappa_t^v(\omega_t))\,;\, v_t^b - \omega_t,\, \kappa_t^v(\omega_t)))\right]. \tag{31}$$

Third, we replace the function π_t^p in (31) with an approximation that is at least twice differentiable. We do so because π_t^p, in general, might have kinks. This makes the second-order analysis, which is necessary to show the concavity of π_t^b, quite cumbersome. This replacement is not restrictive however because any univariate concave function on a closed interval can be approximated arbitrarily closely by a smooth concave function (Azagra 2013). Thus, we will be able to approximate π_t^p, and hence $\tilde{\pi}_t^b$ and π_t^b also, to any desired level of precision. To reduce the notational burden, we refer to the approximations of $\tilde{\pi}_t^b$ and π_t^b as those functions themselves (we treat them more carefully in Appendix C). Now, we obtain the following result.

Proposition 4. *Define* $\omega_t^*(v_t^b) = \arg\max_{\omega_t \in [\underline{\omega}_t, \overline{\omega}_t]} \tilde{\pi}_t^b(v_t^b, \omega_t)$*. For a given* v_t^b*, there exists a unique* $\omega_t^*(v_t^b)$*. The following statements hold:*

(i) *There exists* $\hat{v}_t^b \in [\underline{v}_t^b, \overline{v}_t^b]$ *such that for* $v_t^b \geq \hat{v}_t^b$*,* $\omega_t^*(v_t^b) = \underline{\omega}_t$*. For* $v_t^b < \hat{v}_t^b$*,* $(v_t^b, \omega_t^*(v_t^b))$ *lies in the region* $\mathfrak{R} = \left\{(v_t^b, \omega_t) \in [\underline{v}_t^b, \hat{v}_b^t] \times [\underline{\omega}_t, \overline{\omega}_t] \,\middle|\, \omega_t \leq v_t^b - \underline{v}_t^b,\, \omega_t - \lambda_1 \kappa_t^v(\omega_t) \geq v_t^b - \Lambda_{1,t}\right\}$*. The region* $\mathfrak{R}$ *is convex.*

(ii) $\pi_t^b(v_t^b) = \tilde{\pi}_t^b(v_t^b, \omega_t^*(v_t^b))$ *is concave in* v_t^b*.*

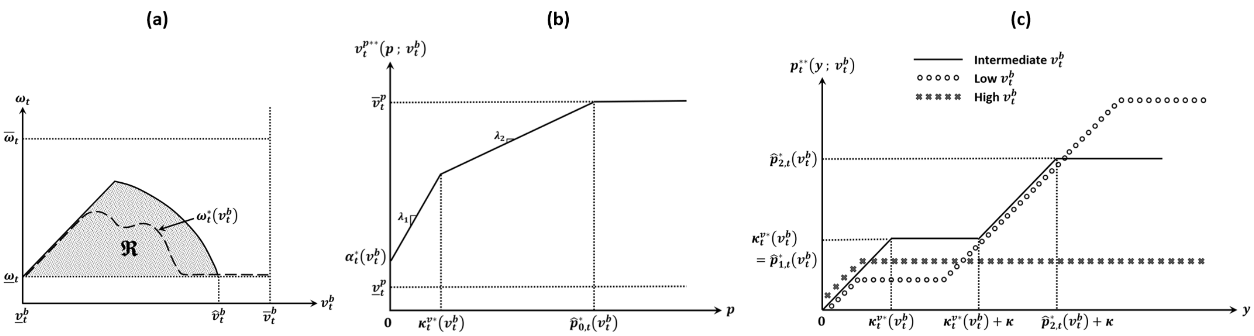

Figure 4 (a) Region $\mathfrak{R}$ and a plausible $\omega_t^*(v_t^b)$ in the (v_t^b, ω_t) plane; (b) the shape of the optimal reward function; and (c) the shapes of the optimal payment functions for different magnitudes of v_t^b.

Some comparative statics. We now proceed to understand how the incentives are shaped as a function of v_t^b, the current standing of the consumer. To that end, we define the following functions: (i) $\alpha_t^*(v_t^b) = v_t^b - \omega_t^*(v_t^b)$, (ii) $\kappa_t^{v*}(v_t^b) = \kappa_t^v(\omega_t^*(v_t^b))$, (iii) $\hat{p}_{j,t}^*(v_t^b) = \hat{p}_{j,t}(\alpha_t^*(v_t^b), \kappa_t^{v*}(v_t^b))$ for $j \in \{0,1,2\}$, (iv) $v_t^{p**}(p\,;\,v_t^b) = v_t^{p*}(p\,;\,\alpha_t^*(v_t^b), \kappa_t^{v*}(v_t^b))$, and (v) $p_t^{**}(y\,;\,v_t^b) = p_t^*(y\,;\,\alpha_t^*(v_t^b), \kappa_t^{v*}(v_t^b))$.

To characterize any of the above functions of v_t^b, we must first understand the behavior of $\omega_t^*(v_t^b)$. As per Proposition 4(i), Figure 4(a) plots the region $\mathfrak{R}$ and a plausible $\omega_t^*(v_t^b)$ in the (v_t^b, ω_t) plane. For $v_t^b \geq \hat{v}_t^b$, $\omega_t^*(v_t^b)$ is equal to the lowest possible value $\underline{\omega}_t$. For $v_t^b < \hat{v}_t^b$, we cannot analytically pin down the shape of $\omega_t^*(v_t^b)$. Yet, the shape of the region $\mathfrak{R}$ suggests that $\omega_t^*(v_t^b)$ must first increase for smaller values of v_t^b (with $\omega_t^*(\underline{v}_t^b) = \underline{\omega}_t$), and that it must decrease as v_t^b approaches $\hat{v}_t^b$. Accordingly, we say that $\omega_t^*(v_t^b)$ exhibits a $\cap$-pattern in v_t^b. By $\cap$-pattern, we do *not* mean that $\omega_t^*(v_t^b)$ is unimodal in v_t^b. Instead, we mean that the function certainly takes smaller values as we approach both ends of its domain when compared to its values in the intermediate part of its domain.

Recall from (28) that ω_t is the component that incentivizes the consumer to make payment. The $\cap$–pattern of $\omega_t^*(v_t^b)$ suggests that it is optimal to offer weaker payment incentives when the consumer's v_t^b is either too low or too high, and stronger incentives for intermediate values of v_t^b. If a consumer's payment history is bad she is pushed closer to the confiscation boundary, and conversely, when it is good she moves closer to ownership, and that itself is incentive enough. But when a consumer is somewhere in between the two extremes, she requires stronger payment incentives. We dub this structure "strong-in-the-middle" incentives.

Such a strong-in-the-middle structure can be envisaged in several other settings too. If an employee is close to either getting fired or getting promoted, that itself provides sufficient incentives, but if she is in the middle of her career progression, she might require stronger incentives to perform. Any task that takes a long time to finish and can potentially have a slump in the middle also might require strong motivation in the middle. Such tasks are pervasive: writing a research paper, becoming an expert sportsman or an artist, climbing a mountain, following through with a diet plan, and finishing a Ph.D. program. With this structure in mind, we present a few more results in the proposition below.

Proposition 5. *The following statements hold:*

(i) $\alpha_t^*(v_t^b)$ *is increasing in* v_t^b.

(ii) $\kappa_t^{v*}(v_t^b)$ *exhibits* $\cap$*–pattern in* v_t^b, *with* $\kappa_t^{v*}(\underline{v}_t^b) = 0$ *and* $\kappa_t^{v*}(v_t^b) = 0$ *for* $v_t^b \geq \hat{v}_t^b$.

(iii) $\hat{p}^*_{0,t}(v^b_t)$ *is decreasing in* v^b_t*, and* $\min\{\kappa^{v*}_t(v^b_t), \hat{p}^*_{0,t}(v^b_t)\} = \kappa^{v*}_t(v^b_t)$.

(iv) $\hat{p}^*_{1,t}(v^b_t) = \kappa^{v*}_t(v^b_t)$*, and hence* $\hat{p}^*_{1,t}(v^b_t)$ *exhibits* $\cap$*–pattern in* v^b_t.

(v) $\hat{p}^*_{2,t}(v^b_t)$ *is decreasing in* v^b_t.

(vi) *There exists* $\grave{v}^b_t \le \hat{v}^b_t$ *such that* $\hat{p}^*_{2,t}(v^b_t) - \hat{p}^*_{1,t}(v^b_t)$ *is decreasing in* v^b_t *for both* $v^b_t \le \grave{v}^b_t$ *and* $v^b_t \ge \hat{v}^b_t$.

Following the above results, we plot $v^{p**}_t(p\,;v^b_t)$ and $p^{**}_t(y\,;v^b_t)$ in Figures 4(b) and 4(c) respectively. As expected, v^{p**}_t is increasing in p; the higher the payment, the higher the promised future reward. When compared to the promised reward at the beginning (i.e., v^b_t), the updated promised reward (i.e., v^{p**}_t) is *not* necessarily always an increment. For example, consider the case when the consumer makes no payment: $v^{p**}_t(0\,;v^b_t) - v^b_t = \alpha^*_t(v^b_t) - v^b_t = -\omega^*_t(v^b_t) < 0$. Small payments from the consumer could either indicate low income realizations or large income diversions (the core of the agency issue). The contract moderates the latter issue by dropping (resp., raising) the promised future reward for relatively small (resp., large) payments.

$\alpha^*_t(v^b_t)$ is the fixed, payment-independent, part of the promised reward. Proposition 5(i) states that it is optimal for the firm to deliver more value through the fixed component as the consumer's payment history improves. This can be operationalized by entitling the consumer to a higher number of grace periods—wherein she can access the device without any payment—as her v^b_t increases.

Because $\kappa^{v*}_t(v^b_t)$ is the kink at which the reward function changes slope from λ_1 to λ_2, we say that the incentives become sharper as κ^{v*}_t increases (and blunter as it decreases). As per Proposition 5(ii), $\kappa^{v*}_t(v^b_t)$ exhibits $\cap$–pattern: for smaller and larger values of v^b_t, the consumer is rewarded for her payment at rate λ_2, and the segment with reward rate λ_1 is more applicable for the intermediate values of v^b_t. This is how the optimal contract operationalizes the strong-in-the-middle incentives; by making the reward function blunter for the consumers with good and bad histories, and sharper otherwise. Furthermore, $\hat{p}^*_{0,t}(v^b_t)$ is the threshold payment beyond which the consumer is promised the highest possible value, i.e., the ownership transfer. Proposition 5(iii) states that $\hat{p}^*_{0,t}(v^b_t)$ is decreasing in v^b_t. A consumer with a relatively good payment history has already made more (and regular) payments and hence must be closer to ownership entitlement.

The suggested payment function $p^{**}_t(y\,;v^b_t)$ in Figure 4(c) is increasing in y. We call this a flexible payment scheme; it suggests the consumer pay a smaller amount when income is low, and a larger amount when it is high. In other words, it entertains *no payment* in the current period, as well as *advance payment* for future periods. The payment function is incentive-compatible, and it also respects the consumer's liquidity constraints, thereby not burdening the consumer with borrowing to cover payments. Interestingly, the firm does not suggest arbitrarily high amounts to be paid in advance, even though we assumed the support of Y_t to be $\mathbb{R}_+$, because there is an upper limit on the value that the firm can deliver to the consumer. The maximum suggested payment $\hat{p}^*_{2,t}(v^b_t)$ is decreasing in v^b_t (see Proposition 5(v)) for the same reason $\hat{p}^*_{0,t}(v^b_t)$ is decreasing in v^b_t. Interestingly, the highest promised reward $v^{p**}_t(\hat{p}^*_{2,t})$ need not always be equal to $\overline{v}^b_t$. In some cases the firm will never promise ownership transfer, resulting in a purely rental business model. We discuss this in more detail in Appendix B.

Notice in Figure 4(c) that the magnitude of $\hat{p}^*_{1,t}(v^b_t)$ is indicative of the payments that the firm suggests for relatively low realizations of income. As per Proposition 5(iv), $\hat{p}_{1,t}(v^b_t)$ exhibits a strong-in-the-middle structure. The firm suggests the consumer pay relatively higher amounts even under low realizations of income (and also rewards those payments at a higher rate) when her v_t^b takes an intermediate value.

Finally, Proposition 5(vi) states that $\hat{p}_{2,t}^*(v_t^b) - \hat{p}_{1,t}^*(v_t^b)$ is decreasing in v_t^b for both smaller and larger values of v_t^b. (We are unable to theoretically pin down the shape of $\hat{p}_{2,t}^*(v_t^b) - \hat{p}_{1,t}^*(v_t^b)$ for $\grave{v}_t^b < v_t^b < \hat{v}_t^b$. But we know that $\hat{p}_{1,t}^*$ is relatively higher for the intermediate values of v_t^b and $\hat{p}_{2,t}^*$ is decreasing in v_t^b, so we surmise that the function is decreasing in this region as well.) We interpret $\hat{p}_{2,t}^*(v_t^b) - \hat{p}_{1,t}^*(v_t^b)$ as the amount of flexibility that the contract offers as a function of v_t^b. If this quantity is zero, then for a large subset of the domain of Y_t, the firm suggests the consumer pay the same amount irrespective of her income realization (see the function with crosses in Figure 4(c)). However, if this quantity is large, then the firm induces the consumer to pay different amounts depending on her income realizations for a wide range of Y_t values (as in the function with circles in Figure 4(c)). As the consumer's payment history improves, it is optimal to offer her lower flexibility. To understand this result, it is important to recognize that flexibility in a payment scheme is required in the first place to accommodate consumer's income uncertainty, and to also take advantage of her high income realizations in the current period to hedge against plausible low income realizations in future periods. Consequently, a consumer with low v_t^b is incentivized to pay higher amounts when the income realizations are high to move her away from the confiscation boundary. In contrast, for a consumer with high v_t^b, there is no longer a strong need to capture her income in the current period because she has already paid a large amount, possibly to cover for her potential non-payment periods in the future.

Remark 3. We have assumed in our analysis that the consumer's utility $u(\cdot)$ is piecewise linear and concave with one kink. By extending the arguments made in this section, it should be evident that if $u(\cdot)$ instead has $k > 1$ kinks, it will result in a piecewise linear and concave reward function with k kinks and a suggested payment function with $k+1$ steps (the corresponding optimization problem will have $k+1$ parameters). We can approximate any smooth concave utility function arbitrarily closely by increasing the number of kinks in a piecewise linear utility function. The resultant reward and payment functions also tend to become smoother concave functions as the number of kinks increases.

4.4. Last Period

In the last period, as the contract does not continue further, $\pi_T^e(v_T^e) = 0$ for all v_T^e. This results in a scenario similar to the one shown in Figure 2(b). So we have:

$$\pi_T^a(v_T^a) = \mathcal{C}_T(\overline{v}_T^a - v_T^a)/(\overline{v}_T^a - \underline{v}_T^a) + \mathcal{O}_T(v_T^a - \underline{v}_T^a)/(\overline{v}_T^a - \underline{v}_T^a). \tag{32}$$

Proposition 6. *$\pi_T^b(v_T^b)$, computed from $\pi_T^a(v_T^a)$ in (32), is concave in v_T^b.*

4.5. Contract Initiation

The problem of initiating the contract involves determining (i) the downpayment d that the firm asks a consumer to pay in period 0 to start the contract, and (ii) upon initiation, the value v_1^b that the firm promises the consumer at the beginning of period 1.

Recall from Section 3.2 that the consumer's ability to make the downpayment is w; we refer to it hereafter as the consumer's *wealth* (not the valuation of all the assets that the consumer holds, but only the liquid portion that could be used to make the downpayment). The distribution of wealth in the market is characterized by the random variable W. We can think of a market in our model as a set of consumers living in a village with similar income profiles but some might have better access to liquid assets than others, or some are ready to partly forego their consumption in the short run to acquire the technology while others are not. Accordingly, we assume that the random variables $\{Y_t\}$ characterize income profiles of all the consumers in the market, whereas the wealth of those consumers in period 0 could be different and are realizations of the random variable W, with support $\mathbb{R}_+$. We assume that the firm cannot perfectly observe a consumer's wealth, but it knows the distribution of W through market surveys.

To model the consumer's take-up decision, we denote her period 0 utility function by $\tilde{u}(x)$. We say that the consumer's value from taking up the contract is $\tilde{u}(w-d)+\delta v_1^b$, whereas the value from not taking it up is $\tilde{u}(w)+\delta \underline{v}_1^b$. We can thus interpret the function $\tilde{u}$ as the value that the consumer gains from utilizing her wealth at will, and taking up the contract reduces her wealth by d. We assume $\tilde{u}$ to have the same structure as u but with different parameters. Formally, $\tilde{u}(x)=\tilde{\lambda}_1 x+\left(\tilde{\lambda}_2-\tilde{\lambda}_1\right)[x-\tilde{\kappa}]^+$. (In our model, W could be the same as Y_0, and $\tilde{u}$ could be the same as u, but we represent them separately because the consumer may manage her money differently to make the downpayment compared to period-wise payments. The purpose of the former is different, i.e., to acquire the technology, and it can also be relatively larger in magnitude.)

We compare the firm's payoff under two methods of initiating the contract: one in which the firm charges a single value of downpayment d to all its consumers and offers them all a promised value of v_1^b, and the other in which the firm offers a menu $\{(d,v_1^b)\}$ and a consumer chooses the menu item that is optimal for her.

First method: The firm's problem in this case is as follows:

$$\pi_0^{(1)}=\max_{d\in\mathbb{R}_+,\, v_1^b\in\left[\underline{v}_1^b,\overline{v}_1^b\right]} \left(d+\gamma\pi_1^b(v_1^b)-\mathcal{K}\right)\times \Pr\left\{W\geq d,\, \tilde{u}(W-d)+\delta v_1^b\geq \tilde{u}(W)+\delta \underline{v}_1^b\right\}. \tag{33}$$

The objective in (33) is the firm's expected payoff, wherein the first term in the product is the revenue that the firm gains from a consumer who takes up the contract minus the cost of the technology, and the second term is the probability that a randomly chosen consumer in the market takes up the contract. Upon takeup, the firm's payoff is d in period 0, and $\pi_1^b(v_1^b)$ in period 1. A consumer in the market takes up the contract if and only if (i) her ability to pay w exceeds downpayment d, and (ii) she is willing to pay for the contract because her value from being in the contract exceeds her outside value.

Second method: The firm designs a *downpayment scheme* that will encourage a consumer with wealth w to pay $d(w)$ for the initiation of a contract that promises her a value of $v_1^b(d(w))$. The firm has the option to turn away some consumers who cannot afford a minimum downpayment; thus, the firm also chooses $\underline{d}\in\bar{\mathbb{R}}_+$, a minimum acceptable downpayment, where $\bar{\mathbb{R}}_+=\mathbb{R}_+\cup\{\infty\}$. (We allow $\underline{d}=\infty$, which is equivalent to not serving the market at all.) The firm solves the following problem:

$$\pi_0^{(2)}=\max_{d(\cdot)\in\mathcal{D},\, v_1^b(\cdot)\in\mathcal{V}_0,\, \underline{d}\in\bar{\mathbb{R}}_+} \mathbb{E}\left[d(W)+\gamma\pi_1^b(v_1^b(d(W)))-\mathcal{K}\;\middle|\; d(W)\geq \underline{d}\right]\times \Pr\left\{d(W)\geq \underline{d}\right\} \tag{34}$$

$$\text{s.t.}\quad d(w)\in \operatorname*{arg\,max}_{0\leq d\leq w} \begin{cases}\tilde{u}(w-d)+\delta v_1^b(d) & \text{if } d\geq \underline{d},\\ \tilde{u}(w)+\delta\underline{v}_1^b & \text{otherwise,}\end{cases} \quad \forall\, w\in\mathbb{R}_+. \tag{35}$$

The term in the brackets in (34) is the firm's payoff from a consumer with wealth W, and the expectation is taken over the wealth levels in the market. The objective is maximized over the space of functions $\mathcal{D}\times\mathcal{V}_0$, where the set $\mathcal{D}$ contains functions $d:\mathbb{R}_+\to\mathbb{R}_+$ that are absolutely continuous, left- and right-differentiable, and satisfy $d(w)\le w$; and set $\mathcal{V}_0$ consists of functions $v_1^b:\mathbb{R}_+\to[\underline{v}_1^b,\overline{v}_1^b]$ that are absolutely continuous, and left- and right-differentiable. The incentive compatibility constraint in (35) states that the functions $d(w)$ and $v_1^b(d)$ must be consistent with the consumer's preference maximization in period 0. The problem (34)–(35) is structurally similar to the problem (19)–(21) studied in Section 4.3, except that now we do not have a promise-keeping constraint, but do have the choice over $\underline{d}$. The following proposition shows that the shapes of the optimal functions in the former problem mimic the shapes of those in the latter.

Proposition 7. *For a given $\underline{d}$, there exist $\alpha_0,\kappa_0^v\in\mathbb{R}_+$ such that an optimal reward function is given by*

$$v_1^{b*}(d\,;\alpha_0,\kappa_0^v)=\begin{cases}\big(\alpha_0+\tilde\lambda_1 d+\big(\tilde\lambda_2-\tilde\lambda_1\big)[d-\kappa_0^v]^+\big)/\delta & \text{for}\quad 0\le d<\hat d_0(\alpha_0,\kappa_0^v),\quad\text{and}\\ \overline{v}_1^b & \text{for}\quad d\ge\hat d_0(\alpha_0,\kappa_0^v),\end{cases}\tag{36}$$

where $\alpha_0\ge\delta\underline{v}_1^b$, and $\hat d_0(\alpha_0,\kappa_0^v)$ solves the equation $\tilde\lambda_1 d+\big(\tilde\lambda_2-\tilde\lambda_1\big)[d-\kappa_0^v]^+=\delta\overline{v}_1^b-\alpha_0$. Let $\underline{w}(\alpha_0,\kappa_0^v,\underline{d})=\inf\big\{w\in[\underline{d},\infty)\mid\tilde u(w-\underline{d})+\delta v_1^{b*}(\underline{d}\,;\alpha_0,\kappa_0^v)\ge\tilde u(w)+\delta\underline{v}_1^b\big\}$, which is the lowest wealth at which the consumer is willing to pay the minimum downpayment. An optimal downpayment function $d^*(\cdot)$ is given by

$$d^*(w\,;\alpha_0,\kappa_0^v,\underline{d})=\begin{cases}[w+\underline{d}-\underline{w}(\alpha_0,\kappa_0^v,\underline{d})]^+ & \text{for}\quad 0\le w<\hat d_1(\alpha_0,\kappa_0^v,\underline{d})+\underline{w}(\alpha_0,\kappa_0^v,\underline{d})-\underline{d},\\ \hat d_1(\alpha_0,\kappa_0^v,\underline{d}) & \text{for}\quad \hat d_1(\alpha_0,\kappa_0^v,\underline{d})+\underline{w}(\alpha_0,\kappa_0^v,\underline{d})-\underline{d}\le w<\hat d_1(\alpha_0,\kappa_0^v,\underline{d})+\tilde\kappa,\\ w-\tilde\kappa & \text{for}\quad \hat d_1(\alpha_0,\kappa_0^v,\underline{d})+\tilde\kappa\le w<\hat d_2(\alpha_0,\kappa_0^v,\underline{d})+\tilde\kappa,\\ \hat d_2(\alpha_0,\kappa_0^v,\underline{d}) & \text{for}\quad w\ge\hat d_2(\alpha_0,\kappa_0^v,\underline{d})+\tilde\kappa,\end{cases}$$

where $\hat d_1(\alpha_0,\kappa_0^v,\underline{d})=\max\big\{\underline{d},\,\min\big\{\underline{w}(\alpha_0,\kappa_0^v,\underline{d})+[\kappa_0^v-\underline{d}]^+,\,\kappa_0^v,\,[\delta\Lambda_{1,0}-\alpha_0]^+/\tilde\lambda_1,\,\hat d_0(\alpha_0,\kappa_0^v)\big\}\big\}$, $\hat d_2(\alpha_0,\kappa_0^v,\underline{d})=\max\big\{\underline{d},\,\big(\delta\Lambda_{2,0}-\alpha_0-\big(\tilde\lambda_1-\tilde\lambda_2\big)\kappa_0^v\big)/\tilde\lambda_2,\,\min\big\{\kappa_0^v,\,\hat d_0(\alpha_0,\kappa_0^v)\big\}\big\}$, and $\Lambda_{j,0}=\max\big\{\underline{v}_1^b,\,\sup\big\{v\in[\underline{v}_1^b,\overline{v}_1^b]\mid\pi_1^{b'}(v)\ge-\delta/\big(\gamma\tilde\lambda_j\big)\big\}\big\}$, for $j\in\{1,2\}$.

