Electronic Companion to "Designing Payment Models for the Poor"
by Bhavani Shanker Uppari and Saša Zorc

This electronic companion (EC) provides supplementary material to the main paper. EC.