As expected, d^* is increasing in w and v_1^{b*} is increasing in d. It is optimal for the firm to suggest a consumer with relatively high (resp., low) wealth to make a higher (resp., lower) downpayment and in turn be promised a higher (resp., lower) value at the start of the contract. Proposition 7 reduces the problem (34)–(35) into a simpler one with three parameters α_0, κ_0^v, and $\underline{d}$:

$$\pi_0^{(2)}=\max_{\alpha_0\in\left[\delta\underline{v}_1^b,\delta\overline{v}_1^b\right],\,\kappa_0^v\in\mathbb{R}_+,\,\underline{d}\in\bar{\mathbb{R}}_+}\mathbb{E}\big[\mathbb{1}_{\{W\ge\underline{d}\}}\big(d^*(W\,;\alpha_0,\kappa_0^v,\underline{d})+\gamma\pi_1^b(v_1^{b*}(d^*(W\,;\alpha_0,\kappa_0^v,\underline{d})\,;\alpha_0,\kappa_0^v))-\mathcal{K}\big)\big].\tag{37}$$

Proposition 8. *The following statements hold:*

(i) *For every $d^\dagger$ that is optimal in* (33) *there exists $\underline{d}^*\le d^\dagger$ that is optimal in* (34)*–*(35)*; thus, the wealth needed to acquire the technology in the second method is no larger than in the first.*

(ii) *The second method is better for the firm: $\pi_0^{(2)}\ge\pi_0^{(1)}$.*

We can think of downpayment as an admission fee that the firm charges its consumers for entry into the contract. In the first method, we see from (33) that the firm admits only a part of the market, i.e., the consumers with enough wealth to afford the fixed downpayment. By contrast, in the second method, the firm offers flexibility in how much downpayment the consumers can make, but will not admit the ones who cannot afford the minimum downpayment $\underline{d}^*$. As per Proposition 8(i), the minimum downpayment under the second method is lower than the optimal fixed method of the first. As a result, more consumers can acquire the technology if it is distributed using the second method.[3] To understand why this happens, notice the trade-off in (33): setting the downpayment low allows the firm to admit more consumers, but that reduces the revenue from each consumer; setting it high achieves the opposite. The second method eliminates this trade-off: admit more consumers, but also charge higher fees to the consumers with relatively higher wealth. This is why the second method is always better for the firm.

Thus, the *flexible* downpayment scheme of the second method allows a consumer with relatively low wealth—who otherwise would not have been able to enter the contract—to take it up by making a smaller downpayment, resulting in a more *inclusionary* method of initiation. In other words, a consumer in period 0 under this scheme could *purchase her position* in the contract by making an appropriate downpayment. Because a low-wealth consumer starts at a lower position, her time-to-ownership would be longer than a consumer with higher wealth. In fact, for the latter, the purchased position could be so high that the firm immediately grants her the ownership in period 0, which is a direct sale of the technology to the consumer. (We elaborate more on this in Appendix B.) Currently, the PAYGo firms sell the technology to consumers who can afford to purchase it right away, and the others are charged the same value of downpayment to initiate the contract. The optimal contract suggests that the initiation method need not be this rigid.

5. Managerial Insights

5.1. Summary and Implementation

A key feature of the optimal contract is that it *scores* the consumer through the value that is promised to her in the future, which we refer to hereafter as the v-score. The v-score summarizes the consumer's payment history; the higher the v-score, the better the payment history. Figure 5(a) shows the evolution of v-score under the optimal contract. In period 0, the firm proposes a flexible downpayment scheme $d^{**}(w)$. The consumer pays d and purchases her position in the contract, which is also her promised future reward and the initial value of her v-score $v_1^b = v_1^{b**}(d)$. (Here v_1^{b**} and d^{**} are the functions v_1^{b*} and d^* evaluated at the optimal parameter values.) After the contract starts, in a representative period t, the consumer enters the period with a v-score of v_t^b. The firm recommends the consumer pay $p_t^{**}(y\,;\,v_t^b)$ if her disposable income in period t is y. This payment scheme is flexible, respects the consumer's liquidity constraint, accommodates her income uncertainty, and is also incentive-compatible. After the consumer makes a payment p, her v-score updates to $v_t^p = v_t^{p**}(p\,;\,v_t^b)$. Thereafter, based on the updated v-score, the consumer is granted fractional access to the technology $f_t^*(v_t^p)$, and the v-score is again updated to $v_t^a = v_t^{a*}(v_t^p)$. Subsequently, the contract determines whether to continue or to terminate based on the current v-score v_t^a: the device is confiscated with probability $c_t^*(v_t^a)$, the ownership is transferred with probability $o_t^*(v_t^a)$, and the contract continues to the next period with probability $1-c_t^*(v_t^a)-o_t^*(v_t^a)$. Under continuation, the v-score updates to $v_t^e = v_t^{e*}(v_t^a)$. This

[3] An implicit assumption that we have made in our analysis is that the random variable W represents the steady state distribution of wealth within the market. Period 0, the period in which the consumer makes the take-up decision, is not a static point in time in reality. A consumer who cannot make the downpayment today might be able to make it at a later point in time by saving some money in the meantime. Such transitions of consumers across wealth levels in the market are assumed to be in a steady state, and in that case, the above argument continues to hold.

end-of-period v-score, when moved to the next period, appreciates by consumer's discount factor; i.e., the v-score at the beginning of period $t+1$ is v_t^e/δ, and the score-update process continues in a similar manner. If the contract is not already terminated, it is definitely terminated in period T: the device is confiscated with probability $c_T^*(v_T^a)$ and the ownership is transferred with probability $1-c_T^*(v_T^a)$.

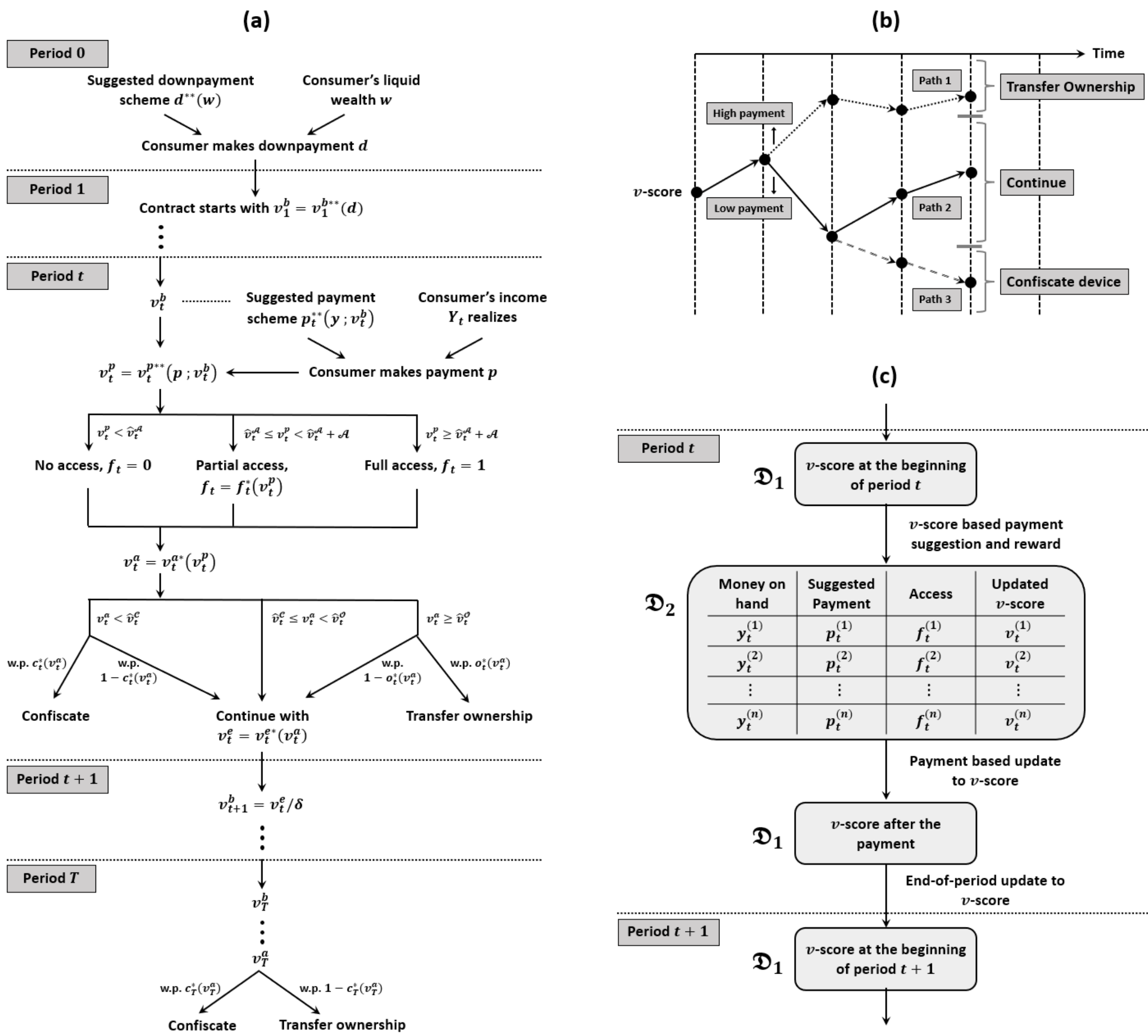

Figure 5 **(a) Summary of the contract dynamics. (b) Three sample paths of v-scores. (c) Implementation of contract using digital displays $\mathfrak{D}_1$ and $\mathfrak{D}_2$.**

To sum, a consumer's v-score starts at a value that increases with her downpayment, and thereafter, as shown in Figure 5(b), evolves with her payments, dropping with low payments and rising with high payments. That process summarizes the payment history. In Figure 5(b), all the three sample paths start at the same value of v-score, but the first path results in ownership transfer, the third in confiscation, and the second path continues further. We show in Appendix B that the commonly-seen business models in the BoP markets (sales, rent-to-own, and rental models) emerge as special cases of this optimal contract.

The ability to effectively summarize a consumer's payment history in a single score has profound practical implications. The flexibility given to consumers in terms of both the amount that they pay and the time at which they pay makes the communication of payment performance difficult. For example, the total amount paid by the consumer thus far does not sufficiently capture her performance: a consumer who paid 50 USD in 3 months is not the same as one who paid the same amount but in 9 months (all else equal), because of the time value of money. By either not communicating anything to consumers or by communicating ineffective metrics such as cumulative payment (both are seen in practice), a disconnect is created between the firm's and the consumer's perceptions of a consumer's payment performance. The surveys by Zollmann et al. (2017) reveal that the consumers who are regularly behind on payments might also think that they are good payers because "they were trying hard or missed only X days, weeks, or months of payments," or even worse, because their device is not yet confiscated (Zollmann et al. 2017, p. 24). Such a disconnect can be resolved by communicating the v-score—a metric that automatically accounts for the time value of money—continuously to consumers through a digital display on their devices.

Moreover, in the current implementations of payment flexibility, consumers are given a reference payment schedule (e.g., pay 1 USD every week for a duration of 3 years), but are allowed to deviate from it as they wish. The firm too can deviate from the initially-communicated repayment term depending on consumers' payment performance (Zollmann et al. 2017, Guajardo 2021, Bonan et al. 2023). In such implementations, a bad performer is provided as much flexibility as a good performer. The consumers are also not offered recommendations on how much they can deviate from the reference schedule, and they are generally not aware of the current and future consequences of their deviations (Zollmann et al. 2017, Kocieniewski and Finch 2022). The flexibility offered in this manner is not only excessive, it is also implicit in nature. Some consumers may not even need so much flexibility, and providing it without effectively communicating the consequences of using it might result in severe payment problems. To resolve these issues, the optimal contract proposes to explicitly communicate payment suggestions and rewards. That way, clear expectations are set on how much flexibility consumers have and what they get in return for their payments. As we discuss next, this too can be implemented using a digital display on devices.

The flexible downpayment scheme can be implemented by offering consumer a (discretized) menu of downpayment suggestions: $\{w^{(j)}, d^{(j)}, v_1^{(j)}\}$, where for a discretized set of liquid wealth levels $\{w^{(j)}\}$, $d^{(j)} = d^{**}(w^{(j)})$ and $v_1^{(j)} = v_1^{b**}(d^{(j)})$. The consumer, with liquid wealth $w^{(j)}$, is advised to pay $d^{(j)}$, and upon paying $d^{(j)}$ the contract starts with a v-score $v^{(j)}$. The v-score in our model can be normalized to between 0 and 100 at the beginning, with 0 indicating no entry into the contract and 100 indicating immediate transfer of ownership. After the contract starts, the necessary details can be communicated to the consumer through two digital displays $\mathfrak{D}_1$ and $\mathfrak{D}_2$ on her device, see Figure 5(c). $\mathfrak{D}_1$ displays the current v-score of the consumer, indicating her payment performance, while $\mathfrak{D}_2$ displays payment suggestions and rewards as a lookup table. For a discretized set of disposable income values $\{y_t^{(j)}\}_{j=1}^n$ in period t, $\mathfrak{D}_2$ displays in a table $\{(y_t^{(j)}, p_t^{(j)}, f_t^{(j)}, v_t^{(j)})\}$, where (i) $p_t^{(j)} = p_t^{**}(y_t^{(j)}; v_t^b)$, (ii) $f_t^{(j)} = f_t^*(v_t^{p**}(p_t^{(j)}; v_t^b))$, and (iii) $v_t^{(j)} = v_t^{a*}(v_t^{p**}(p_t^{(j)}; v_t^b))$. (The size n could be chosen based on the available display space, and the access could be displayed in terms of hours.) Thus, $\mathfrak{D}_2$ offers a payment suggestion: "if you have $y_t^{(j)}$ on hand, you are advised to pay $p_t^{(j)}$," and informs the resultant current and future rewards: "if you pay $p_t^{(j)}$, you will be provided access $f_t^{(j)}$ in the current period and your v-score will then be updated to $v_t^{(j)}$." At any point in time, the contract requires the consumer to make only a single-period payment decision, with the knowledge of how her v-score will be updated upon payment (see Section 4.3), which is displayed in $\mathfrak{D}_2$.

Figure 5(c) also shows how the displays must be updated over time. At the beginning of period t, $\mathfrak{D}_1$ displays v_t^b, and the payment suggestions and rewards in $\mathfrak{D}_2$ are based on this value of v-score. After the consumer makes a payment $p_t^{(j)}$, $\mathfrak{D}_1$ displays the corresponding updated v-score $v_t^{(j)}$ ($= v_t^a$ in our model). At the end of the period, if the contract is not terminated, then the v-score is updated from its current value to v_{t+1}^b while moving to the next period. The same update process continues thereafter. All the updates over time can be done automatically and remotely using IoT technology, and at a frequency that the firm deems appropriate (e.g., updates on a daily basis offer flexibility at the lowest possible denomination).

The v-scoring mechanism offers efficient confiscation and ownership transfer rules that can be easily communicated through the color coding of digital displays. The rows in $\mathfrak{D}_2$ that put the consumer at risk of confiscation can be displayed in red, the rows that result in continuation can be displayed in green, and blue can be used for the rows that offer ownership. $\mathfrak{D}_1$ can dynamically change its display color depending on what the current v-score entails. We believe that this dynamic scoring mechanism can resolve the issues associated with fixed termination rules that are currently seen in practice (Zollmann et al. 2017, Kocieniewski and Finch 2022), such as fixed confiscation rules, e.g., "we will confiscate the device if there is no payment for 90 days." Although it is easy to make the consumer aware of such a rule, it may not be efficient and is therefore not necessarily enforced: a consumer who paid 80 USD cannot be treated the same as a consumer who paid only 30 USD in the same amount of time, and now both with 90 days of no payment (all else equal). The confiscation threat must be stronger for the latter, and the threat may need to intensify long before the 90 day deadline is reached. The v-scoring mechanism automatically takes these factors into account by dynamically offering different incentives to these two consumers and by communicating the threat via a red light on the displays.

Similarly, fixed ownership rules such as "pay 150 USD over 3 years to obtain ownership," again seen in practice, are also inefficient. Some consumers could have paid faster, but they are given perverse incentives to pay slowly. Moreover, if a consumer is paying too slowly, then because of time value of money, the firm might want to charge her more than 150 USD, resulting in common situations wherein the consumer has already paid 150 USD but ownership has yet to be transferred. Under the optimal contract, the consumers are given ownership when their v-score goes above a certain threshold, which can be communicated at the beginning, as well as dynamically using a blue light on the displays. We leave it to future research to test the efficacy of the proposed contract and its implementation in the field.

5.2. Comparison with ToP Contracts

There are some features that are shared between the contract that we studied and the contracts that we see in the ToP markets. For example, the mortgage contracts offered by the banks and the service contracts offered by the utility companies evaluate consumers by tracking their payment history, the way the consumer is evaluated in our model through her v-score. If a consumer's payment history is significantly bad, severe measures are taken such as claiming the collateral or terminating the service contract, as in our model where the consumer faces the risk of confiscation if her v-score falls below the threshold $\hat{v}_t^{\mathcal{C}}$. If the consumer has a good payment history, the organization might tolerate a few non-payment periods; our model also entitles the consumer to grace periods if her v-score is sufficiently high. Similar to the firm in our model, utility companies can regulate consumer access by providing fractional services based on payment levels. Next, we compare the contract in this paper with some ToP contracts along a few specific dimensions, to further enhance the understanding of some of its key features.

Credit scores versus *v*-scores. The v-score emerging from the optimal contract summarizes the consumer's payment history: it is higher for good payers and lower for bad payers. In this sense, we can interpret a consumer's v-score as her *credit score*. The consumers in the BoP markets hardly have any credit history that they can show to financial organizations to obtain monetary support, and those organizations do not usually offer loans to these consumers because they do not have a credit history; a vicious circle that v-scores could help break. Provided that the policy environment is supportive, BoP consumers could use their v-scores as credit scores while applying to other organizations for financial support (e.g., microfinance institutions, development banks, and other private firms), and those organizations in turn could treat v-scores as indicators of consumers' creditworthiness while extending support.

Despite this similarity, the functionality of v-score in our context is somewhat different from the functionality of credit score that we see in the ToP markets. Credit scores are usually constructed by third parties using multiple sources of consumer transactions, and they act as a central repository of information that different organizations use to assess the creditworthiness of consumers. A consumer generally knows that certain actions increase her credit score, while certain others decrease it, but while taking those actions, she may not necessarily be aware of (or may not take into consideration) what her credit score is, and what the exact impact of her actions will be on her score. Therefore, the incentives induced by the credit-scoring mechanism to elicit good payment behavior are implicit and imperfect. By contrast, the purpose of v-scoring in our model is to *explicitly* offer consumers the (nonmonetary) payment incentives. The efficacy of the contract hinges on the perfect knowledge of v-score and how it updates with the payments, which, as we discussed in Section 5.1, can be communicated to the consumer through digital displays on her device.

Flexibility as a contractual feature. A distinctive feature of our setting is that the consumer is contractually offered payment flexibility. This feature arises because of the consumer's liquidity constraints and income uncertainty—the factors that are specific to the BoP. If we remove these factors, then the contract becomes inflexible—it becomes the kind of fixed installment contract that we see in the ToP markets. To see this, refer to the payment function with crosses in Figure 4(c): it takes that shape for high values of v_t^b. On this payment function, if we move the minimum income, call it $\underline{y}_t$, from zero to a value beyond the kink, then in the domain $[\underline{y}_t, \infty)$, the payment function $p_t^{**}(y)$ is a flat line. In other words, we get this fixed installment contract—in which the firm asks for the same amount irrespective of the consumer's income—if the disposable income (here, $\underline{y}_t$) and the credit score (here, v_t^b) are sufficiently high; the typical characteristics of a ToP consumer. Thus, payment flexibility is intimately tied to the fundamental features of the BoP market.

Attitude to advance payments. A lending organization like a bank does not prefer the consumer to pay in advance. A bank earns money by charging an interest rate that is higher than its own discount rate (Choudhry 2022), and hence a consumer paying in advance hurts the bank's earnings. The utility companies allow their consumers to pay in advance for their services, but these advance payments are simply treated as a store credit, and the consumer does not earn interest on it; the consumer is therefore not explicitly incentivized to pay in advance. In contrast, the firm in our setting encourages the consumer to pay in advance to accommodate the uncertainty in her income. The contract rewards the consumer for her advance payments: her remaining credit at the end of a period (i.e., v_t^e) is appreciated by her discount factor. Without this increase of the v-score, the consumer has no incentive to pay in advance.

Servitization. Some firms in the ToP markets adopt a servitization strategy, where instead of selling a product (e.g., printers, heavy machinery, automotives), its outcome is sold as a service. The most common service contracts that are offered under this strategy are *subscription* and *pay per use*. A subscription contract requires the consumer to pay a (relatively high) fee and in return, she is offered uninterrupted service for a (relatively long) specified duration. In a pay-per-use contract, the consumer pays for the service whenever she requires it and the amount paid corresponds to the duration of usage. The choice between the two in the ToP market generally depends on the intensity of demand for the service: if intensity is high, then a subscription contract is better; otherwise, a pay-per-use contract is better (see, e.g., Fishburn and Odlyzko 1999). In a BoP market however, there is another factor at play: the consumer's liquidity constraints, which makes the contract design nontrivial. The products and services discussed in our paper are necessities, so demand intensity is high. Despite that, a PAYGo firm cannot simply offer a subscription contract because that requires the consumer to pay a higher amount upfront, which she may not always be able to.

Interestingly, the optimal contract integrates the features of both subscription and pay-per-use contracts dynamically, depending on the consumer's v-score. If a consumer pays high amounts in advance, then her v-score shoots up and the contract allows her to access the device fully without any payments for the next few periods. Thus, a farmer who earns money seasonally after harvests can plausibly use the device on a subscription basis, while a truck driver who may earn money in small amounts spread over time can use the device on a pay-per-use basis, as and when the money arrives. (The constraining factor in this case is not the demand for the service—that always exists—but the ability to pay for it.)

Remote asset control. Several ToP firms and merchants sell electronic goods using payment models—such as buy-now-pay-later and rent-to-own—that allow consumers to pay over time, like in our setting. However, they do not incorporate the remote access-control feature into their devices. Doing the same in a BoP market, although theoretically possible, might be ineffective because of practical limitations. If the firm lacks the ability to control the device remotely, then it can discipline the consumer only through the threat of confiscation. The (continuation) contract, which earlier was a triplet (f^t, c^t, o^t), now reduces to (c^t, o^t). It is easy to see that the resulting incentives will be weaker than before, and there will therefore be more confiscations. The ToP organizations can avoid such dire consequences because they have alternative mechanisms for influencing consumer behavior: in addition to confiscating the asset, they can take the consumer to court and sue her, block her bank account, or report her to credit-scoring agencies, which then negatively affect her future purchasing power and prospects of getting a loan. Thus, the strong regulatory framework in a ToP market does not necessitate remote asset control. By contrast, BoP markets usually have weak and broken regulatory frameworks, which severely limits the set of available discipline mechanisms. PAYGo firms therefore need to invest in IoT technologies and design devices that are remotely controllable in order to maintain the ability to influence consumer behavior.

Contract initiation and adverse selection. Our contract initiation problem assumes that consumers in the market differ only in terms of their liquid wealth, which is not observable to the firm; this problem is solved by offering a menu of contracts that accepts varying levels of downpayment, starting consumers at different values of v-scores. This assumption may not be far from reality in remote villages where the households usually share characteristics and rely on similar means to make their living. However, we acknowledge that our formulation of the initiation problem is incomplete. Consumers could differ on other characteristics at the time of initiation—e.g., their financial discipline and consumption patterns, which are difficult to observe but influence their willingness and ability to repay—thus resulting in a complex adverse selection problem. Banks and other lending organizations in the ToP markets face similar problems, and solve them by resorting to techniques such as screening and risk-based pricing (Choudhry 2022). We do not study these aspects of the initiation problem in our paper, as our primary focus has been on the problem of efficiently managing payment flexibility—the dynamic moral hazard problem that ensues after the contract starts. That said, the techniques deployed by the ToP organizations can be practically useful in our setting too. For example, extensive surveys could be used to cluster consumers into different homogeneous categories. The firm may then decide to screen out some categories and offer different menus to the remaining categories depending on their characteristics and perceived risk. Future research could explore these aspects in more detail.

6. Concluding Remarks

In this paper, we have carried out a rigorous analysis of the payment problem stemming from payment flexibility in a PAYGo model. We used the optimal contracting approach to solve the problem and generated insightful and implementable results on how to score consumers and optimal approaches to offering flexibility. Our work has implications for firms, academics, policymakers, and consumers.

Our contracting model could act as a structured framework for the PAYGo firms to understand and analyze the payment problems within the industry. They can conduct field experiments to test the efficacy of different features suggested by the optimal contract: importantly, the v-scoring mechanism and offering suggestions to consumers on how to make use of flexibility. The model can also form the basis for a dynamic learning framework that learns more about consumer characteristics as more data is collected using IoT technology and then adjusts the policies as necessary.

In terms of academic relevance, to the best of our knowledge, there has been no systematic theoretical analysis of PAYGo contracts; we believe our paper is a first. The problem studied here is not only practically relevant but also technically challenging. The approach that we have developed to solve the incentive problem can be applied broadly to other contexts. Future research could explore other related problems within the context of PAYGo. A third party (a policymaker or a donor) could offer different types of subsidy dynamically, either to the firm or the consumers, with the aim of improving societal outcomes. The design of optimal subsidy programs deserves a detailed investigation. Another crucial problem is related to the sales agents of PAYGo firms. The agents might share misleading information about the contract or device capabilities to bring consumers on board quickly. This harms consumers and the firm alike (Zollmann et al. 2017, Kocieniewski and Finch 2022). It is therefore necessary to design incentives for agents that link their commissions to consumers' payment performance, yet still motivate them to reach out to low-income consumers.

Several low-income countries are resource-constrained and cannot extend their centralized supply networks to remote rural areas. That is why these countries have been welcoming decentralized off-grid service provision by private firms. While extending support to these firms or regulating them, policymakers can enforce transparent scoring mechanisms and clear communication about the consequences of using flexibility, as guided by the optimal contract. Policymakers must note that the v-scores emanating from our contract are akin to credit scores that quantify the creditworthiness of consumers. They can therefore facilitate the creation of a seamless framework across different organizations to extend further financial support to these consumers based on their v-scores.

If there is no light in a household, its members are vulnerable to theft and reptile attacks in the nighttime, while their productive time and the children's study time are reduced. If a household does not use clean cooking methods, the smoke from firewood or charcoal can impact their respiratory health, and they face fire hazards that can burn their house down and even take their lives. If there is no access to clean water, they are vulnerable to several water-borne diseases. A lack of access to smartphones and televisions means that households cannot access critical information on current affairs and important opportunities. Depriving the poor of such basic products and services traps them in poverty. The flexible financing available in PAYGo models offers the possibility of bringing valuable assets within the reach of a large number of low-income households and releasing them from the poverty trap. Our paper considers efficient mechanisms to bring life-improving goods and services into the lives of the poor, which in turn contribute to increasing consumers' productivity, improving their health, alleviating poverty, and promoting economic growth.

References

Acemoglu, D., M. Arellano, E. Dekel. 2013. *Advances in Economics and Econometrics: Tenth World Congress*, vol. 1. Cambridge University Press.

Azagra, D. 2013. Global and Fine Approximation of Convex Functions. *Proceedings of the London Mathematical Society* **107**(4) 799–824.

Balasubramanian, K., D.F. Drake, J. Acimovic, D. Fearing. 2023. Inventory Management for Mobile Money Agents in the Developing World. *Forthcoming in Manufacturing & Service Operations Management.* Available at SSRN: 2987591.

Balke, N., T. Lamadon. 2022. Productivity Shocks, Long-Term Contracts, and Earnings Dynamics. *American Economic Review* **112**(7) 2139–2177.

Balseiro, S.R., H. Gurkan, P. Sun. 2019. Multiagent Mechanism Design Without Money. *Operations Research* **67**(5) 1417–1436.

Barboni, G. 2017. Repayment flexibility in microfinance contracts: Theory and experimental evidence on take up and selection. *Journal of Economic Behavior & Organization* **142** 425–450.

Biais, B., T. Mariotti, J. Rochet, S. Villeneuve. 2010. Large Risks, Limited Liability, and Dynamic Moral Hazard. *Econometrica* **78**(1) 73–118.

Bonan, J., G. d'Adda, M. Mahmud, F. Said. 2023. Nudging Payment Behaviour: Evidence from a field experiment on pay-as-you-go off-grid electricity. *Forthcoming in World Bank Economic Review.* URL `https://ore.exeter.ac.uk/repository/handle/10871/132791`.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press.

Calmon, A., D. Jue-Rajasingh, G. Romero, J. Stenson. 2022. Operational Strategies for Distributing Durable Goods in the Base of the Pyramid. *Manufacturing & Service Operations Management* **24**(4) 1887–1905.

Choudhry, M. 2022. *The Principles of Banking*. Wiley Finance, Wiley.

Clementi, G.L., H.A. Hopenhayn. 2006. A Theory of Financing Constraints and Firm Dynamics. *The Quarterly Journal of Economics* **121**(1) 229–265.

Collins, D., J. Morduch, S. Rutherford, O. Ruthven. 2009. *Portfolios of the Poor: How the World's Poor Live on $2 a Day*. Princeton University Press.

de Zegher, J.F., D.A. Iancu, E. Plambeck. 2018. Sustaining Rainforests and Smallholders by Eliminating Payment Delay in a Commodity Supply Chain–It takes a Village. Working paper, Stanford University.

DeMarzo, P.M., M.J. Fishman. 2007a. Agency and Optimal Investment Dynamics. *The Review of Financial Studies* **20**(1) 151–188.

DeMarzo, P.M., M.J. Fishman. 2007b. Optimal Long-Term Financial Contracting. *The Review of Financial Studies* **20**(6) 2079–2128.

Field, E., R. Pande, J. Papp, Y.J. Park. 2012. Repayment Flexibility Can Reduce Financial Stress: A Randomized Control Trial with Microfinance Clients in India. *PLoS ONE* **7**(9) 1–7.

Fischer, G.W. 1975. Experimental Applications of Multi-Attribute Utility Models. D. Wendt, C. Vlek, eds., *Utility, Probability, and Human Decision Making*. Springer Netherlands, 7–46.

Fischer, G.W. 1976. Multidimensional utility models for risky and riskless choice. *Organizational Behavior and Human Performance* **17**(1) 127–146.

Fishburn, P.C., A.M. Odlyzko. 1999. Competitive Pricing of Information Goods: Subscription Pricing versus Pay-per-Use. *Economic Theory* **13**(2) 447–470.

Gao, J., F. Zhao, Y. Gu. 2018. Sufficient conditions of stochastic dominance for general transformations and its application in option strategy. *Economics* **12**(1) 20180001.

Gernert, A., A. Calmon, G. Romero, L.N. Van Wassenhove. 2023. Business Model Innovation for Ambulance Systems in Developing Economies: 'Coordination and Competition'. *Forthcoming in Production and Operations Management.* Available at SSRN: 4049997.

Green, E.J. 1987. Lending and the Smoothing of Uninsurable Income. E.C. Prescott, N. Wallace, eds., *Contractual Arrangements for Intertemporal Trade*. University of Minnesota Press, 3–25.

Guajardo, J.A. 2021. Repayment performance for pay-as-you-go solar lamps. *Energy for Sustainable Development* **63** 78–85.

Hopenhayn, H.A., J.P. Nicolini. 1997. Optimal Unemployment Insurance. *Journal of Political Economy* **105**(2) 412–438.

Jain, S., G. Mansuri. 2003. A little at a time: the use of regularly scheduled repayments in microfinance programs. *Journal of Development Economics* **72**(1) 253–279.

Kocieniewski, D., G. Finch. 2022. Tesla-Backed Startup Made Cheap Power a Debt Burden for the World's Poorest. URL `https://www.bloomberg.com/news/features/2022-04-07/how-pay-as-you-go-solar-made-the-world-s-cheapest-new-energy-unaffordable`.

Kundu, A., K. Ramdas. 2022. Timely After-Sales Service and Technology Adoption: Evidence from the Off-Grid Solar Market in Uganda. *Manufacturing & Service Operations Management* **24**(3) 1329–1348.

Labie, M., C. Laureti, A. Szafarz. 2017. Discipline and flexibility: a behavioural perspective on microfinance product design. *Oxford Development Studies* **45**(3) 321–337.

Laureti, C. 2012. Flexibility and Payment Discipline in Microfinance. Working paper, Université de Mons and Université Libre de Bruxelles.

Levy, H. 1992. Stochastic Dominance and Expected Utility: Survey and Analysis. *Management Science* **38**(4) 555–593.

Lighting Global. 2022. *Off-Grid Solar Market Trends Report*. World Bank Group.

Mas-Colell, A., M.D. Whinston, J.R. Green. 1995. *Microeconomic Theory*. Oxford University Press.

Milgrom, P., I. Segal. 2002. Envelope Theorems for Arbitrary Choice Sets. *Econometrica* **70**(2) 583–601.

Piskorski, T., A. Tchistyi. 2010. Optimal Mortgage Design. *The Review of Financial Studies* **23**(8) 3098–3140.

Piskorski, T., A. Tchistyi. 2011. Stochastic House Appreciation and Optimal Mortgage Lending. *The Review of Financial Studies* **24**(5) 1407–1446.

Ramdas, K., A. Sungu. 2022. The Digital Lives of the Poor: Entertainment Traps and Information Isolation. *Forthcoming in Management Science.* Available at SSRN: 4012172.

Rothschild, M., J.E. Stiglitz. 1970. Increasing Risk. I. A Definition. *Journal of Economic Theory* **2** 225–243.

Spear, S.E., S. Srivastava. 1987. On Repeated Moral Hazard with Discounting. *The Review of Economic Studies* **54**(4) 599–617.

Tedeschi, G.A. 2006. Here today, gone tomorrow: Can dynamic incentives make microfinance more flexible? *Journal of Development Economics* **80**(1) 84–105.

Tian, F., P. Sun, I. Duenyas. 2021. Optimal Contract for Machine Repair and Maintenance. *Operations Research* **69**(3) 916–949.

Uppari, B.S., S. Netessine, I. Popescu, R.P. Clarke. 2023. Design of Off-Grid Lighting Business Models to Serve the Poor: Field Experiments and Structural Analysis. *Forthcoming in Management Science.* Available at SSRN: 4326920.

Uppari, B.S., I. Popescu, S. Netessine. 2019. Selling Off-Grid Light to Liquidity-Constrained Consumers. *Manufacturing & Service Operations Management* **21**(2) 308–326.

Wang, S., P. Sun, F. de Véricourt. 2016. Inducing Environmental Disclosures: A Dynamic Mechanism Design Approach. *Operations Research* **64**(2) 371–389.

Wiemann, M., D. Lecoque. 2019. High-profile bankruptcies in the off-grid sector: Where do we go from here? Tech. rep., Alliance for Rural Electrification.

Winiecki, J. 2019. The power of flexibility: How innovative payment models can expand access to energy services. URL `https://bfaglobal.com/fibr/insights/the-power-of-flexibility-how-innovative-payment-models-can-expand-access-to-energy-services`.

Zollmann, J., D. Waldron, A. Sotiriou, A. Gachoka. 2017. Escaping Darkness: Understanding Consumer Value in PAYGo Solar. Tech. rep., CGAP.

Appendix A: Repayment Term

We assumed in the paper that the repayment term (the end time of the contract) is T, the lifetime of the device. But this assumption may not always be realistic, especially when the device's lifetime is too long. For example, a solar home system's lifetime usually exceeds 5 years, but most of the payment contracts that we see only run for up to 3 years. So the repayment term θ need not coincide with the device's lifetime T.

In our model, the firm can set the repayment term to any arbitrary length $\theta < T$ by exogenously setting the last period for the contractual relationship to be θ. The choice of θ may depend on factors that are external to the model, such as the expiration period of funds or the deadlines set by the involved parties. In such a case, we can redo the analysis in Section 4 with θ, instead of T, as the last period. Doing so does not qualitatively affect the contract structure: in Figure 5(a), period θ replaces period T—the period in which there is no option of contract continuation. Interestingly, a repayment term less than T can also endogenously emerge from our model. Recall Figure 2(b): if the firm sees no value in continuing the contract in a certain period, it will terminate the contract in that period (by randomizing between confiscation and ownership transfer). The following result establishes a sufficient condition for the existence of such a period θ ($< T$). To derive this result, we mute the impact of ownership incentives by setting $\mathcal{O}_t = 0$ for all t, because we are interested in the endogenous emergence of a shorter repayment term *without* the influence of a third party (e.g., a policymaker or a donor who sets $\mathcal{O}_t$).

Proposition 9. *If there exists a time period θ such that*

$$\frac{\mathcal{C}_{t-1}}{\mathcal{C}_t} \leq \frac{\gamma}{\delta}\frac{\overline{v}^e_{t-1}-\underline{v}^e_{t-1}}{\overline{v}^e_t-\underline{v}^e_t}, \qquad \text{for } 1<t\leq\theta, \quad \text{and} \tag{38}$$

$$\frac{\mathcal{C}_{t-1}}{\mathcal{C}_t} \geq \frac{\gamma}{\delta}\frac{\overline{v}^e_{t-1}-\underline{v}^e_{t-1}}{\overline{v}^e_t-\underline{v}^e_t}\max\left\{1,\ \frac{\overline{v}^e_t-\underline{v}^e_t}{\lambda_2\,\mathcal{C}_t}\right\} \quad \text{for } \theta<t\leq T, \tag{39}$$

then it is optimal to terminate in period θ.

We expect the confiscation payoffs to weakly decrease over time (i.e., $\mathcal{C}_{t-1}\geq\gamma\mathcal{C}_t$). Proposition 9 imposes a condition on how $\mathcal{C}_t$ decreases over time: the relative fall in the confiscation payoffs (or equivalently, the decrease in the log of confiscation payoffs) is bounded below by a time-varying threshold in (38) until a period θ, and thereafter the relative fall is bounded above by a time-varying threshold in (39). It is easy to show that the time-varying thresholds on the right-hand sides of (38) and (39) are increasing in t. Thus, Proposition 9 states that if the confiscation payoffs fall at a moderate pace until a point in time, but thereafter fall at an increasing pace, then the firm will surely terminate the contract before the end of the device's lifetime. In other words, if the technology depreciates at an accelerating pace beyond a point in time or if the opportunity costs of the firm become increasingly intensive after a certain period (which both correlate with a steep drop in $\mathcal{C}_t$), then the contract's repayment term θ will be shorter than T. Moreover, the faster the technology depreciates, the shorter the repayment term.

Appendix B: Business Models

Three different business models that are prominently seen in the BoP markets emerge as special cases in our contracting framework: (i) a *sales model*, in which the firm charges its consumers a fixed price and gives ownership of (i.e., sells) the technology right after the payment; (ii) a *rental model*, wherein the firm never gives the option of technology ownership (i.e., it only rents the technology) to its consumers; and (iii) a *rent-to-own model*, which gives the option of ownership to consumers but does not immediately sell the device. In this section, we examine the relationships between the characteristics of a technology and the business model used to offer that technology.

As in Appendix A, we set the ownership incentive to zero in this section as well. By giving sufficiently high ownership incentives, a policymaker or a donor can exogenously increase the possibility of ownership in any period (see Figure 2(a)). However, we are interested in the endogenous emergence of different business models (and the option of ownership) within the contracting framework, so we set $\mathcal{O}_t=0$ for all t.

With this in mind, we now formally characterize the option of ownership in our model. Recall from Figures 4(b) and 4(c) that in period t, the highest reward that the firm offers the consumer is $v^{p**}_t(\hat{p}^*_{2,t})$. If this highest reward is equal to $\overline{v}^p_t$, which is the value that the consumer derives if she was granted the ownership in period t, then we say that the firm gives the option of ownership in period t. In contrast, if $v^{p**}_t(\hat{p}^*_{2,t})<\overline{v}^p_t$ for all $t\leq\theta$ for a given repayment term θ, then the firm never gives the option of ownership to the consumer—it only rents the device. Furthermore, if the option of ownership exists in period 0 itself, then the firm sells the technology to consumers who can afford it. We define this using the second method of contract initiation in Section 4.5. The option of ownership exists in period 0 if $v^{b**}_1(\hat{d}^*_2)=\overline{v}^b_1$, where the threshold $\hat{d}^*_2$ is $\hat{d}_2$ in Proposition 7 evaluated at the optimal values of α_0, κ^v_0, and $\underline{d}$. Specifically, to sell the technology, the firm charges a price of $\hat{d}_2^*$, and the consumers with wealth above $\hat{d}_2^* + \tilde{\kappa}$ pay that price and purchase the technology in period 0.

The following result establishes simple conditions on the confiscation payoff in the terminal period of the contract for the emergence of different business models. We make two assumptions to derive this result. First, the confiscation payoffs satisfy (38) and (39). Thus, the structure of the result holds for a repayment term θ^* that is either endogenously determined as per Proposition 9, or exogenously chosen but is smaller than the endogenous repayment term θ. Second, $\tilde{\lambda}_2 \geq \lambda_2$: this assumption is mild because $\tilde{\lambda}_2$ characterizes consumption utility over wealth, which is relatively larger in magnitude, as opposed to λ_2, which characterizes utility of consuming disposable income, which is relatively smaller in magnitude.

Proposition 10. *For a firm with the set of confiscation values* $\{\mathcal{C}_t\}$ *that satisfy* (38) *and* (39) *and repayment term* $\theta^* \leq \theta$*, there exist thresholds* $\hat{\mathcal{C}}^{(1)}$ *and* $\hat{\mathcal{C}}^{(2)}$ *such that:*

(i) *if* $\mathcal{C}_{\theta^*} \leq \hat{\mathcal{C}}^{(1)}$*, then the firm deploys a rent-to-own model for consumers with wealth below a threshold* $\hat{w}$*, and a sales model for the consumers with wealth above* $\hat{w}$*,*

(ii) *if* $\hat{\mathcal{C}}^{(1)} \leq \mathcal{C}_{\theta^*} < \hat{\mathcal{C}}^{(2)}$*, then the firm deploys a rent-to-own model, and*

(iii) *if* $\mathcal{C}_{\theta^*} \geq \hat{\mathcal{C}}^{(2)}$*, then the firm deploys a rental model.*

If the confiscation payoff in the terminal period is too low, then the firm sells the device to consumers who can afford it and offers a rent-to-own model to consumers who cannot purchase it right away. In general, $\mathcal{C}_{\theta^*}$ is low if the technology is either cheap or depreciates too fast once it is in the hands of the consumer, and significant value cannot be obtained from it in a secondary market. We know from Appendix A that fast-depreciating technologies are associated with shorter repayment terms. This result is consistent with the fact that technologies such as clean cookstoves and small solar lamps, which cost less and depreciate fast, are either sold directly or are offered on a rent-to-own basis with relatively short repayment terms that span only a few weeks or months.

For moderate values of $\mathcal{C}_{\theta^*}$, the firm offers only rent-to-own contracts. Solar home systems, smartphones, and computing systems, which cost relatively more and depreciate at a relatively moderate pace, are predominantly offered on a rent-to-own basis with relatively longer repayment terms that span multiple years.

If $\mathcal{C}_{\theta^*}$ is too high, the firm only rents the technology. This will be the case for technologies that either cost more or depreciate slowly, and remain quite valuable in a secondary market. In reality, these characteristics map well to two settings. First, if the firm takes the responsibility of regularly maintaining the device and replacing the parts when necessary, then it derives high value from that device even upon confiscation. We usually see this happening with large solar home systems (e.g., with fans and television), which are only rented out to consumers and maintained by the firm. Second is the case of services, such as providing cooking gas through cylinders and purified water through prepaid meters. Besides the fact that these technologies can never actually be owned by the consumer, they also have high a $\mathcal{C}_{\theta^*}$ because the confiscation payoff comes, not from the resale of any device, but from simply shifting the provision of service from one consumer to another relatively easily without incurring a high cost, so the firm's opportunity costs vary little over time. We conclude this section with a remark on how we can accommodate pure services like these (with no ownership option) in our modeling framework.

Modeling pure services. It is possible to transfer the ownership of technologies such as solar home systems and smartphones to a consumer because they can provide services (i.e., energy and information) in a self-sustaining manner, without the firm's involvement, after the ownership. By contrast, purified water and cooking gas cannot be provided in a self-sustaining manner (at least not with the prevailing technologies). For there to be a continuous supply of these services, the firms must operate water purification stations with piped connections, and source LPG and distribute it in cylinders. These technologies cannot therefore be owned by a single consumer. Such pure services can be accommodated within our modeling framework, with two minor modifications. First, a firm in this case incurs an operational cost to deliver the service (e.g., it needs to maintain the purification systems, or replace all the cylinders once a month). This can be accounted for in the firm's objective at the beginning of each period by rewriting it as $\pi_t^b(v_t^b) - c_t^o$, where c_t^o is the operational cost in period t. We can easily see from the analyses in Section 4 that such a modification might change the values of the thresholds in the optimal contract but does not affect its overall structure.

Second, given that the services operate over a much longer time period, perhaps it is appropriate to model service delivery as an infinite-horizon problem. This too can be accommodated easily in our model by taking $T \to \infty$ and making $\mathcal{C}_t$, G_t, and c_t^o stationary. Interestingly, making $\mathcal{C}_t$ stationary is consistent with what we just discussed above: when a technology does not depreciate over time, the firm offers it on a purely rental basis, which indeed is the case here. This modification will make all the threshold policies and the payment and reward functions in the optimal contract stationary (and also simpler).

The v-scoring mechanism in this case too can be implemented using digital displays, as we discussed in Section 5.1, but without the blue light indicators as there is no ownership option here. It would be appropriate to specify the access f_t in terms of quantities (e.g., liters) of water or gas that will be provided. Overall, the insights that we developed in the paper on how to efficiently manage payment flexibility—which has been our primary focus—apply equally well to the case of pure services.

Appendix C: Proofs of Propositions in the Main Paper and the Appendices

For the proofs, we introduce some notational conventions:

1. We denote extended real numbers by $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$.
2. We use $'^{+}$ ($'^{-}$) or $\frac{d}{d^+}$ ($\frac{d}{d^-}$) to denote right (left) derivatives.
3. We extend the traditional derivative notation to borders of closed sets and kinks in concave functions as follows. For a function π defined on any closed interval $[\underline{x}, \bar{x}]$, we define $\pi'(\underline{x}) := \pi'^{+}(\underline{x})$ and $\pi'(\bar{x}) := \pi'^{-}(\bar{x})$. For a concave function π, we interpret the statement "$\pi'(v) = k$" as "k is a super-derivative of π at v." Similarly, we interpret inequality statements such as "$\pi'(v) > k$" as "there exists an element of the super-differential of π at v, which is greater than k." For reference, for a concave function $\pi(x)$ defined on $[\underline{x}, \bar{x}]$, we say that k is a super-derivative of π at v if $\pi(x) - \pi(v) \le k(x - v)$ for all $x \in (\underline{x}, \bar{x})$. The set of all superderivatives of π at v is called the superdifferential of π at v and is guaranteed (by converse of the mean value theorem) to be equal to $[\pi'^{+}(v), \pi'^{-}(v)]$.
4. For a real function $F(\cdot)$, we define its generalized inverse as $F^{-1*}(y) := \sup\{x \in \mathrm{dom}F | F(x) \le y\}$.

5. Strict (weak) first-order stochastic dominance is denoted by $\succ_{FSD}$ ($\succeq_{FSD}$), while second-order stochastic dominance is denoted by $\succ_{SSD}$ ($\succeq_{SSD}$).

6. For any cdf $G(\cdot)$, we denote the corresponding decumulative distribution function by $\bar{G}(\cdot) := 1 - G(\cdot)$

Throughout the proofs, we assume $\mathcal{Y}_t = [\underline{y}_t, \infty)$ for some $\underline{y}_t \geq 0$. This is more general than the assumptions used in the paper (where $\underline{y}_t = 0$), but this generalization facilitates the proofs of Propositions 7–8.

Proof of Proposition 1. The proposition follows directly from the geometric construction outlined below its statement in the main text. Furthermore, if $\hat{v}_t^{\mathcal{C}} > \underline{v}_t^a$ (resp. $\hat{v}_t^{\mathcal{O}} < \overline{v}_t^a$), the slope of the tangent passing through $(\underline{v}_t^a, \mathcal{C}_t)$ and $(\hat{v}_t^{\mathcal{C}}, v_t^e(\hat{v}_t^{\mathcal{C}}))$ (resp., through $(\overline{v}_t^a, \mathcal{O}_t)$ and $(\hat{v}_t^{\mathcal{O}}, v_t^e(\hat{v}_t^{\mathcal{O}})))$ is given by

$$s_t^{\mathcal{C}} := \frac{\pi_t^e(\hat{v}_t^{\mathcal{C}}) - \mathcal{C}_t}{\hat{v}_t^{\mathcal{C}} - \underline{v}_t^a} \qquad \left(s_t^{\mathcal{O}} := \frac{\mathcal{O}_t - \pi_t^e(\hat{v}_t^{\mathcal{O}})}{\overline{v}_t^a - \hat{v}_t^{\mathcal{O}}} \right).$$

With this in mind, we can write the recursive expression for the firm's value function as

$$\pi_t^a(v_t^a) = \begin{cases} \mathcal{C}_t + s_t^{\mathcal{C}}(v_t^a - \underline{v}_t^a) & \text{if } \underline{v}_t^a \leq v_t^a < \hat{v}_t^{\mathcal{C}}, \\ \pi_t^e(v_t^a) & \text{if } \hat{v}_t^{\mathcal{C}} \leq v_t^a \leq \hat{v}_t^{\mathcal{O}}, \\ \pi_t^e(\hat{v}_t^{\mathcal{O}}) + s_t^{\mathcal{O}}(v_t^a - \hat{v}_t^{\mathcal{O}}) & \text{if } \hat{v}_t^{\mathcal{O}} < v_t^a \leq \overline{v}_t^a. \end{cases} \qquad \square \tag{40}$$

Proof of Proposition 2. Substituting the expression for v_t^a from (16) into (15) reduces the problem (15)–(16) to $\max_{f_t \in [0,1]} \pi_t^a(v_t^p - f_t A)$. Using the property established in Proposition 1 that π_t^a is a concave function gives (17). Inserting the optimal fractional access (17) into (16) yields the optimal continuation value (18). Lastly, inserting (17) into (16) yields $\pi_t^p(v_t^p) = \pi_t^a(v_t^p - \min\{[v_t^p - \hat{v}_t^{\mathcal{A}}]^+, A\})$, which is also concave. $\square$

The next lemma is a general (model-independent) result and is used in the proofs of Propositions 3 and 4. Similar results are given in Levy (1992) and Gao et al. (2018).

Lemma 1 (Stochastic Dominance via Slope). *Let X be a continuous random variable with convex support $\mathcal{X} \subseteq \mathbb{R}$ and cdf F. Let $g, h : \mathcal{X} \to \mathbb{R}$ be absolutely continuous and non-decreasing functions such that $\mathbb{E}[h(X)] \geq \mathbb{E}[g(X)]$. If g and h are right-differentiable and $h'^{+}(x) \leq g'^{+}(x)$, $\forall x \in \mathcal{X} \setminus \{\sup \mathcal{X}\}$ then $h(X)$ weakly dominates $g(X)$ in terms of second-order stochastic dominance ($h(X) \succeq_{SSD} g(X)$). Alternatively, if g and h are left-differentiable and $h'^{-}(x) \leq g'^{-}(x)$, $\forall x \in \mathcal{X} \setminus \{\inf \mathcal{X}\}$ then also $h(X) \succeq_{SSD} g(X)$.*

Proof. Denoting $\xi = \inf \mathcal{X}$, $\psi = \sup \mathcal{X}$, define $\bar{g}, \bar{h} : [\xi, \psi] \to \bar{\mathbb{R}}$ by

$$\bar{g}(x) = \begin{cases} \lim_{y \to \xi^+} g(y) & | x = \xi, \\ g(x) & | x \in \mathcal{X} \setminus \{\xi, \psi\}, \\ \lim_{y \to \psi^-} g(y) & | x = \psi, \end{cases} \qquad \bar{h}(x) = \begin{cases} \lim_{y \to \xi^+} h(y) & | x = \xi, \\ h(x) & | x \in \mathcal{X} \setminus \{\xi, \psi\}, \\ \lim_{y \to \psi^-} h(y) & | x = \psi. \end{cases}$$

These two functions are just the natural extensions (by continuity) of g and h to the closure of their domains. By definition, the cdf of $h(X)$ is $F_{h(X)}(x) = P(h(X) \leq x) = \int_{\xi}^{h^{-1*}(x)} dF(u) = F(h^{-1*}(x))$, and analogously $F_{g(X)}(x) = F(g^{-1*}(x))$. Then, $(h(X) \succeq_{SSD} g(X)) \Leftrightarrow (\int_{-\infty}^{x} F_{h(X)}(u) - F_{g(X)}(u) du \leq 0, \forall x \in \mathbb{R})$. Notice that $\bar{g}(\xi) \leq \bar{h}(\xi)$, otherwise $h'^{+}(x) \leq g'^{+}(x)$ or $h'^{-}(x) \leq g'^{-}(x)$ would imply $\mathbb{E}[h(X)] > \mathbb{E}[g(X)]$. If $\bar{g}(\psi) \leq \bar{h}(\psi)$ or $\bar{g}(\xi) = \bar{h}(\xi)$, then as $h'^{+}(x) \leq g'^{+}(x)$ or $h'^{-}(x) \leq g'^{-}(x)$ we have $g(x) \leq h(x), \forall x \in \mathcal{X}$, thus $F_{h(X)}(x) \leq F_{g(X)}(x), \forall x \in \mathcal{X}$, which is equivalent to $h(X) \succeq_{FSD} g(X)$, from which $h(X) \succeq_{SSD} g(X)$ follows.

It remains to show that the statement of the lemma holds when both $\bar{g}(\xi) < \bar{h}(\xi)$ and $\bar{g}(\psi) > \bar{h}(\psi)$. In this case, $(h(X) \succeq_{SSD} g(X)) \Leftrightarrow (\int_{\bar{g}(\xi)}^{x} F_{h(X)}(u) - F_{g(X)}(u) du, \forall x \in \{g(y) | y \in \mathcal{X}\})$. Denoting $I(x) := \int_{\bar{g}(\xi)}^{x} F_{h(X)}(u) -$

$F_{g(X)}(u)du$, we proceed to show that the integral condition for SSD holds, i.e., $I(x) \le 0, \forall x \in \{g(y)|y \in \mathcal{X}\}$. We have $I(\bar{g}(\xi)) = 0$, because the integral vanishes. Next, we can establish that $I(\bar{g}(\psi)) \le 0$, by using the property that for any real random variable Y with cdf F_Y it holds that $\mathbb{E}[Y] = \int_0^\infty (1 - F_Y(u))du + \int_{-\infty}^0 F_Y(u)du$, which then implies $0 \ge \mathbb{E}[g(X)] - \mathbb{E}[h(X)] = \int_{-\infty}^\infty F_{h(X)}(u) - F_{g(X)}(u)du = I(\bar{g}(\psi))$. Next, we check the interior of $I(x)$. Because $\bar{g}(\cdot)$ and $\bar{h}(\cdot)$ are continuous functions such that $\bar{g}(\xi) < \bar{h}(\xi)$ and $\bar{g}(\psi) > \bar{h}(\psi)$, there needs to exist a point where they are equal (by the intermediate value theorem). Furthermore, as $h'^{+}(x) \le g'^{+}(x)$ or $h'^{-}(x) \le g'^{-}(x)$, the set of all points on which these functions are equal is a closed interval ($\exists \xi^*, \psi^*$ such that $\xi < \xi^* \le \psi^* < \psi$, where $g(x) = h(x)$, $\forall x \in [\xi^*, \psi^*]$, and $\bar{g}(x) \ne \bar{h}(x), \forall x \in [\xi, \xi^*) \cup (\psi^*, \psi]$). Thus, the cdf-s $F_{h(X)}$ and $F_{g(X)}$ have the single-crossing property where $F(h^{-1*}(x)) \le F(g^{-1*}(x)), \forall x < g(\xi^*)$ and $F(h^{-1*}(x)) \ge F(g^{-1*}(x)), \forall x \ge g(\xi^*)$. Because of this, the integrand in $I(x) = \int_{\bar{g}(\xi)}^x F_{h(X)}(u) - F_{g(X)}(u)du$ is weakly negative for $u \le g(\xi^*)$ and weakly positive for $u \ge g(\xi^*)$. Hence, $I(x)$ is non-increasing up to $x = g(\xi^*)$ and non-decreasing afterwards, which combined with $I(\bar{g}(\xi)) = 0$ and $I(\bar{g}(\psi)) \le 0$ implies that $I(x) \le 0, \forall x \in \{g(y)|y \in \mathcal{X}\}$, completing the proof. $\square$

Proof of Proposition 3. *Part 1: shape of* v_t^{p*}. We establish the shape of v_t^{p*} by showing that the choice of $(p_t(\cdot), v_t^p(\cdot))$ in (19)–(21) can be be restricted to $\mathcal{P}_t \times \bar{\mathcal{V}}_t$ without loss of optimality, where $\bar{\mathcal{V}}_t := \{v_t^{p*}(p; \alpha_t, \kappa_t^v) | \alpha_t \ge \underline{v}_t^p, \kappa_t^v \in \bar{\mathbb{R}}_+\}$, and $v_t^{p*}(p; \alpha_t, \kappa_t^v)$ is given by (22). The proof is based on showing that for any pair $(p_t(\cdot), v_t^p(\cdot))$ that solves (19)–(21), we can construct $v_t^{p*}(\cdot) \in \bar{\mathcal{V}}_t$ such that $(p_t(\cdot), v_t^{p*}(\cdot))$ also solves (19)–(21). Assume $(p_t(\cdot), v_t^p(\cdot))$ solves (19)–(21). Then, from (20) we have $p_t(y) \in \underset{0 \le p \le y}{\operatorname{argmax}}\, u(y-p) + v_t^p(p), \forall y \in \mathcal{Y}_t$, which implies (since $p_t(y)$ is a local maximum)

$$\frac{d}{d^- p}\left(v_t^p(p) + u(y-p)\right)\Big|_{p=p_t(y)} \ge 0, \quad \forall y \in \mathcal{Y}_t \text{ such that } p_t(y) \ne 0, \tag{41}$$

$$\frac{d}{d^+ p}\left(v_t^p(p) + u(y-p)\right)\Big|_{p=p_t(y)} \le 0, \quad \forall y \in \mathcal{Y}_t \text{ such that } p_t(y) \ne y. \tag{42}$$

Next, we will narrow down the region in which $p_t(y)$ lies. Define $\hat{\hat{p}}_t := \sup\{p_t(y)|y \in \mathcal{Y}_t\}$, and $\kappa_t^{v*} := \inf\{p | 0 \le p \le \hat{\hat{p}}_t, \exists r_p > p \text{ such that } v_t^p(x) < v_t^p(p) + \lambda_1(x-p), \forall x \in (p, r_p)\}$. Since both one-sided derivatives of $u(y)$ are either λ_1 or λ_2, from (41) it follows that $\frac{d}{d^- p} v_t^p(p) \ge \lambda_2$, for all p such that $0 < p \le \hat{\hat{p}}_t$. Note that $\hat{\hat{p}}_t < \infty$ for any incentive compatible $p_t(y)$, as the amount of utility the principal can offer the agent is bounded ($v_t^p(y) \le \bar{v}_t^p, \forall y \in \mathcal{Y}_t$) but consumption utility is not ($\lim_{y\to\infty} u(y) = \infty$).

First, since from the definition of κ_t^{v*} we have $\frac{d}{d^+ p} v_t^p(p) \ge \lambda_1, \forall p \in [0, \kappa_t^{v*})$, for incomes below κ_t^{v*} any incentive compatible payment policy will leave at most κ for consumption. Thus, $\min\{\hat{\hat{p}}_t, [y-\kappa]^+\} \le p_t(y) \le \min\{\hat{\hat{p}}_t, y\}, \forall y \le \kappa_t^{v*}$. Second, from the definition of κ_t^{v*} we also have that for $y_t \in (\kappa_t^{v*}, r_p)$ it holds that $p_t(y_t) < y_t$, i.e. will be optimal for the consumer to divert for personal consumption the marginal income above κ_t^{v*}. Since (42) implies that $\frac{d}{d^+ p} v_t^p(p) \le \lambda_1$ for every p for which there exists $y_p \in \mathcal{Y}_t$ such that $p = p_t(y_p) < y_p$, we have that $\frac{d}{d^+ p} v_t^p(p) \le \lambda_1$ for all $p > \kappa_t^{v*}$, so $\min\{\hat{\hat{p}}_t, [y-\kappa]^+\} \le p_t(y) \le \min\{\kappa_t^{v*}, \hat{\hat{p}}_t\}\}, \forall y \in (\kappa_t^{v*}, \kappa_t^{v*} + \kappa]$, i.e. the consumer will divert for consumption *all* income above κ_t^{v*} until she exhausts her ability to receive marginal utility λ_1 from consumption. Third, for incomes above $\kappa_t^{v*} + \kappa$, because of the $\frac{d}{d^+ p} v_t^p(p) \le \lambda_1$ for all $p > \kappa_t^{v*}$ condition, the consumer will either prefer consuming any marginal income or be indifferent between consuming it and paying it to the firm, so $p_t(y) \in [\min\{\kappa_t^{v*}, \hat{\hat{p}}_t\}, \min\{\hat{\hat{p}}_t, y-\kappa\}], \forall y \ge \kappa_t^{v*}$. Noticing that $\hat{\hat{p}}_t \ge \kappa_t^{v*}$ and combining the three cases yields that $p_t(y) \in [\min\{[y-\kappa]^+, \kappa_t^{v*}\}, \max\{y-\kappa, \hat{\hat{p}}_t\}], \forall y \in \mathcal{Y}_t$.

Now, consider $v_t^{p*}(p;\alpha_t,\kappa_t^{v*})$, as given by (22). Noticing that $\hat{p}_{0,t}(\alpha_t,\kappa_t^{v*}) \geq \hat{\hat{p}}_t, \forall \alpha_t \in \mathbb{R}$ we can see that *any* payment policy $\acute{p}_t$ that satisfies $\acute{p}_t(y) \in [\min\{[y-\kappa]^+,\kappa_t^{v*}\}, \max\{y-\kappa, \hat{p}_{0,t}(\alpha_t,\kappa_t^{v*})], \forall y \in \mathcal{Y}_t$ satisfies the incentive compatibility constraint (20). (This region is the Space of Incentive Compatible Payment Functions illustrated Figure 3, panel (b), when $\kappa_t^v = \kappa_t^{v*}$; it can also be written using notation of footnote 2 as $\underline{p}_t(y) \leq \acute{p}(y) \leq \bar{p}_t(y)$.) Thus the starting optimal payment policy p_t, which was just shown to fall in this region, is also incentive compatible under $v_t^{p*}(p;\alpha_t,\kappa_t^{v*})$. Furthermore, we have that $\forall y \in \mathcal{Y}_t, p \in [\min\{[y-\kappa]^+,\kappa_t^{v*}\}, \max\{y-\kappa,\kappa_t^{v*}\}]$:

$$v_t^{p*}(p;\alpha_t,\kappa_t^{v*}) + u(y-p) = \begin{cases} \lambda_1 y & \text{if } y \leq \kappa + \kappa_t^{v*}, \\ \lambda_1(\kappa_t^{v*}+\kappa) + \lambda_2(y-\kappa_t^{v*}-\kappa) & \text{otherwise.} \end{cases} \tag{43}$$

Taking a left derivative of (43) with respect to p yields that (41) holds for $v_t^p(\cdot) = v_t^{p*}(\cdot;\alpha_t,\kappa_t^{v*})$ with equality. Hence, from (41) we have that $\frac{d}{d^-p}v_t^p(p)|_{p=p_t(y)} \geq \frac{d}{d^-p}v_t^{p*}(p;\alpha_t,\kappa_t^{v*})\big|_{p=p_t(y)}, \forall y \in \mathcal{Y}_t$ such that $p_t(y) \neq 0, \forall \alpha_t \in \mathbb{R}$, thus also $\frac{d}{d^-y}v_t^p(p_t(y)) \geq \frac{d}{d^-y}v_t^{p*}(p_t(y);\alpha_t,\kappa_t^{v*}), \forall y \in \mathcal{Y}_t$ such that $p_t(y) \neq 0, \forall \alpha_t \in \mathbb{R}$. Now, we identify the value of α_t that will make the promise-keeping constraint (21) binding. Denoting this value by α_t^*, we can find it by solving $\mathbb{E}[u(Y_t - p_t(Y_t)) + v_t^{p*}(p_t(Y_t);\alpha_t,\kappa_t^{v*})] = \mathbb{E}[u(Y_t - p_t(Y_t)) + v_t^p(p_t(Y_t))]$, which yields $\alpha_t^* = \mathbb{E}[v_t^p(p_t(Y_t)) - v_t^{p*}(p_t(Y_t);0,\kappa_t^{v*})]$. So, we have that $\mathbb{E}[v_t^{p*}(p_t(Y_t);\alpha_t^*,\kappa_t^{v*})] = \mathbb{E}[v_t^p(p_t(Y_t))]$ and $\frac{d}{d^-y}v_t^p(p_t(y)) \geq \frac{d}{d^-y}v_t^{p*}(p_t(y);\alpha_t^*,\kappa_t^{v*}), \forall y \in \mathcal{Y}_t$ such that $p_t(y) \neq 0$, which allows us to apply Lemma 1 and obtain $v_t^{p*}(p_t(Y_t);\alpha_t^*,\kappa_t^{v*}) \succeq_{SSD} v_t^p(p_t(y))$.

Here, we can use the result of Rothschild and Stiglitz (1970) that for any two random variables X and Y, it holds that ($X \succeq_{SSD} Y$ and $\mathbb{E}[X] = \mathbb{E}[Y]$) $\Leftrightarrow$ (Y is a mean preserving spread of X) $\Leftrightarrow$ ($\mathbb{E}[q(X)] \geq \mathbb{E}[q(Y)], \forall$ concave $q(\cdot)$). Thus, from concavity of π_t^p we have $\mathbb{E}[\pi_t^p(v_t^{p*}(p_t(Y_t);\alpha_t^*,\kappa_t^{v*})) + p_t(Y_t)] \geq \mathbb{E}[\pi_t^p(v_t^p(p_t(Y_t))) + p_t(Y_t)]$, so the pair (p_t, v_t^{p*}) performs at least as good as (p_t, v_t^p) in the objective function (19), while both pairs satisfy the constraints (20)–(21), implying the optimality of (p_t, v_t^{p*}).

Lastly, a technical remark is that without loss of optimality, we can restrict attention to finite κ_t^v (so $\kappa_t^v \in \mathbb{R}_+$ rather than $\bar{\mathbb{R}}_+$), because in problem (19)–(21), for every α_t, both the firm and the consumer are indifferent between all $v_t^{p*}(p;\alpha_t,\kappa_t^v)$ such that $\kappa_t^v \geq (\bar{v}_t - \alpha_t)/\lambda_1$.

Part 2: construction of p_t^*. We present the proof of this part of the proposition by construction, but note that the same result can be also be obtained by pointwise optimization, analagously to Part 2 of Proposition 7. The intution of how this alternative method works is also given in the discussion following Proposition 3.

First, we establish an intermediary result. Let, $(p_t(y), v_t^{p*}(p;\alpha_t,\kappa_t^v)) \in \mathcal{P}_t \times \bar{\mathcal{V}}_t$ satisfy (20)–(21). We will construct a piecewise-linear $p_t^\dagger(y;\alpha_t;\kappa_t^v) \in \mathcal{P}_t$, such that $(p_t^\dagger(y;\alpha_t;\kappa_t^v), v_t^{p*}(p;\alpha_t,\kappa_t^v))$ performs equal or better to $(p_t(y), v_t^{p*}(p;\alpha_t,\kappa_t^v))$ in (19)–(21). Note that under $v_t^{p*}(p;\alpha_t,\kappa_t^v)$, the consumer is indifferent between all $\hat{p}_t(y)$ such that $\hat{p}_t(y) = p_t(y)$ for $y \leq \kappa + \kappa_t^v$ and $\hat{p}_t(y) \in [\kappa_t^v, y-\kappa]$ for $y > \kappa + \kappa_t^v$. This is because once the consumer has paid κ_t^v to the firm and has consumed κ, she will derive marginal utility λ_2 from any additional money, irrespective of whether she consumes it or pays it to the firm. Let $\hat{p}_{2,t}$ be as given by (25), and define

$$p_t^\dagger(y;\alpha_t;\kappa_t^v) := \begin{cases} p_t(y) & \text{if } y \leq \kappa + \kappa_t^v, \\ y - \kappa & \text{if } \kappa_t^v + \kappa \leq y \leq \hat{p}_{2,t}(\alpha_t,\kappa_t^v) + \kappa, \\ \hat{p}_{2,t}(\alpha_t,\kappa_t^v) & \text{if } y \geq \hat{p}_{2,t}(\alpha_t,\kappa_t^v) + \kappa. \end{cases} \tag{44}$$

The idea behind this construction is that under $v_t^{p*}(\cdot;\alpha_t,\kappa_t^v)$, after the firm has received κ_t^v money, it needs to promise λ_2 utility to the agent for every additional unit of money received; as $\pi_t^p(\cdot)$ is concave, the "cost of fulfilling promises" is convex, thus the consumer giving more money to the firm is beneficial to the firm, but only up to some point ($\hat{p}_{2,t}$). Thus, $p_t^\dagger(y;\alpha_t;\kappa_t^v)$ is constructed as a modification of $p_t(y)$ that changes how the consumer uses money in excess of $\kappa+\kappa_t^v$: the indifference between consuming money at marginal utility λ_2 and or giving it to the firm (at marginal promised future utility λ_2) is broken in favor of giving the money to the firm, but only until the consumer has given $\hat{p}_{2,t}$ money to the firm, after which consumer consumes any additional money. Notice that $p_t(y)$ satisfies $p_t(\kappa+\kappa_t^v)=\kappa_t^v$. Thus $p_t^\dagger(y;\alpha_t;\kappa_t^v)$ is continuous, so $p_t^\dagger(y;\alpha_t;\kappa_t^v)\in\mathcal{P}_t$. Because of the aforementioned indifference property, under $v_t^{p*}(p;\alpha_t,\kappa_t^v)$ the agent derives the same utility when using $p_t^\dagger(y;\alpha_t;\kappa_t^v)$ as with $p_t(y)$, thus the pair $(p_t^\dagger(y;\alpha_t;\kappa_t^v),v_t^{p*}(p;\alpha_t,\kappa_t^v))$ also satisfies the constraints (20)–(21).

Now, to show that that the firm prefers $p_t^\dagger$ over p_t, we first establish one property: that $p_t\le y-\kappa,\forall y>\kappa_t^v$. If this were not the case ($\exists y^*>\kappa_t^v$, s.t. $p_t(y^*)>y^*-\kappa$), following p_t when her income is y^* would give the consumer expected utility $v_t^{p*}(p_t(y^*);\alpha_t,\kappa_t^v)+u(y^*-p_t(y^*))=\lambda_1\kappa_t^v+\lambda_2(p_t(y^*)-\kappa_t^v)+\lambda_1\min\{y^*-p_t(y^*),\kappa\}+\lambda_2[y^*-p_t(y^*)-\kappa]^+=\lambda_1(\kappa_t^v+y^*-p_t(y^*))+\lambda_2(p_t(y^*)-\kappa_t^v)<\lambda_1(\kappa+\kappa_t^v)+\lambda_1(y^*-\kappa-\kappa_t^v)$, but the consumer could get utility $\lambda_1(\kappa+\kappa_t^v)+\lambda_2(y^*-\kappa-\kappa_t^v)$ by paying the firm κ_t^v and consuming the rest, in violation of the incentive compatibility constraint (20). Now, we shall establish that for any $y\in\mathcal{Y}_t$,

$$\pi_t^p(v_t^{p*}(p_t^\dagger(y;\alpha_t;\kappa_t^v);\alpha_t,\kappa_t^v))+p_t^\dagger(y;\alpha_t;\kappa_t^v)\ge\pi_t^p(v_t^{p*}(p_t(y);\alpha_t,\kappa_t^v))+p_t(y). \tag{45}$$

For $y\in[0,\kappa+\kappa_t^v]\cap\mathcal{Y}_t$, we have $p_t^\dagger(y;\alpha_t;\kappa_t^v)=p_t(y)$ from (44), so (45) follows. For $y\in[\kappa+\kappa_t^v,\hat{p}_{2,t}+\kappa]\cap\mathcal{Y}_t$, from $p_t(y)\le y-\kappa$ and (44) we have $y-\kappa=p_t^\dagger(y;\alpha_t;\kappa_t^v)\ge p_t(y)\ge\kappa_t^v$, from which (45) follows again as $\pi_t^p(v_t^{p*}(p;\alpha_t,\kappa_t^v))+p$ is non-decreasing on $p\in[\kappa_t^v,\hat{p}_{2,t}(\alpha_t,\kappa_t^v)]$ because it is concave with $\hat{p}_{2,t}(\alpha_t,\kappa_t^v)\in\arg\max_p\pi_t^p(v_t^{p*}(p;\alpha_t,\kappa_t^v))+p$. Finally, for $y\in[\hat{p}_{2,t}(\alpha_t,\kappa_t^v)+\kappa,\infty)\cap\mathcal{Y}_t$, (45) follows from $p_t^\dagger(y;\alpha_t;\kappa_t^v)=\hat{p}_{2,t}(\alpha_t,\kappa_t^v)\in\arg\max_p\pi_t^p(v_t^{p*}(p;\alpha_t,\kappa_t^v))+p$. So, as (45) holds for all y, we also have $\pi_t^p(v_t^{p*}(p_t^\dagger(Y_t;\alpha_t;\kappa_t^v);\alpha_t,\kappa_t^v))+p_t^\dagger(Y_t;\alpha_t;\kappa_t^v))\succeq_{FSD}\pi_t^p(v_t^{p*}(p_t(Y_t);\alpha_t,\kappa_t^v))+p_t(Y_t)$, thus $(p_t^\dagger,v_t^{p*})$ performs at least as good as (p_t,v_t^{p*}) in the objective function (19).

We constructed $p_t^\dagger(y)$ from $p_t(y)$ by linearizing the segment where payments and consumption both yield marginal utility λ_2 to the agent. We do the same process again to transform the segment where payments and consumption both yield marginal utility λ_1 (for $y\in[0,\kappa+\kappa_t^v]$). The resulting function p_t^* will have the property that (p_t^*,v_t^{p*}) performs at least as well as (p_t,v_t^{p*}) in (19)–(21). So, let $p_t^*(y;\alpha_t,\kappa_t^v)$ be the one given by (23). Analogously to the proof of the same properties for $p_t^\dagger(y;\alpha_t;\kappa_t^v)$, it follows that (p_t^*,v_t^{p*}) satisfies constraints (20)–(21) (due to the agent's indifference between $p_t^\dagger$ and p_t^* under v_t^{p*}), while $\pi_t^p(v_t^{p*}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v))+p_t^*(Y_t;\alpha_t,\kappa_t^v)\succeq_{FSD}\pi_t^p(v_t^{p*}(p_t^\dagger(Y_t;\alpha_t;\kappa_t^v));\alpha_t,\kappa_t^v))+p_t^\dagger(Y_t;\alpha_t;\kappa_t^v))$, so $p_t^*(y;\alpha_t,\kappa_t^v)$ is the optimal payment policy when using the reward function $v_t^{p*}(p;\alpha_t;\kappa_t)$. □

Proof of Proposition 4. *Preliminaries.* We first introduce some auxiliary notation to be used in the proof. Let $\Psi(\omega_t):=\omega_t-\lambda_1\kappa_t^v(\omega_t)$; let $\hat{\omega}_{1,t}(v_t^b)$ be an implicit function defined by $\Psi(\omega_t)-v_t^b+\Lambda_{1,t}=0$; let $\hat{\omega}_{2,t}(v_t^b)$ be an implicit function defined by $\Psi(\omega_t)-v_t^b+\Lambda_{2,t}=0$; let $h_t(\omega_t):=\omega_t-(\lambda_1-\lambda_2)\kappa_t^v(\omega_t)$; and let $p_{1,t}(v_t^b,\omega_t):=\max_{p\in\mathbb{R}_+}\pi_t^p(v_t^b-\omega_t+\lambda_1 p)+p$ subject to $v_t^b-\omega_t+\lambda_1 p\in[\underline{v}_t^p,\overline{v}_t^p]$.

Part (i): if $v_t^b \le \hat{v}_t^b$ then $\omega_t^*(v_t^b) \in \Re$, otherwise $\omega_t^*(v_t^b) = \underline{\omega}_t$. We start by showing that the optimal values of (α_t, κ_t^v) in (26)–(27) satisfy $\kappa_t^v = \hat{p}_{1,t}(\alpha_t, \kappa_t^v)$ (thus also the optimal ω_t satisfies $\kappa_t^v(\omega_t) = \hat{p}_{1,t}(v_t^b - \omega_t, \kappa_t^v(\omega_t))$). To do that, we first demonstrate that for any $p_t^*(y;\alpha_t,\kappa_t^v)$ and $v_t^{p*}(p;\alpha_t,\kappa_t^v)$, as given by Proposition 3, that solve (19)–(21), we can construct $\alpha_t^{**}(\alpha_t,\kappa_t^v)$ such that $p_t^*(y;\alpha_t,\kappa_t^v), v_t^{p*}(p;\alpha_t^{**}(\alpha_t,\kappa_t^v),\hat{p}_{1,t}(\alpha_t,\kappa_t^v))$ solve the same problem. Let $v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v) := v_t^{p*}(p;\alpha_t^{**}(\alpha_t,\kappa_t^v),\hat{p}_{1,t}(\alpha_t,\kappa_t^v))$ where $\alpha_t^{**}(\alpha_t,\kappa_t^v) := \mathbb{E}[v_t^{p*}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v) - v_t^{p*}(p_t^*(Y_t;\alpha_t,\kappa_t^v);0,\hat{p}_{1,t}(\alpha_t,\kappa_t^v))]$. In other words, $v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v)$ is a modified version of $v_t^{p*}(p)$ which (a) changes the agent's marginal reward for paying money to the firm from λ_1 to λ_2 on the $p \in [\hat{p}_{1,t}(\alpha_t,\kappa_t^v),\kappa_t^v]$ segment, (b) changes the constant α_t by increasing it to compensate for the loss of expected utility resulting from the change under (a) (to ensure the promise-keeping constraint (21) still binds). Then, combining incentive compatibility (20) with (43) we have that under $v_t^{p\dagger\dagger}(p)$, any payment function $\check{p}_t \in \mathcal{P}_t$ that satisfies $\check{p}_t(y) \in [\min\{(y-\kappa)^+, \hat{p}_{1,t}(\alpha_t,\kappa_t^v)\}, \max\{y-\kappa, \hat{p}_{1,t}(\alpha_t,\kappa_t^v)\}]$ is incentive compatible. From (23) it follows that $p_t^*(y;\alpha_t,\kappa_t^v)$ satisfies this criterion, so $p_t^*(y;\alpha_t,\kappa_t^v), v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v)$ satisfy (20). That $p_t^*(y;\alpha_t,\kappa_t^v), v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v)$ satisfy the promise-keeping constraint (21) follows directly from $p_t^*(y;\alpha_t,\kappa_t^v), v_t^{p*}(p;\alpha_t,\kappa_t^v)$ satisfying it and the construction of $\alpha_t^{**}(\alpha_t,\kappa_t^v)$. It remains to show that the firm receives at least as much expected utility under $p_t^*(y;\alpha_t,\kappa_t^v)$ and $v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v)$ as under $p_t^*(y;\alpha_t,\kappa_t^v)$ and $v_t^{p*}(p;\alpha_t,\kappa_t^v)$. We have that $\frac{d}{d^+p} v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v) \le \frac{d}{d^+p} v^{p*}(p;\alpha_t,\kappa_t^v), \forall y \in \mathcal{Y}_t$ (from the construction of $v_t^{p\dagger\dagger}$, the slopes of $v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v)$ and $v_t^{p*}(p;\alpha_t,\kappa_t^v)$ are the same everywhere except on $[\hat{p}_{1,t}(\alpha_t,\kappa_t^v),\kappa_t^v)$, where $v_t^{p\dagger\dagger}$ is less steep). Then, $\frac{d}{d^+y}\left(v_t^{p\dagger\dagger}(p_t^*(y;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)\right) \le \frac{d}{d^+y}\left(v_t^{p*}(p_t^*(y;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)\right), \forall y \in \mathcal{Y}_t$, thus from Lemma 1 it follows that $v_t^{p\dagger\dagger}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v) \succeq_{SSD} v_t^{p*}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)$. As for any two random variables X, Y, ($X \succeq_{SSD} Y$ and $\mathbb{E}[X] = \mathbb{E}[Y]$) $\Leftrightarrow$ (Y is a mean preserving spread of X) $\Leftrightarrow$ ($\mathbb{E}[q(X)] \ge \mathbb{E}[q(Y)], \forall$ concave $q(\cdot)$) (Rothschild and Stiglitz 1970), from concavity of $\pi_t^p(v_t^p)$ we have $\mathbb{E}[\pi_t^p(v_t^{p\dagger\dagger}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)) + p_t^*(Y_t;\alpha_t,\kappa_t)] \ge \mathbb{E}[\pi_t^p(v_t^{p\dagger\dagger}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)) + p_t^*(Y_t;\alpha_t,\kappa_t^v)]$, thus the pair $(p_t^*(y;\alpha_t,\kappa_t^v), v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v))$ performs at least as good as $(p_t^*(y;\alpha_t,\kappa_t^v), v_t^{p*}(p;\alpha_t,\kappa_t^v))$ in the objective function (19), while both pairs satisfy the constraints (20)–(21), implying the optimality of $(p_t^*(y;\alpha_t,\kappa_t^v), v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v))$. However, notice from Proposition 3 that the optimal payment function, given a reward function $v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v)$, is $p_t^*(y;\alpha_t^{**}(\alpha_t,\kappa_t^v),\hat{p}_{1,t}(\alpha_t,\kappa_t^v))$, implying that the initial choice of optimal (α_t,κ_t^v) satisfies $\kappa_t^v = \hat{p}_{1,t}(\alpha_t,\kappa_t^v)$.

Then, the optimal ω_t satisfies $\kappa_t^v(\omega_t) = \hat{p}_{1,t}(v_t^b - \omega_t, \kappa_t^v(\omega_t)) \le p_{1,t}(v_t^b,\omega_t) = (\Lambda_{1,t} - (v_t^b - \omega_t))/\lambda_1$, which implies $\omega_t - \lambda_1\kappa_t^v(\omega_t) \ge v_t^b - \Lambda_{1,t} \Rightarrow \Psi(\omega_t) \ge \Psi(\hat{\omega}_{1,t}(v_t^b)) = 1 - \Lambda_{1,t} \Rightarrow \omega_t \le \hat{\omega}_{1,t}(v_t^b)$. Since $\hat{\omega}_{1,t}(v_t^b)$ is given implicitly by the ω that solves $\omega - \lambda_1\kappa_t^v(\omega) - v_t^b + \Lambda_{1,t} = 0$, we have $\omega_t - \lambda_1\kappa_t^v(\omega_t) - v_t^b + \Lambda_{1,t} \ge 0$ (since $\omega - \lambda_1\kappa_t^v(\omega)$, is decreasing in ω), rearranging which yields the $\omega_t - \lambda_1\kappa_t^v(\omega_t) \ge v_t^b - \Lambda_{1,t}$ condition in the definition of $\Re$ (which is equivalent to $\omega_t \le \hat{\omega}_{1,t}(v_t^b)$). Differentiating $\hat{\omega}_{1,t}(v_t^b)$ implicitly yields $\hat{\omega}'_{1,t}(v_t^b) = 1/\Psi'(\hat{\omega}_{1,t}(v_t^b))$, so $\hat{\omega}_{1,t}(v_t^b)$ is non-increasing and concave since $\Psi(\omega)$ is non-increasing and concave. Defining $\hat{v}_t^b := \max\{\underline{v}_t^b, \sup\{v \in [\underline{v}_t^b, \bar{v}_t^b] | \hat{\omega}_{1,t}(v_t^b) > \underline{\omega}_t\}\}$ and observing that $\alpha_t \ge \underline{v}_t^b$ implies $\omega_t \le v_t^b - \underline{v}_t^b$ yields that the optimal $(v_t^b, \omega_t) \in \Re$ if $v_t^b \le \hat{v}_t^b$ and $\omega_t = \underline{v}_t^b$ otherwise. Lastly, since $\hat{\omega}_{1,t}(v_t^b)$ is concave on $v_t^b \in [\underline{v}_t^b, \hat{v}_t^b]$, the convexity of $\Re$ follows from it being an intersection of two convex sets, more precisely $\Re = \Re_1 \cap \Re_2$, where $\Re_1$ is the hypograph of $\hat{\omega}_{1,t}(v_t^b)|_{[\underline{v}_t^b,\hat{v}_t^b]}$, and $\Re_2 := \{(v_t^b,\omega_t) \in [\underline{v}_t^b, \hat{v}_t^b] \times [\underline{\omega}_t, \bar{\omega}_t] | \omega_t \le v_t^b - \underline{v}_t^b\}$.

Part (ii): concavity of π_t^{b*}. We leverage the property that any univariate concave function on a closed interval can be approximated arbitrarily closely by a smooth concave function (the Convex Smoothing; e.g., see Azagra 2013). Specifically, the property implies that for any $\epsilon > 0$ there exists a smooth (i.e., of class C^∞) and strictly concave function $\pi_t^{p\epsilon}(v)$ such that $\pi_t^p(v) - \epsilon \le \pi_t^{p\epsilon}(v) \le \pi_t^p(v)$, for all $v \in [\underline{v}_t^p, \bar{v}_t^p]$. We will replace π_t^p in (30)–(31) with $\pi_t^{p\epsilon}$; the solution of this problem will then be π_t^{b*}. First, we note that the first part of this proof works in an identical way if we replace $\pi_t^p(v)$ with $\pi_t^{p\epsilon}(v)$ (assuming $\pi_t^p(v)$ in the definition $\Lambda_{j,t}, j \in \{1,2\}$ is also replaced with $\pi_t^{p\epsilon}(v)$). Second, because of the property that under the optimum of (30)–(31), $\kappa_t^v = \hat{p}_{1,t}(\alpha_t, \kappa_t^v)$ holds, we can replace the $\hat{p}_{1,t}(v_t^b - \omega_t, \kappa_t^v(\omega_t))$ in p_t^* with $\kappa_t(\omega_t)$ without changing the optimum; or more explicitly, we can replace $p_t^*(y; \alpha_t, \kappa_t)$ with

$$p_t^{\kappa *}(y; \alpha_t, \kappa_t) := \begin{cases} y & \text{for} \quad 0 \le y < \kappa_t^v, \\ \kappa_t^v & \text{for} \quad \kappa_t^v \le y < \kappa_t^v + \kappa, \\ y - \kappa & \text{for} \quad \kappa_t^v + \kappa \le y < \hat{p}_{2,t}(\alpha_t, \kappa_t^v) + \kappa, \\ \hat{p}_{2,t}(\alpha_t, \kappa_t^v) & \text{for} \quad y \ge \hat{p}_{2,t}(\alpha_t, \kappa_t^v) + \kappa. \end{cases}$$

Putting these changes together, we will solve the following modified version of (30)–(31):

$$\pi_t^{b*}(v_t^b) = \max_{\omega_t \in [\underline{\omega}_t, \overline{\omega}_t]} \tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t), \quad \text{where} \tag{46}$$

$$\tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t) = \mathbb{E}\big[p_t^{\kappa *}(Y_t\,;\, v_t^b - \omega_t,\, \kappa_t^v(\omega_t)) + \pi_t^{p\epsilon}(v_t^{p*}(p_t^{\kappa *}(Y_t\,;\, v_t^b - \omega_t,\, \kappa_t^v(\omega_t))\,;\, v_t^b - \omega_t,\, \kappa_t^v(\omega_t)))\big]. \tag{47}$$

The proof proceeds in three steps: first, we will show that $\pi_t^{b*}(v_t^b)$ is concave on $[\underline{v}_t^b, \hat{v}_t^b]$; second, we will show concavity on $[\hat{v}_t^b, \bar{v}_t^b]$; and finally, we will show that the transition between these two segments preserves concavity over the whole domain as $(\pi_t^{b*})'^-(\hat{v}_t^b) \ge (\pi_t^{b*})'^+(\hat{v}_t^b)$.

Proceeding with the first step, we note that when $(v_t^b, \omega_t) \in \Re$, then $\omega_t \le \hat{\omega}_{1,t}(v_t^b) \le \hat{\omega}_{2,t}(v_t^b)$, thus $\Psi(\omega_t) \ge \Psi(\hat{\omega}_{2,t}) = v_t^b - \Lambda_{2,t}$, so $\omega(t) - (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t) - \lambda_2 \kappa_t^v(\omega_t) \ge \omega_t - \Lambda_{2,t}$, rearranging which yields $\kappa_t^v(\omega_t) \le (\Lambda_{2,t} - v_t^b + \omega_t - (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t))/\lambda_2$. Thus, from (25) we have $\hat{p}_{2,t}(v_t^b - \omega_t, \kappa_t^v(\omega_t)) = (\Lambda_{2,t} - v_t^b + \omega_t - (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t))/\lambda_2 =: p_{2,t}(v_t^b, \omega_t)$. Inserting this last equation and the expression for $p_t^{\kappa *}$ into (47) we obtain

$$\begin{aligned} \tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t) = &\int_0^{\kappa_t^v(\omega_t)} \big[\pi_t^{p\epsilon}(v_t^b - \omega_t + \lambda_1 y) + y\big]\, dG_t(y) \\ &+ \int_{\kappa_t^v(\omega_t)}^{\kappa_t^v(\omega_t)+\kappa} \big[\pi_t^{p\epsilon}(v_t^b - \omega_t + \lambda_1 \kappa_t^v(\omega_t) + \kappa_t^v(\omega_t)\big]\, dG_t(y) \\ &+ \int_{\kappa_t^v(\omega_t)+\kappa}^{p_{2,t}(v_t^b,\omega_t)+\kappa} \big[\pi_t^{p\epsilon}(v_t^b - \omega_t + (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t) + \lambda_2(y - \kappa)) + y - \kappa\big]\, dG_t(y) \\ &+ \int_{p_{2,t}(v_t^b,\omega_t)+\kappa}^{\infty} \big[\pi_t^{p\epsilon}(v_t^b - \omega_t + (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t) + \lambda_2 p_{2,t}(v_t^b, \omega_t)) + p_{2,t}(v_t^b, \omega_t)\big]\, dG_t(y). \end{aligned} \tag{48}$$

Taking a partial derivative of this expression with respect to v_t^b yields

$$\begin{aligned} \frac{\partial \tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t)}{\partial v_t^b} = &\int_0^{\kappa_t^v(\omega_t)} (\pi_t^{p\epsilon})'(v_t^b - \omega_t + \lambda_1 y) dG_t(y) \\ &+ (\pi_t^{p\epsilon})'(v_t^b - \omega_t + \lambda_1 \kappa_t^v(\omega_t))[G_t(\kappa + \kappa_t^v(\omega_t)) - G_t(\kappa_t^v(\omega_t))] \\ &+ \int_{\kappa_t^v(\omega_t)+\kappa}^{p_{2,t}(v_t^b,\omega_t)+\kappa} (\pi_t^{p\epsilon})'(v_t^b - h_t(\omega_t) + \lambda_2(y - \kappa)) dG_t - \frac{1}{\lambda_2} \bar{G}_t(p_{2,t}(v_t^b, \omega_t) + \kappa), \end{aligned} \tag{49}$$

$$\begin{aligned} \frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t)}{\partial (v_t^b)^2} = &\int_0^{\kappa_t^v(\omega_t)} (\pi_t^{p\epsilon})''(v_t^b - \omega_t + \lambda_1 y) dG_t(y) + (\pi_t^{p\epsilon})''(v_t^b - \omega_t + \lambda_1 \kappa_t^v(\omega_t))[G_t(\kappa + \kappa_t^v(\omega_t)) - G_t(\kappa_t^v(\omega_t))] \\ &+ \int_{\kappa_t^v(\omega_t)+\kappa}^{p_{2,t}(v_t^b,\omega_t)+\kappa} (\pi_t^{p\epsilon})''(v_t^b - h_t(\omega_t) + \lambda_2(y - \kappa)) dG_t(y) + \frac{1}{\lambda_2} g_t(p_{2,t}(v_t^b, \omega_t) + \kappa) \frac{\partial p_{2,t}(v_t^b, \omega_t)}{\partial v_t^b} \\ &+ (\pi_t^{p\epsilon})'(v_t^b - h_t(\omega_t) + \lambda_2 p_{2,t}(v_t^b, \omega_t)) g_t(p_{2,t}(v_t^b, \omega_t) + \kappa) \frac{\partial p_{2,t}(v_t^b, \omega_t)}{\partial v_t^b} \le 0. \end{aligned}$$

Here, the second derivative is negative because (a) the $(\pi_t^{p\epsilon})''$ terms are negative since $\pi_t^{p\epsilon}$ is concave, while $[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))]$ is positive as G_t is increasing, making the first three additive terms are all negative; (b) the sum of the last two additive terms is also negative because $(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2 p_{2,t}(v_t^b,\omega_t))\geq -1/\lambda_2$ and $\partial p_{2,t}(v_t^b,\omega_t)/\partial v_t^b=-1/\lambda_2$. Differentiating with respect to ω_t yields

$$\begin{aligned}\frac{\partial\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial\omega_t}=&-\int_0^{\kappa_t^v(\omega_t)}[(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1 y)-(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))]dG_t(y)\\&+[\lambda_1(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))+1]\kappa_t^{v\prime}(\omega_t)[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))]\\&+h_t'(\omega_t)\bigg[(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))\bar{G}_t(\kappa+\kappa_t^v(\omega_t))+\frac{1}{\lambda_2}\bar{G}_t(p_{2,t}(v_t^b,\omega_t)+\kappa)\\&-\int_{\kappa+\kappa_t^v(\omega_t)}^{p_{2,t}(v_t^b,\omega_t)+\kappa}(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2(y-\kappa))dG_t(y)\bigg].\end{aligned}\tag{50}$$

Denote the term in square brackets in the last two rows above (the one that multiplies $h_t'(\omega_t)$) by Υ and notice that $\Upsilon\geq(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))\bar{G}_t(\kappa+\kappa_t^v(\omega_t))-(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2\kappa_t^v(\omega_t))[\bar{G}_t(\kappa+\kappa_t^v(\omega_t))-\bar{G}_t(\kappa+p_{2,t}(v_t^b,\omega_t))]+\bar{G}_t(\kappa+p_{2,t}(v_t^b,\omega_t))/\lambda_2\geq[(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2\kappa_t^v(\omega_t))+1/\lambda_2]\bar{G}_t(\kappa+p_{2,t}(v_t^b,\omega_t))\geq 0$; this will be useful to establish the sign of the second derivative. Also, denote $M_t(z):=(\bar{G}_t(z)-\bar{G}_t(z+\kappa))/((\lambda_1-\lambda_2)\bar{G}_t(z+\kappa))$, which will be a recurring expression in equations that follow. Note that our assumption that $G_t(\cdot)$ has a decreasing hazard rate implies that $M_t(\cdot)$ is decreasing. Taking second-order partial derivatives of $\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)$ yields

$$\begin{aligned}\frac{\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial\omega_t^2}=&\int_0^{\kappa_t^v(\omega_t)}(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1 y)dG_t(y)+[\lambda_1(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))+1]\frac{\partial}{\partial\omega_t}M_t(\kappa_t^v(\omega_t))\\&+h_t''(\omega_t)\Upsilon+(h_t'(\omega_t))^2\int_{\kappa+\kappa_t^v(\omega_t)}^{p_{2,t}(v_t^b,\omega_t)+\kappa}(\pi_t^{p\epsilon})''(v_t^b-h_t(\omega_t)+\lambda_2(y-\kappa))dG_t(y)\\&+(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))(-1+\lambda_1\kappa_t^v(\omega_t))^2[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))]<0,\\\frac{\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial v_b^t\partial\omega_t}=&-\int_0^{\kappa_t^v(\omega_t)}(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1 y)dG_t(y)\\&+(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))](\lambda_1\kappa_t^{v\prime}(\omega_t)-1)\\&+h_t'(\omega_t)\int_{\kappa+\kappa_t^v(\omega_t)}^{p_{2,t}(v_t^b,\omega_t)+\kappa}(\pi_t^{p\epsilon})''(v_t^b-h_t(\omega_t)+\lambda_2(y-\kappa))dG_t(y).\end{aligned}$$

The $\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)/\partial\omega_t^2<0$ result also guarantees uniqueness of $\omega_t^*(v_t^b)$. We will use these derivatives to establish joint concavity by examining the Hessian matrix. For notational brevity, we introduce notation for the recurring expressions:

$$\begin{aligned}K_1:=&\int_0^{\kappa_t^v(\omega_t)}(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1 y)dG_t(y),\\K_2:=&[\lambda_1(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))+1]\frac{\partial}{\partial\omega_t}M_t(\kappa_t^v(\omega_t)),\\K_3:=&\int_{\kappa+\kappa_t^v(\omega_t)}^{p_{2,t}(v_t^b,\omega_t)+\kappa}(\pi_t^{p\epsilon})''(v_t^b-h_t(\omega_t)+\lambda_2(y-\kappa))dG_t(y),\\K_4:=&h_t''(\omega_t)\Upsilon,\\K_5:=&(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))],\\K_6:=&\left(\frac{1}{\lambda_2}+(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2 p_{2,t}(v_t^b,\omega_t))\right)g_t(p_{2,t}(v_t^b,\omega_t)+\kappa)\frac{\partial p_{2,t}(v_t^b,\omega_t)}{\partial v_t^b}.\end{aligned}\tag{51}$$

Note that $K_i \le 0, \forall i \in \{1,2,\dots,6\}$. Using this notation, we can write the second partial derivatives of $\tilde{\pi}_t^{b\epsilon}$ as

$$\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial \omega_t^2} = K_1 + K_2 + h_t'(\omega_t)^2 K_3 + K_4 + K_5(\lambda_1 \kappa_t^{v\prime}(\omega_t) - 1)^2,$$

$$\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial (v_t^b)^2} = K_1 + K_3 + K_5 + K_6,$$

$$\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial v_t^b \partial \omega_t} = -K_1 + K_5(\lambda_1 \kappa_t^{v\prime}(\omega_t) - 1) - h_t'(\omega_t) K_3.$$

To show that the Hessian of $\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)$ is negative-semidefinite, we need that all the first-order principal minors are weakly negative (so, $\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial \omega_t^2} \le 0$ and $\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial (v_t^b)^2} \le 0$, which is already shown above), while the determinant of the Hessian is weakly positive. Calculating the determinant yields

$$\begin{aligned}\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial \omega_t^2}\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial (v_t^b)^2} - \left(\frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial v_t^b \partial \omega_t}\right)^2 =& (K_2+K_4)(K_1+K_3+K_5) + K_1K_3(1-h_t'(\omega_t))^2 \\ &+ K_1K_5(\lambda_1(\kappa_t^v)'(\omega_t))^2 + K_3K_5(h_t'(\omega_t) - 1 + \lambda_1(\kappa_t^v)'(\omega_t))^2 \\ &+ K_6 \frac{\partial^2 \tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial \omega_t^2} \ge 0.\end{aligned}$$

Here, the last inequality holds since all of the additive terms are weakly positive. Thus, the Hessian is negative-semidefinite and hence $\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)$ is jointly concave. From joint concavity of $\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)$ and convexity of $\Re$ it follows from (Boyd and Vandenberghe 2004, pp. 87–88) that $\pi_t^{b*}(v_t^b) = \max_{\omega_t \in [\underline{\omega}_t, \bar{\omega}_t]} \tilde{\pi}_t^{b\epsilon}$ is concave on $v_t^b \in [\underline{v}_t^b, \hat{v}_t^b] = \text{proj}_1 \Re$ (the projection of $\Re$ on its v_t^b component).

Proceeding to step two: establishing concavity of $\pi_t^{b*}(v_t^b)$ on $v_t^b \in [\hat{v}_t^b, \overline{v}_t^b]$. Since $\omega_t = \underline{\omega}_t$ is optimal in this region (by part (i) of this proof), and $\kappa_t(\underline{\omega}_t) = 0$, we have that on $v_t^b \in [\hat{v}_t^b, \overline{v}_t^b]$:

$$\begin{aligned}\pi_t^{b*}(v_t^b) = \tilde{\pi}_t^{b\epsilon}(v_t^b, \underline{\omega}_t) = &\int_0^\kappa \pi_t^{p\epsilon}(v_t^b - \underline{\omega}_t)\, dG_t(y) \\ &+ \int_\kappa^{\hat{p}_{2,t}(v_t^b - \underline{\omega}_t, 0)+\kappa} \left[\pi_t^{p\epsilon}(v_t^b - \underline{\omega}_t + \lambda_2(y-\kappa)) + y - \kappa\right] dG_t(y) \\ &+ \int_{\hat{p}_{2,t}(v_t^b - \underline{\omega}_t, 0)+\kappa}^\infty \left[\pi_t^{p\epsilon}(v_t^b - \underline{\omega}_t + \lambda_2 \hat{p}_{2,t}(v_t^b - \underline{\omega}_t, 0)) + \hat{p}_{2,t}(v_t^b - \underline{\omega}_t, 0)\right] dG_t(y).\end{aligned}$$

From (25) we have $\hat{p}_{2,t}(v_t^b - \underline{\omega}_t, 0) = \max\{(\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2, 0\}$; therefore there exists $\hat{\hat{v}}_t^b \in [\hat{v}_t^b, \overline{v}_t^b]$ such that $\hat{p}_{2,t}(v_t^b - \underline{\omega}_t, 0) = (\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2$ if $v_t^b \in [\hat{v}_t^b, \hat{\hat{v}}_t^b]$ and $\hat{p}_{2,t}(v_t^b - \underline{\omega}_t, 0) = 0$ if $v_t^b \in [\hat{\hat{v}}_t^b, \overline{v}_t^b]$. Thus, on $v_t^b \in (\hat{v}_t^b, \hat{\hat{v}}_t^b)$,

$$\begin{aligned}\frac{d\pi_t^{b*}(v_t^b)}{dv_t^b} =& (\pi_t^{p\epsilon})'(v_t^b - \underline{\omega}_t) G_t(\kappa) + \int_\kappa^{(\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa} (\pi_t^{p\epsilon})'(v_t^b - \underline{\omega}_t + \lambda_2(y-\kappa))\, dG_t(y) \\ &- \frac{1}{\lambda_2}\bar{G}_t((\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa),\end{aligned} \tag{52}$$

$$\begin{aligned}\frac{d^2\pi_t^{b*}(v_t^b)}{d(v_t^b)^2} =& (\pi_t^{p\epsilon})''(v_t^b - \underline{\omega}_t) G_t(\kappa) + \int_\kappa^{(\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa} (\pi_t^{p\epsilon})''(v_t^b - \underline{\omega}_t + \lambda_2(y-\kappa))\, dG_t(y) \\ &- \frac{1}{\lambda_2}(\pi_t^{p\epsilon})'(\Lambda_{2,t}) g_t((\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa) \\ &- \frac{1}{\lambda_2}\bar{G}_t((\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa) \le 0.\end{aligned}$$

Here, the second derivative is weakly negative because all the $(\pi_t^{p\epsilon})''$ terms are weakly negative (from concavity of $(\pi_t^{p\epsilon})''$), while $(\pi_t^{p\epsilon})'(\Lambda_{2,t})$ is positive by definition of $\Lambda_{2,t}$. Hence, $\pi_t^{b*}(v_t^b)$ is concave on $[\hat{v}_t^b, \hat{\hat{v}}_t^b]$. We can do the same check on $[\hat{v}_t^b, \overline{v}_t^b]$, which yields $\pi_t^{b*}(v_t^b) = \pi_t^{p\epsilon}(v_t^b - \underline{\omega}_t)$, which is concave as $\pi_t^{p\epsilon}(\cdot)$ is concave. Furthermore, concavity of $\pi_t^{b*}(v_t^b)$ on $[\hat{v}_t^b, \overline{v}_t^b]$ follows from

$$\begin{aligned}\left.\frac{d\pi_t^{b*}(v_t^b)}{d^- v_t^b}\right|_{v_t^b=\hat{v}_t^b} =& \lim_{v_t^b \to \hat{v}_t^{b-}} \Big[(\pi_t^{p\epsilon})'(v_t^b - \underline{\omega}_t)G_t(\kappa) \\ &+ \int_\kappa^{(\Lambda_{2,t}-v_t^b+\underline{\omega}_t)/\lambda_2+\kappa} (\pi_t^{p\epsilon})'(v_t^b - \underline{\omega}_t + \lambda_2(y-\kappa))\, dG_t(y) - \frac{1}{\lambda_2}\bar{G}_t((\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa)\Big] \\ =& (\pi_t^{p\epsilon})'(\hat{v}_t^b - \underline{\omega}_t)G_t(\kappa) - \frac{1}{\lambda_2}\bar{G}_t(\kappa) \\ \geq& (\pi_t^{p\epsilon})'(\hat{v}_t^b - \underline{\omega}_t) \\ =& \left.\frac{d\pi_t^{b*}(v_t^b)}{d^- v_t^b}\right|_{v_t^b=\hat{v}_t^b}.\end{aligned}$$

Finally, only the third step remains, where we need to ensure that the possible kink at $\hat{v}_t^b$ does not break concavity. Since $\pi_t^{b*}(v_t^b)$ is concave on $[\underline{v}_t^b, \hat{v}_t^b]$ (as shown in the first step of this proof), it is left-differentiable on $(\underline{v}_t^b, \hat{v}_t^b]$ and right-differentiable on $[\underline{v}_t^b, \hat{v}_t^b)$ with $(\pi_t^{b*})'^-(v_t^b) \geq (\pi_t^{b*})'^+(v_t^b)$. Thus, using the Envelope Theorem for one-sided derivatives (Milgrom and Segal 2002, Theorem 1) and recalling that the optimal ω_t at $\hat{v}_t^b$ is $\underline{\omega}_t$, from (49) we have that

$$\left.\frac{d\pi_t^{b*}(v_t^b)}{d^- v_t^b}\right|_{v_t^b=\hat{v}_t^b} \geq (\pi_t^{p\epsilon})'(\hat{v}_t^b - \underline{\omega}_t)G_t(\kappa) + \int_\kappa^{p_{2,t}(\hat{v}_t^b, \underline{\omega}_t)+\kappa} (\pi_t^{p\epsilon})'(\hat{v}_t^b - \underline{\omega}_t + \lambda_2(y-\kappa))dG_t - \frac{1}{\lambda_2}\bar{G}_t(p_{2,t}(\hat{v}_t^b, \underline{\omega}_t) + \kappa),$$

while from (52) it follows that

$$\left.\frac{d\pi_t^{b*}(v_t^b)}{d^+ v_t^b}\right|_{v_t^b=\hat{v}_t^b} = (\pi_t^{p\epsilon})'(\hat{v}_t^b - \underline{\omega}_t)G_t(\kappa) + \int_\kappa^{p_{2,t}(\hat{v}_t^b, \underline{\omega}_t)+\kappa} (\pi_t^{p\epsilon})'(\hat{v}_t^b - \underline{\omega}_t + \lambda_2(y-\kappa))dG_t - \frac{1}{\lambda_2}\bar{G}_t(p_{2,t}(\hat{v}_t^b, \underline{\omega}_t) + \kappa);$$

thus $(\pi_t^{b*})'^-(\hat{v}_t^b) \geq (\pi_t^{b*})'^+(\hat{v}_t^b)$ and so $\pi_t^{b*}(\cdot)$ is concave on its entire domain. □

Proof of Proposition 5. *Part (i).* We will show that $\alpha_t^*(v_t^b)$ is increasing by demonstrating that it is absolutely continuous (thus also almost-everywhere differentiable), with a positive derivative wherever it exists.

First, we establish absolute continuity. From the proof of Proposition 4, part (ii), we have that $\tilde{\pi}^{b\epsilon}(v_t^b, \omega_t) : [\underline{v}_t^b, \bar{v}_t^b] \times [\underline{\omega}_t, \bar{\omega}_t] \to \mathbb{R}$ is a smooth strictly concave function. We define its extension in the ω_t direction $\tilde{\pi}_t^{b\epsilon X}(v_t^p, \omega_t) : [\underline{v}_t^b, \bar{v}_t^b] \times \mathbb{R} \to \mathbb{R}$ by

$$\tilde{\pi}_t^{b\epsilon X}(v_t^b, \omega_t) := \begin{cases} \tilde{\pi}_t^{b\epsilon}(v_t^b, \underline{\omega}_t) - (\underline{\omega}_t - \omega_t)^2 - (\underline{\omega}_t - \omega_t)\frac{\partial \tilde{\pi}_t^{b\epsilon}(v_t^b, \underline{\omega}_t)}{\partial^+ \omega_t} & \text{if } \omega_t \in (-\infty, \underline{\omega}_t), \\ \tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t) & \text{if } \omega_t \in [\underline{\omega}_t, \bar{\omega}_t], \\ \tilde{\pi}_t^{b\epsilon}(v_t^b, \bar{\omega}_t) - (\omega_t - \bar{\omega}_t)^2 + (\omega_t - \bar{\omega}_t)\frac{\partial \tilde{\pi}_t^{b\epsilon}(v_t^b, \bar{\omega}_t)}{\partial^- \omega_t} & \text{if } \omega_t \in (\bar{\omega}_t, \infty). \end{cases}$$

Thus defined $\tilde{\pi}_t^{p\epsilon X}$ is strictly concave in ω_t (although not necessarily jointly concave), smooth, and ensures existence of $\max_{\omega_t \in \mathbb{R}} \tilde{\pi}_t^{b\epsilon X}(v_t^b, \omega_t)$. Let $\omega_t^\dagger(v_t^b)$ be given implicitly by $\partial \tilde{\pi}_t^{b\epsilon X}(v_t^b, \omega_t)/\partial \omega_t = 0$; $\omega_t^\dagger(v_t^b)$ can be thought of as the unconstrained version of $\omega_t^*(v_t^b)$ (while ω_t^* solves $\max_{\omega_t \in [\underline{\omega}_t, \bar{\omega}_t]} \tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t) = \max_{\omega_t \in [\underline{\omega}_t, \bar{\omega}_t]} \tilde{\pi}_t^{b\epsilon X}(v_t^b, \omega_t)$, $\omega_t^\dagger$ solves $\max_{\omega_t \in \mathbb{R}} \tilde{\pi}_t^{b\epsilon X}(v_t^b, \omega_t)$).

By the Implicit Function Theorem, $\omega_t^\dagger(v_t^b)$ is continuously differentiable, thus also Lipschitz continuous, so there exists $K \in \mathbb{R}$ such that $|\omega_t^\dagger(v_1) - \omega_t^\dagger(v_2)| \leq K|v_1 - v_2|, \forall v_1, v_2 \in [\underline{v}_t^b, \bar{v}_t^b]$. Since $\omega_t^*(v_t^b) = \min\{v_t^b - \underline{v}_t^b, \max\{\underline{\omega}_t, \omega_t^\dagger(v_t^b)\}\}$, we have $|\omega_t^*(v_1) - \omega_t^*(v_2)| \leq |\omega_t^\dagger(v_1) - \omega_t^\dagger(v_2)| \leq K|v_1 - v_2|, \forall v_1, v_2 \in [\underline{v}_t^b, \bar{v}_t^b]$, so $\omega_t^*(v_t^b)$ is also Lipschitz continuous. From $\alpha_t^*(v_t^b) = v_t^b - \omega_t^*(v_t^b)$ it then follows that $\alpha_t^*(v_t^b)$ is Lipschitz continuous, thus also absolutely continuous, which is what we wanted to show.

Now we need to find the derivatives of $\alpha_t^*(v_t^b)$ (where they exist). Since $\tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t)$ is jointly concave (established in Part (ii) of the proof of Proposition 4), whenever $(v_t^b, \omega_t^*(v_t^b))$ is in the interior of $\Re$, $\omega_t^*(v_t^b)$ is the solution of $\partial\tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t)/\partial\omega_t = 0$. Thus, at any such point, we can differentiate $\omega_t^*(v_t^b)$ implicitly, which yields

$$\begin{aligned}\frac{d\omega_t^*(v_t^b)}{dv_t^b} &= -\left(\frac{\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial\omega_t\partial v_t^b}\right)\Big/\left(\frac{\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial\omega_t^2}\right)\\ &= -\frac{-K_1 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t) - 1) - h_t'(\omega_t)K_3}{K_1 + K_2 + h_t'(\omega_t)^2K_3 + K_4 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t) - 1)^2},\end{aligned}$$

where all K_j-s are given by (51). Since $\alpha_t^*(v_t^b) = v_t^b - \omega_t^*(v_t^b)$, at those points we also have

$$\begin{aligned}\frac{d\alpha_t^*(v_t^b)}{dv_t^b} &= 1 - \frac{d\omega_t^*(v_t^b)}{dv_t^b}\\ &= \frac{K_2 + h_t'(\omega_t)^2K_3 + K_4 + 2K_5(\lambda_1\kappa_t^{v\prime}(\omega_t) - 1) - h_t'(\omega_t)K_3}{K_1 + K_2 + h_t'(\omega_t)^2K_3 + K_4 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t) - 1)^2}.\end{aligned}$$

Since all K_j-s are weakly negative, $h_t'(\omega_t)$ is positive (see the proof of Proposition 4, part (ii)), and $\lambda_1\kappa_t^{v\prime}(\omega_t) - 1 = \lambda_1/[(\lambda_1 - \lambda_2)\bar{G}_t(\kappa_t^v(\omega_t) + \kappa)] - 1 = [\lambda_1 G_t(\kappa_t^v(\omega_t) + \kappa) + \lambda_2\bar{G}_t(\kappa_t^v(\omega_t) + \kappa)]/[(\lambda_1 - \lambda_2)\bar{G}_t(\kappa_t^v(\omega_t) + \kappa)] > 0$, we have that $\frac{d\alpha_t^*(v_t^b)}{dv_t^b} \geq 0$.

There are two more corner possibilities to check: First, it is possible that $\omega_t^*(v_t^b) = \underline{\omega}_t$. This is always the case for $v_t^b \in [\hat{v}_t^b, \overline{v}_t^b]$, but it may also be true for some $v_t^b \in [\underline{v}_t^b, \hat{v}_t^b)$. For all such points where $\omega_t^*(v_t^b)$ is differentiable, we have $(\omega_t^*)'(v_t^b) = 0$ thus also $\frac{d\alpha_t^*(v_t^b)}{dv_t^b} = 1$. Second, for $v_t^b \in [\underline{v}_t^b, \hat{v}_t^b]$ a corner solution of $\omega_t^*(v_t^b) = v_t^b - \underline{v}_t^b$ is possible. Once again, for any such point where $\omega_t^*(v_t^b)$ is differentiable, we have $(\omega_t^*)'(v_t^b) = 1$ thus also $\frac{d\alpha_t^*(v_t^b)}{dv_t^b} = 0$. (Note that for any v_t^b such that $(v_t^b, \omega_t^*(v_t^b))$ is in the part of the closure of $\Re$ that satisfies $\omega_t^*(v_t^b) - \lambda_1\kappa_t^v(\omega_t^*(v_t^b)) = v_t^b - \Lambda_{1,t}$, $\omega_t^*(v_t^b)$ also solves $\partial\tilde{\pi}_t^{b\epsilon}(v_t^b, \omega_t)/\partial\omega_t = 0$, thus that case will not need to be checked separately.) Putting all the cases together we have that $\frac{d\alpha_t^*(v_t^b)}{dv_t^b} \geq 0$ at all points where $\alpha_t^*(v_t^b)$ is differentiable, which together absolute continuity, established earlier in the proof, implies that $\alpha_t^*(v_t^b)$ is increasing.

Part (ii). From the shape of $\Re$ in Proposition 4 we have that that $\omega_t^*(v_t^b)$ exhibits a $\cap$-pattern where $\omega_t^*(v_t^b) = \underline{\omega}_t, \forall v_t^b \in \{\underline{v}_t^b\} \cup [\hat{v}_t^b, \overline{v}_t^b]$. The statement of Part (ii) then follows from $\kappa_t^v(\omega_t)$ being a non-decreasing function with $\kappa_t^v(\underline{\omega}_t) = 0$.

Part (iii). As shown in part (i) of the proof of Proposition 4, for $v_t^b \in [\underline{v}_t^b, \hat{v}_t^b]$ we have $k_t^{v*}(v_t^b) = \hat{p}_{1,t}^*(v_t^b)$ while (24) implies $\hat{p}_{1,t}^*(v_t^b) \leq \hat{p}_{0,t}^*(v_t^b)$. On the other hand, for $v_t^b \in [\hat{v}_t^b, \overline{v}_t^b]$, from Proposition 4 we obtain $\kappa_t^*(v_t^b) = \kappa_t^*(\omega_t^*(v_t^b)) = \kappa_t^*(\underline{\omega}_t) = 0$. So, on entire $v_t^b \in [\underline{v}_t^b, \overline{v}_t^b]$, we have $\min\{\kappa_t^{v*}(v_t^b), \hat{p}_{0,t}^*(v_t^b)\} = \kappa_t^{v*}(v_t^b)$.

Since $\hat{p}_{0,t}(\alpha_t, \kappa_t^v)$ is given as the solution of $\lambda_1 p + (\lambda_2 - \lambda_1)[p - \kappa_t^v]^+ - \overline{v}_t^p + \alpha_t = 0$, by Implicit Function Theorem, the partial derivatives of $\hat{p}_{0,t}(\alpha_t, \kappa_t^v)$ are given by

$$\frac{\partial\hat{p}_{0,t}(\alpha_t,\kappa_t^v)}{\partial\alpha_t} = \begin{cases}-1/\lambda_1 & \text{if } \hat{p}_{0,t}(\alpha_t,\kappa_t^v) \leq \kappa_t^v,\\ -1/\lambda_2 & \text{if } \hat{p}_{0,t}(\alpha_t,\kappa_t^v) > \kappa_t^v,\end{cases}$$

$$\frac{\partial\hat{p}_{0,t}(\alpha_t,\kappa_t^v)}{\partial\kappa_t^v} = \begin{cases}0 & \text{if } \hat{p}_{0,t}(\alpha_t,\kappa_t^v) < \kappa_t^v,\\ (\lambda_2 - \lambda_1)/\lambda_2 & \text{if } \hat{p}_{0,t}(\alpha_t,\kappa_t^v) > \kappa_t^v.\end{cases}$$

As $\kappa_t^{v*}(v_t^b) \le \hat{p}_{0,t}^*(v_t^b)$, the slope of $\hat{p}_{0,t}(\alpha_t^*(v_t^b), \kappa_t^{v*}(v_t^b))$ does not depend on $\kappa_t^{v*}(v_t^b)$; and as $\alpha_t^*(v_t^b)$ is increasing in v_t^b, $\hat{p}_{0,t}(\alpha_t^*(v_t^b), \kappa_t^{v*}(v_t^b))$ is decreasing in v_t^b.

Part (iv). $\hat{p}_{1,t}^*(v_t^b) = \kappa_t^{v*}(v_t^b)$ is established in the proof of Proposition 4, part (i), after which $\hat{p}_{1,t}^*(v_t^b)$ exhibiting a ∩–pattern in v_t^b follows from part (ii) of this proposition.

Part (v). From the proof of Proposition 4, part (ii), we have that on $[\underline{v}_t^b, \hat{v}_t^b]$: $\hat{p}_{2,t}^*(v_t^b) = (\Lambda_{2,t} - \alpha_t^*(v_t^b) - (\lambda_1 - \lambda_2)\kappa_t^{v*}(v_t^b))/\lambda_2 = (\Lambda_{2,t} - v_t^b + \omega_t^*(v_t^b) - (\lambda_1 - \lambda_2)\kappa_t^{v*}(v_t^b))/\lambda_2 = (\Lambda_{2,t} - v_t^b + h_t(\omega_t^*(v_t^b)))/\lambda_2$, where $h_t(\omega_t) := \omega_t - (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t)$ (as in the proof of Proposition 4). Thus, for all $v_t^b \in [\underline{v}_t^b, \hat{v}_t^b]$ for which $d\omega_t^*(v_t^b)/dv_t^b$ exists, $d\hat{p}_{2,t}^*(v_t^b)/dv_t^b$ also exists and is given by

$$\begin{aligned}
\frac{d\hat{p}_{2,t}^*(v_t^b)}{dv_t^b} =& \frac{1}{\lambda_2}\left(-1 + h_t'(\omega_t^*(v_t^b))\frac{d\omega_t^*(v_t^b)}{dv_t^b}\right) \\
=& \frac{1}{\lambda_2}\left(-1 - h_t'(\omega_t^*(v_t^b))\frac{-K_1 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1) - h_t'(\omega_t^*(v_t^b))K_3}{K_1 + K_2 + h_t'(\omega_t^*(v_t^b))^2K_3 + K_4 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1)^2}\right) \\
=& -\frac{1}{\lambda_2}\left(1 + \frac{-K_1h_t'(\omega_t^*(v_t^b)) + K_5h_t'(\omega_t^*(v_t^b))(\lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1) - (h_t'(\omega_t^*(v_t^b)))^2K_3}{K_1 + K_2 + h_t'(\omega_t^*(v_t^b))^2K_3 + K_4 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1)^2}\right) \\
=& -\frac{1}{\lambda_2}\left(\frac{K_1(1 - h_t'(\omega_t^*(v_t^b))) + K_2 + K_4 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1)(h_t'(\omega_t^*(v_t^b)) + \lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1)}{K_1 + K_2 + (h_t'(\omega_t^*(v_t^b)))^2K_3 + K_4 + K_5(\lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1)^2}\right).
\end{aligned}$$

Here, the signs of all terms that multiply K_i-s are clearly positive except $(1 - h_t'(\omega_t^*(v_t^b)) = (\lambda_1 - \lambda_2)\kappa_t^{v\prime}(\omega_t) > 0$ and the sign of $h_t'(\omega_t^*(v_t^b)) + \lambda_1\kappa_t^{v\prime}(\omega_t^*(v_t^b)) - 1$, which is unclear. Expanding that term yields $-1 + \lambda_1/[(\lambda_1 - \lambda_2)\bar{G}_t(\kappa + \kappa_t^v(\omega_t^*(v_t^b)))] - G_t(\kappa + \kappa_t^v(\omega_t^*(v_t^b)))/\bar{G}_t(\kappa + \kappa_t^v(\omega_t^*(v_t^b))) = \lambda_2/[(\lambda_1 - \lambda_2)\bar{G}_t(\kappa + \kappa_t^{v*}(v_t^b)) > 0$. Thus, as all K_i-s are negative and the terms multiplying them are positive, we have that $d\hat{p}_{2,t}^*(v_t^b)/dv_t^b \le 0$.

From the proof of Proposition 4, part (ii), we also have that there exists $\hat{\hat{v}}_t^b \in [\hat{v}_t^b, \overline{v}_t^b]$ such that $\hat{p}_{2,t}^*(v_t^b) = (\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2$ if $v_t^b \in [\hat{v}_t^b, \hat{\hat{v}}_t^b]$ and $\hat{p}_{2,t}^*(v_t^b) = 0$ if $v_t^b \in [\hat{\hat{v}}_t^b, \overline{v}_t^b]$. In both of those cases $p_{2,t}^*(v_t^b)$ is also (weakly) decreasing. Finally, that $p_2^*, t(v_t^b)$ is absolutely continuous follows from a Lipschitz continuity argument analogous to the one used in part (i) of this proof. Thus, it is decreasing on its entire domain.

Part (vi). For $v_t^b \in [\underline{v}_t^b, \hat{v}_t^b]$, $p_{1,t}^*(v_t^b) = \kappa_t^{v*}(v_t^b)$, so

$$\begin{aligned}
p_{2,t}^*(v_t^b) - p_{1,t}^*(v_t^b) &= (\Lambda_{2,t} - \alpha_t^*(v_t^b) - (\lambda_1 - \lambda_2)\kappa_t^{v*}(v_t^b))/\lambda_2 - \kappa_t^{v*}(v_t^b) \\
&= (\Lambda_{2,t} - \alpha_t^*(v_t^b) - \lambda_1\kappa_t^{v*}(v_t^b))/\lambda_2.
\end{aligned}$$

Here $\alpha_t^*(v_t^b)$ is increasing (by part (i) of this proposition), while $\kappa_t^{v*}(v_t^b)$ is ∩-shaped (by part (ii) of this proposition). Thus there exists $\grave{v}_t^b \in [\underline{v}_t^b, \hat{v}_t^b]$ such that both $\alpha_t^*(v_t^b)$ and $\kappa_t^{v*}(v_t^b)$ are increasing on $v_t^b \in [\underline{v}_t^b, \grave{v}_t^b]$, in which case we also obtain that $p_{2,t}^*(v_t^b)$ is decreasing on $v_t^b \in [\underline{v}_t^b, \grave{v}_t^b]$.

For $v_t^b \in [\hat{v}_t^b, \overline{v}_t^b]$ we have $p_{1,t}^*(v_t^b) = 0$, while $\hat{p}_{2,t}^*(v_t^b) = (\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2$ if $v_t^b \in [\hat{v}_t^b, \hat{\hat{v}}_t^b]$ and $\hat{p}_{2,t}^*(v_t^b) = 0$ if $v_t^b \in [\hat{\hat{v}}_t^b, \overline{v}_t^b]$. In both of those cases $p_{2,t}^*(v_t^b)$ is (weakly) decreasing, thus so is $p_{2,t}^*(v_t^b) - p_{1,t}^*(v_t^b)$. □

Proof of Proposition 6. The result follows directly from noting that (32) is linear, thus also concave, and applying Propositions 2–4. □

Proof of Proposition 7.

Part 1: optimality of $v_1^{b*}(d; \alpha_0, \kappa_0^v)$. First, we note an intuitive property: (35) implies that without loss of optimality we can restrict attention in (34)–(35) to non-decreasing functions d and v_1^b. Consider an arbitrary triple $(d, \hat{v}_1^b, \underline{d}) \in \mathcal{D} \times \mathcal{V}_0 \times \mathbb{R}_+$ that satisfies the IC constraint (35), and denote the consumer's expected utility under this triple by $v^\dagger$. Can the firm benefit from choosing a different reward function v_1^b, while keeping $d, \underline{d}$, and $v^\dagger$ constant? From (34)–(35), this problem can be formulated as

$$\begin{aligned} \max_{v_1^b \in \mathcal{V}_0} \quad & \mathbb{E}\big[d(W) + \gamma \pi_1^b(v_1^b(d(W))) - \mathcal{K} | d(W) \geq \underline{d}\big] \times \Pr\{d(W) \geq \underline{d}\} \\ \text{s.t.} \quad & d(w) \in \operatorname*{argmax}_{0 \leq d < w} \begin{cases} \tilde{u}(w-d) + \delta v_1^b(d) & \text{if } d \geq \underline{d}; \\ \tilde{u}(w) + \delta \underline{v}_1^b & \text{otherwise,} \end{cases} \quad \forall w \in \mathbb{R}_+, \\ & v^\dagger = \mathbb{E}\big[\mathbb{1}_{W \geq \underline{d}}(\tilde{u}(W - d(W)) + \delta v_1^b(d(W))) + \mathbb{1}_{W < \underline{d}}(\tilde{u}(W) + \delta \underline{v}_1^b)\big]. \end{aligned}$$

Denoting by $W_{d,\underline{d}}$ a random variable with distribution equal to the conditional distribution of W, conditional on $d(W) \geq \underline{d}$; defining $v_{d,\underline{d}}^{\dagger\dagger} := (v^\dagger - \mathbb{E}[\mathbb{1}_{W<\underline{d}}(\tilde{u}(W) + \delta \underline{v}_1^b)])/\Pr\{W \geq \underline{d}\}$; and dropping the terms from the objective function that do not depend on v_1^b, we can restate the problem above as

$$\max_{v_1^b \in \mathcal{V}_0} \quad \mathbb{E}\big[d(W_{d,\underline{d}}) + \gamma \pi_1^b(v_1^b(d(W_{d,\underline{d}})))\big] \tag{53}$$

$$\text{s.t.} \quad d(w) \in \operatorname*{argmax}_{\underline{d} \leq d < w} \tilde{u}(w-d) + \delta v_1^b(d), \quad \forall w \in \operatorname{supp} W_{d,\underline{d}}, \tag{54}$$

$$\tilde{u}(w - d(w)) + \delta v_1^b(d(w)) \geq \tilde{u}(w) + \delta \underline{v}_1^b, \quad \forall w \in \operatorname{supp} W_{d,\underline{d}}, \tag{55}$$

$$v_{d,\underline{d}}^{\dagger\dagger} = \mathbb{E}\big[\tilde{u}(W_{d,\underline{d}} - d(W_{d,\underline{d}})) + \delta v_1^b(d(W_{d,\underline{d}}))\big]. \tag{56}$$

Now, consider the closely related problem

$$\max_{v_1^b \in \mathcal{V}_0} \quad \mathbb{E}\big[d(W_{d,\underline{d}}) + \gamma \pi_1^b(v_1^b(d(W_{d,\underline{d}})))\big] \tag{57}$$

$$\text{s.t.} \quad d(w) \in \operatorname*{argmax}_{0 \leq d < w} \tilde{u}(w-d) + \delta v_1^b(d), \quad \forall w \in \operatorname{supp} W_{d,\underline{d}}, \tag{58}$$

$$v_{d,\underline{d}}^{\dagger\dagger} = \mathbb{E}\big[\tilde{u}(W_{d,\underline{d}} - d(W_{d,\underline{d}})) + \delta v_1^b(d(W_{d,\underline{d}}))\big]. \tag{59}$$

This is useful as (57)–(59) is analogous to (19)–(21) that we already solved. Thus, it follows from the proof of Proposition 3, part 1, that there exists $v_1^{b*}(d; \alpha_0, \kappa_0^v)$, as given by (36), such that it satisfies (58)–(59) and performs as well as $\hat{v}_1^b$ or better in (57). It is easily verifiable that v_1^{b*} also satisfies (55) and—because (58) is tighter than (54)—also satisfies (54). Thus, v_1^{b*} also performs as well or better than $\hat{v}_1^b$ in (53)–(56). Consequently, in (34)–(35), we can restrict attention to reward functions that satisfy the functional form in (36) without loss of optimality, i.e., $v_1^b \in \{v_1^{b*}(\cdot; \alpha_0, \kappa_0^v) | \alpha_0, \kappa_0^v \in \mathbb{R}_+\}$.

Part 2: optimality of $d^*(w; \alpha_0, \kappa_0^v, \underline{d})$. We will solve the problem of finding the optimal d, for any given $v_1^b(\cdot; \alpha_0, \kappa_0^v)$ and $\underline{d}$, by pointwise optimization of (34)–(35), and afterword verify that this solution really belongs to $\mathcal{D}$. To do that, we first introduce some notation needed to rewrite (35) in a more convenient form. Denote $\underline{w}(\alpha_0, \kappa_0^v, \underline{d}) = \inf\{w \in [\underline{d}, \infty) | \tilde{u}(w - \underline{d}) + \delta v_1^{b*}(\underline{d}; \alpha_0, \kappa_0^v) \geq \tilde{u}(w) + \delta \underline{v}_1^b\}$—this is the lowest wealth at which the consumer is willing to pay the minimum downpayment. (It is possible that $\underline{w}(\alpha_0, \kappa_0^v, \underline{d}) = \infty$, in which case the consumer is never willing to pay the downpayment, regardless of the realization of W.) Note that without loss of otimality, in (34)–(35) we can restrict attention to triples $(d, v_b^1, \underline{d})$ such that $d(\underline{w}) = \underline{d}$. (To see why this is the case, assume there exists an optimal triple $(d, v_b^1, \underline{d})$ that does not satisfy this property. But then, changing the minimum downpayment to $\underline{d}^{\text{alt}} := \inf\{w \in \mathbb{R}_+ | d(w) = \underline{d}\}$ will still be incentive compatible and achieve the same payout for the firm, implying that $(d, v_1^b, \underline{d}^{\text{alt}})$ is optimal as well.) With this in mind, for

$w \geq \underline{w}(\alpha_0, \kappa_0^v, \underline{d})$, we can write down the upper and lower bounds of incentive compatible payment functions (ones that satisfy (35)) as

$$d^\times(w;\alpha_0,\kappa_0^v,\underline{d}) := \begin{cases} \min\{\underline{d}+w-\underline{w}, \max\{\hat{d}_0(\alpha_0,\kappa_0^v),\underline{d}\}\} & \text{if } w < \underline{w}+[\kappa_0^v-\underline{d}]^+, \\ \min\{\underline{d}+[\kappa_0^v-\underline{d}]^+, \max\{\hat{d}_0(\alpha_0,\kappa_0^v),\underline{d}\}\} & \text{if } \underline{w}+[\kappa_0^v-\underline{d}]^+ \leq w < \underline{d}+\tilde{\kappa}+[\kappa_0^v-\underline{d}]^+, \\ \min\{w-\tilde{\kappa}, \max\{\hat{d}_0(\alpha_0,\kappa_0^v),\underline{d}\}\} & \text{if } \underline{w}+[\kappa_0^v-\underline{d}]^+ \leq w < \hat{d}_0(\alpha_0,\kappa_0^v)+\tilde{\kappa}, \\ \max\{\hat{d}_0(\alpha_0,\kappa_0^v),\underline{d}\} & \text{if } w \geq \hat{d}_0(\alpha_0,\kappa_0^v)+\tilde{\kappa}, \end{cases}$$

$$d^\circ(w;\alpha_0,\kappa_0^v,\underline{d}) := \begin{cases} \underline{d} & \text{if } w < \tilde{\kappa}+\underline{d}, \\ \min\{w-\tilde{\kappa}, \max\{\hat{d}_0(\alpha_0,\kappa_0^v),\underline{d}\}\} & \text{if } \tilde{\kappa}+\underline{d} \leq w < \tilde{\kappa}+\underline{d}+[\kappa_0^v-\underline{d}]^+, \\ \min\{\underline{d}+[\kappa_0^v-\underline{d}]^+, \max\{\hat{d}_0(\alpha_0,\kappa_0^v),\underline{d}\}\} & \text{if } w \geq \tilde{\kappa}+\underline{d}+[\kappa_0^v-\underline{d}]^+. \end{cases}$$

(For $w < \underline{w}(\alpha_0, \kappa_0^v, \underline{d})$, any $d(w) \leq w$ is incentive compatible, but the firm will not accept any such payment.) This allows us to express the pointwise optimization of (34)–(35) for each $w \geq \underline{w}$ by

$$\max_{d(w)} \quad d(w) + \gamma\pi_1^b(v_1^{b*}(d(w);\alpha_0,\kappa_0^v)) - \mathcal{K} \tag{60}$$

$$\text{s.t.} \quad d^\circ(w;\alpha_0,\kappa_0^v,\underline{d}) \leq d(w) \leq d^\times(w;\alpha_0,\kappa_0^v,\underline{d}). \tag{61}$$

From Proposition 4, π_1^b is concave, and thus, using the shape of v_1^{b*} from part 1 of this proposition, the objective function of the problem above is either uni- or bi-modal in $d(w)$, being weakly increasing up to $d_1^{\dagger\dagger}(\alpha_0,\kappa_0^v,\underline{d}) := \min\{\kappa_0^v, [\delta\Lambda_{1,0}-\alpha_0]^+/\tilde{\lambda}_1\}$, after which it is weakly decreasing up to κ_0^v (this region is empty if $[\delta\Lambda_{1,0}-\alpha_0]^+/\tilde{\lambda}_1 > \kappa_0^v$), weakly increasing up to $d_2^{\dagger\dagger}(\alpha_0,\kappa_0^v,\underline{d}) := (\delta\Lambda_{2,0}-\alpha_0-(\tilde{\lambda}_1-\tilde{\lambda}_2)\kappa_0^v)/\tilde{\lambda}_2$, and weakly decreasing afterwards. Here, $\Lambda_{1,0}$ and $\Lambda_{2,0}$ are as given in the statement of the proposition. Given that there is no $w \geq \underline{w}$, such that $d^\circ(w;\alpha_0,\kappa_0^v,\underline{d}) < \kappa_0 < d^\times(w;\alpha_0,\kappa_0^v,\underline{d})$, the maximizer of (60)–(61) is either going to be an interior one where $d^*(w) \in \{d_1^{\dagger\dagger}(\alpha_0,\kappa_0^v,\underline{d}), d_2^{\dagger\dagger}(\alpha_0,\kappa_0^v,\underline{d})\}$, or a corner one where $d^*(w) \in \{d^\circ(w;\alpha_0,\kappa_0^v,\underline{d}), d^\times(w;\alpha_0,\kappa_0^v,\underline{d})\}$. More precisely, for $w \geq \underline{w}$, inserting the expressions for $d^\times$ and d° into (60)–(61), we have that

$$d^*(w;\alpha_0,\kappa_0^v,\underline{d}) = \begin{cases} \underline{d}+w-\underline{w}(\alpha_0,\kappa_0^v,\underline{d}) & \text{if } w \leq \hat{d}_1(\alpha_0,\kappa_0^v,\underline{d})+\underline{w}(\alpha_0,\kappa_0^v,\underline{d})-\underline{d}, \\ \hat{d}_1(\alpha_0,\kappa_0^v,\underline{d}) & \text{if } \hat{d}_1(\alpha_0,\kappa_0^v,\underline{d})+\underline{w}(\alpha_0,\kappa_0^v,\underline{d})-\underline{d} \leq w \leq \hat{d}_1(\alpha_0,\kappa_0^v,\underline{d})+\tilde{\kappa}, \\ w-\tilde{\kappa} & \text{if } \hat{d}_1(\alpha_0,\kappa_0^v,\underline{d})+\tilde{\kappa} \leq w \leq \hat{d}_2(\alpha_0,\kappa_0^v,\underline{d})+\tilde{\kappa}, \\ \hat{d}_2(\alpha_0,\kappa_0^v,\underline{d}) & \text{if } w \geq \hat{d}_2(\alpha_0,\kappa_0^v,\underline{d})+\tilde{\kappa}, \end{cases}$$

where $\hat{d}_1(\alpha_0,\kappa_0^v,\underline{d}) = \max\{\underline{d}, \min\{\underline{w}(\alpha_0,\kappa_0^v,\underline{d}) + [\kappa_0^v - \underline{d}]^+, d_1^{\dagger\dagger}(\alpha_0,\kappa_0^v,\underline{d}), \hat{d}_0(\alpha_0,\kappa_0^v)\}\}$, $\hat{d}_2(\alpha_0,\kappa_0^v,\underline{d}) = \max\{\underline{d}, d_2^{\dagger\dagger}(\alpha_0,\kappa_0^v,\underline{d}), \min\{\kappa_0^v, \hat{d}_0(\alpha_0,\kappa_0^v)\}\}$. Lastly, for completeness, we note that thus defined $d^*(w;\alpha_0,\kappa_0^v,\underline{d})$ is piecewise-linear (thus also left- and right- differentiable) as well as continuous, so $d^*(w;\alpha_0,\kappa_0^v,\underline{d}) \in \mathcal{D}$, completing the proof. □

Proof of Proposition 8. *Part (i).* Denote by $d^\dagger, v_1^{b\dagger}$ the solution of (33), and assume the opposite: there exist $\alpha_0^*, \kappa_0^{v*}, \underline{d}^*$, such that $d^{**}, v_1^{b**}, \underline{d}^*$ solve (34)–(35) and $d^\dagger < \underline{d}^*$. (As introduced in the paragraph after Proposition 7, d^{**}, v_1^{b**} are the functions d^*, v_1^{b*} given by Proposition 7, evaluated the optimal parameters $\alpha_0^*, \kappa_0^{v*}, \underline{d}^*$.) Then, denoting the cdf of W by F, from (34) we have that the expected profit of the firm under $d^{**}, v_1^{b**}, \underline{d}^*$ is

$$(\underline{d}^* + \gamma\pi_1^b(v_1^{b**}(\underline{d}^*)) - \mathcal{K}) \times \Pr\{W \geq \underline{w}(\alpha_0^*, \kappa_0^{v*}, \underline{d}^*)\} \\ + \int_{\inf\{x\in\mathbb{R}_+ | d^{**}(x) > \underline{d}^*\}}^{\infty} \left(\pi_1^b(v_1^{b**}(d^{**}(w))) + d^{**}(w) - \pi_1^b(v_1^{b**}(\underline{d}^*)) - d^{**}(\underline{d}^*)\right) dF(w). \tag{62}$$

Changing the minimum downpayment from $\underline{d}^*$ to $d^\dagger$ while keeping $\alpha_0^*, \kappa_0^{v*}$ the same would change the firm's payoff to

$$\begin{aligned}&(\underline{d}^\dagger + \gamma\pi_1^b(v_1^{b**}(\underline{d}^\dagger)) - \mathcal{K}) \times \Pr\{W \geq \underline{w}(\alpha_0^*, \kappa_0^{v*}, \underline{d}^\dagger)\}\\ &+ \int_{\inf\{x\in\mathbb{R}_+|d^{**}(x)>\underline{d}^*\}}^{\infty} \left(\pi_1^b(v_1^{b**}(d^{**}(w))) - \pi_1^b(v_1^{b**}(\underline{d}^*))\right) dF(w)\\ &+ \int_{\inf\{x\in\mathbb{R}_+|d^*(x;\alpha_0^*,\kappa_0^{v*},\underline{d}^\dagger)>\underline{d}^\dagger\}}^{\infty} \left(\min\{\pi_1^b(v_1^{b**}(d^*(w;\alpha_0^*,\kappa_0^{v*},\underline{d}^\dagger)), \pi_1^b(v_1^{b**}(d(\inf\{x\in\mathbb{R}_+|d^{**}(x)>\underline{d}^*\})))\} - \pi_1^b(v_1^{b**}(\underline{d}^\dagger))\right) dF(w).\end{aligned}$$

Here, the first line is no larger than the first line of (62) (because of the optimality of $d^\dagger, v_1^{b\dagger}$ in (33)), the second line is the same as in (62), while the third line is weakly positive (because $\pi_1^b(v_1^{b**}(\cdot))$ is non-decreasing). Therefore, this change did not decrease the firm's payoff, and then either we have a contradiction to the optimality of $(\alpha_0^*, \kappa_0^{v*}, \underline{d}^*)$ or $(\alpha_0^*, \kappa_0^{v*}, \underline{d}^\dagger)$ achieves the same optimum. Lastly, that the wealth needed to acquire the technology in the second method is also no larger than in the first follows from $\Pr\left\{W \geq \underline{d},\ \tilde{u}(W - \underline{d}) + \delta v_1^b \geq \tilde{u}(W) + \delta \underline{v}_1^b\right\} = 1 - F(\underline{w}(\alpha_0, \kappa_0^v, \underline{d}))$ for any incentive compatible $(\alpha_0, \kappa_0^v, \underline{d})$ such that $v_1^b(\underline{d}; \alpha_0, \kappa_0) = v_1^b$, as well as $\underline{w}(\alpha_0, \kappa_0^v, \underline{d})$ being weakly increasing in $\underline{d}$.

Part (ii). The statement follows directly from (33) being the same optimization problem as (34)–(35), just over a smaller parameter space. More precisely, restricting the firm's choice in (34)–(35) to reward functions that satisfy $v_1^b(d) = v_1^b, \forall d \geq \underline{d}$ retrieves (33). □

Proof of Proposition 9. *Preliminaries.* We first derive some conditions which will be helpful for showing the statement of the proposition. Recall that in periods before last, ownership can only be possible if $(\pi_t^p)'(\overline{v}_t^p) \geq -1/\lambda_2$, or equivalently $(\pi_t^a)'(\overline{v}_t^a) \geq -1/\lambda_2$. We start by determining that slope, which will also be instrumental to characterize the situations where the firm wants to terminate irrespective of the payment.

From $\pi_{t-1}^e(v_{t-1}^e) = \gamma\pi_t^b(v_{t-1}^e/\delta)$ we have $(\pi_{t-1}^e)'(v_{t-1}^e) = \gamma/\delta(\pi_t^b)'(v_{t-1}^e/\delta)$. From the proof of Proposition 4, part (i), if $\hat{v}_t^b < \overline{v}_t^b$ then $\pi_t^b(v_t^b) = \pi_t^p(v_t^b - \underline{\omega}_t)$, so from $(\pi_{t-1}^e)'(\overline{v}_{t-1}^e) = \gamma/\delta(\pi_t^b)'(\overline{v}_{t-1}^e/\delta)$ in this case we have $(\pi_{t-1}^e)'(\overline{v}_{t-1}^e) = \gamma/\delta(\pi_t^a)'(\overline{v}_t^a)$. On the other hand, if $\hat{v}_t^b = \overline{v}_t^b$, by inserting $\omega_t^*(v_t^b)$ into (52) we obtain $(\pi_{t-1}^e)'(\overline{v}_{t-1}^e) = \gamma/\delta\left[(\pi_t^a)'(\overline{v}_t^a)G_t(\kappa) - \bar{G}_t(\kappa)/\lambda_2\right]$.

Applying Proposition 1, the situations where the principal desires to terminate irrespective of payment arise if and only if (9) is satisfied, which here becomes equivalent to $-\mathcal{C}_t/(\overline{v}_t^e - \underline{v}_t^e) \leq (\pi_t^e)'(\overline{v}_t^e)$. Checking whether this condition is satisfied in different time periods will be the core of the proof.

Now, consider a period θ as defined in the statement of this proposition. Then, for $t \leq \theta$ we have

$$\frac{\mathcal{C}_{t-1}}{\mathcal{C}_t} \leq \frac{\gamma}{\delta}\frac{\overline{v}_{t-1}^e - \underline{v}_{t-1}^e}{\overline{v}_t^e - \underline{v}_t^e} \Leftrightarrow -\frac{\mathcal{C}_{t-1}}{\overline{v}_{t-1}^e - \underline{v}_{t-1}^e} \geq -\frac{\gamma}{\delta}\frac{\mathcal{C}_t}{\overline{v}_t^e - \underline{v}_t^e}.$$

If $t > \theta$, then (39) holds, so

$$-\frac{\mathcal{C}_{t-1}}{\overline{v}_{t-1}^e - \underline{v}_{t-1}^e} \leq -\frac{\gamma}{\delta}\frac{\mathcal{C}_t}{\overline{v}_t^e - \underline{v}_t^e} \quad \text{and} \quad -\frac{\mathcal{C}_{t-1}}{\overline{v}_{t-1}^e - \underline{v}_{t-1}^e} \leq -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_t}{\overline{v}_t^e - \underline{v}_t^e}G_t(\kappa) - \frac{1}{\lambda_2}\bar{G}_t(\kappa)\right].$$

Part 1: optimality of terminating when $t \geq \theta$. We will show this by backward induction. From the last period payoff function (32), we have

$$(\pi_T^a)'(v_T^a) \geq -\frac{\mathcal{C}_T}{\overline{v}_T^a - \underline{v}_T^a}, \quad \text{so} \quad (\pi_{T-1}^e)'(v_{T-1}^e) \geq -\frac{\mathcal{C}_{T-1}}{\overline{v}_{T-1}^e - \underline{v}_{T-1}^e},$$

which forms the basis of the induction. Assume there exists $t > \theta$ such that $(\pi_t^e)'(v_t^e) \geq -\frac{\mathcal{C}_t}{\bar{v}_t^e - \underline{v}_t^e}$. Then because, as established in the preliminaries, $(\pi_{t-1}^e)'(\bar{v}_{t-1}^e)$ is equal to either $\gamma/\delta(\pi_t^a)'(\bar{v}_t^a)$ or $\gamma/\delta\left[(\pi_t^a)'(\bar{v}_t^a)G_t(\kappa) - \bar{G}_t(\kappa)/\lambda_2\right]$, both of which are greater than $-\frac{\mathcal{C}_t}{\bar{v}_t^e - \underline{v}_t^e}$, we have that $(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) \geq -\frac{\mathcal{C}_{t-1}}{\bar{v}_{t-1}^e - \underline{v}_{t-1}^e}$ as well, completing the induction.

Part 2: optimality of not terminating when $t < \theta$. Once again, we will show this by backward induction. First, we show that it will be optimal not to terminate at time $\theta - 1$, which will form the basis of the induction. From part 1 of this proposition, it is optimal to terminate at θ so $(\pi_\theta^e)'(\bar{v}_\theta^e) \geq -\frac{\mathcal{C}_\theta}{\bar{v}_\theta^e - \underline{v}_\theta^e}$. Thus, using expressions from the preliminaries we obtain that, if $-\mathcal{C}_\theta/(\bar{v}_\theta^e - \underline{v}_\theta^e) < -1/\lambda_2$,

$$(\pi_{\theta-1}^e)'(\bar{v}_{\theta-1}^e) = -\frac{\gamma}{\delta}\frac{\mathcal{C}_\theta}{\bar{v}_\theta^e - \underline{v}_\theta^e} \leq -\frac{\mathcal{C}_{\theta-1}}{\bar{v}_{\theta-1}^e - \underline{v}_{\theta-1}^e}.$$

If $-\mathcal{C}_\theta/(\bar{v}_\theta^e - \underline{v}_\theta^e) \geq -1/\lambda_2$, we have

$$\begin{aligned}(\pi_{\theta-1}^e)'(\bar{v}_{\theta-1}^e) &= -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_\theta}{\bar{v}_\theta^e - \underline{v}_\theta^e}G_\theta(\kappa) + \frac{1}{\lambda_2}\bar{G}_\theta(\kappa)\right] \\ &\leq -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_\theta}{\bar{v}_\theta^e - \underline{v}_\theta^e}G_\theta(\kappa) + \frac{\mathcal{C}_\theta}{\bar{v}_\theta^e - \underline{v}_\theta^e}\bar{G}_\theta(\kappa)\right] \\ &= -\frac{\gamma}{\delta}\frac{\mathcal{C}_\theta}{\bar{v}_\theta^e - \underline{v}_\theta^e} \leq -\frac{\mathcal{C}_{\theta-1}}{\bar{v}_{\theta-1}^e - \underline{v}_{\theta-1}^e}.\end{aligned}$$

Combining the two exhaustive cases, we have $(\pi_{\theta-1}^e)'(\bar{v}_{\theta-1}^e) \leq -\frac{\mathcal{C}_{\theta-1}}{\bar{v}_{\theta-1}^e - \underline{v}_{\theta-1}^e}$, so it is optimal not to terminate in period $\theta - 1$. Assume there exists period $t \leq \theta$ such that $(\pi_t^e)'(\bar{v}_t^e) \leq -\frac{\mathcal{C}_t}{\bar{v}_t^e - \underline{v}_t^e}$. Then, if $-\mathcal{C}_t/(\bar{v}_t^e - \underline{v}_t^e) < -1/\lambda_2$,

$$(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) = \frac{\gamma}{\delta}(v_{t-1}^e)'(\bar{v}_{t-1}^e) \leq -\frac{\mathcal{C}_{t-1}}{\bar{v}_{t-1}^e - \underline{v}_{t-1}^e}.$$

If $-\mathcal{C}_t/(\bar{v}_t^e - \underline{v}_t^e) \geq -1/\lambda_2$, we have

$$\begin{aligned}(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) &= -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_t}{\bar{v}_t^e - \underline{v}_t^e}G_t(\kappa) + \frac{1}{\lambda_2}\bar{G}_t(\kappa)\right] \\ &\leq -\frac{\gamma}{\delta}\frac{\mathcal{C}_t}{\bar{v}_t^e - \underline{v}_t^e} \leq -\frac{\mathcal{C}_{t-1}}{\bar{v}_{t-1}^e - \underline{v}_{t-1}^e}.\end{aligned}$$

Thus, $(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) \leq -\frac{\mathcal{C}_{t-1}}{\bar{v}_{t-1}^e - \underline{v}_{t-1}^e}$ in all cases, completing the induction. □

Proof of Proposition 10. From the proof of Proposition 9 we have that for all $1 \leq t < \theta : (\pi_t^e)'(\bar{v}_t^e) \leq -\mathcal{C}_t/(\bar{v}_t^e - \underline{v}_t^e)$ and

$$(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) = \begin{cases} \frac{\gamma}{\delta}(\pi_t^e)'(\bar{v}_t^e) & \text{if } (\pi_t^e)'(\bar{v}_t^e) < -\frac{1}{\lambda_2}, \\ -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_t}{\bar{v}_t^e - \underline{v}_t^e}G_t(\kappa) + \frac{1}{\lambda_2}\bar{G}_t(\kappa)\right] & \text{if } (\pi_t^e)'(\bar{v}_t^e) \geq -\frac{1}{\lambda_2}. \end{cases} \tag{63}$$

Note that this expression is strictly negative. Since in any period t, $(\pi_t^e)'(\bar{v}_t^e) < -1/\lambda_2$ implies that no ownership can be attained in period t, while also implying that $(\pi_{\hat{t}}^e)'(\bar{v}_{\hat{t}}^e) < -1/\lambda_2$ for all $\hat{t} < t$ (using (63)), it follows that $(\pi_t^e)'(\bar{v}_t^e) < -1/\lambda_2$ also implies that ownership cannot be attained in any period up to t. Thus for a model to become a rent-only model, it is sufficient that $(\pi_{\theta^*-1}^e)'(\bar{v}_{\theta^*-1}^e) < -1/\lambda_2$ and that direct sales do not happen at contract initiation (period 0). From the proof of Proposition 7 we have that a no-direct-sales model is optimal if $(\pi_1^b)'(\bar{v}_1^b) \leq -\delta/(\gamma\tilde{\lambda}_2)$.

Thus, if $\tilde{\lambda}_2 > \delta/(\gamma\lambda_2)$ then $(\pi_{\theta^*-1}^e)'(\bar{v}_{\theta^*-1}^e) < -1/\lambda_2$ will also imply no direct sales. Since $-\mathcal{C}_{\theta^*}/(\bar{v}_{\theta^*}^e - \underline{v}_{\theta^*}^e) \leq -1/\lambda_2$ implies $(\pi_{\theta^*-1}^e)'(\bar{v}_{\theta^*-1}^e) < -1/\lambda_2$, rent-only will be optimal in all periods if $\mathcal{C}_{\theta^*} \geq (\bar{v}_{\theta^*}^e - \underline{v}_{\theta^*}^e)/\lambda_2$. Setting $\hat{\mathcal{C}}^{(2)} = (\bar{v}_{\theta^*}^e - \underline{v}_{\theta^*}^e)/\lambda_2$ completes part (iii) of the proposition.

Now, we examine what happens if $\mathcal{C}_{\theta^*} < (\bar{v}^e_{\theta^*} - \underline{v}^e_{\theta^*})/\lambda_2$. Then,

$$(\pi^e_{\theta^*-1})'(\bar{v}^e_{\theta^*-1}) = -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_{\theta^*}}{\bar{v}^e_{\theta^*} - \underline{v}^e_{\theta^*}} G_{\theta^*}(\kappa) + \frac{1}{\lambda_2}\bar{G}_{\theta^*}(\kappa)\right], \tag{64}$$

which could be either smaller or greater than $-1/\lambda_2$. Note that (63) implies that for all $0 \le t < \theta$, $(\pi^e_t)'(\bar{v}^e_t)$ is increasing in $(\pi^e_{\theta^*-1})'(\bar{v}^e_{\theta^*-1})$. Therefore, there exists a threshold B_{θ^*} such that $(\pi^b_1)'(\bar{v}^b_1) < -\delta/(\gamma\tilde{\lambda}_2)$ if $(\pi^e_{\theta^*-1})'(\bar{v}^e_{\theta^*-1}) < B_{\theta^*}$ and $(\pi^b_1)'(\bar{v}^b_1) \ge -\delta/(\gamma\tilde{\lambda}_2)$ otherwise. More precisely, to ensure uniqueness of B_{θ^*}, we define it as $B_{\theta^*} = \inf\{B \in \mathbb{R} | (\pi^e_{\theta^*-1})'(\bar{v}^e_{\theta^*-1}) \ge B \Rightarrow (\pi^b_1)'(\bar{v}^b_1) \ge -\delta/(\gamma\tilde{\lambda}_2)\}$. Then, using (64) we can verify that $-\mathcal{C}_{\theta^*-1}/(\bar{v}^e_{\theta^*-1} - \underline{v}^e_{\theta^*-1}) \ge (\pi^e_{\theta^*-1})'(\bar{v}^e_{\theta^*-1}) \ge -\gamma/(\delta\lambda_2)$. Consequently, if B_{θ^*} is equal to the upper bound $-\mathcal{C}_{\theta^*-1}/(\bar{v}^e_{\theta^*-1} - \underline{v}^e_{\theta^*-1})$, there will be no possibility of attaining ownership in the first period; otherwise, ownership is possible if $(\pi^e_{\theta^*-1})'(\bar{v}^e_{\theta^*-1}) \ge B_{\theta^*}$. From (64), the statement of parts (i) and (ii) of the proposition then follow by setting

$$\hat{\mathcal{C}}^{(1)} = \left[-\frac{\delta}{\gamma}B_{\theta^*} - \frac{1}{\lambda_2}\bar{G}_{\theta^*}(\kappa)\right]^+ \frac{\bar{v}^e_{\theta^*} - \underline{v}^e_{\theta^*}}{\lambda_2}. \qquad \square$$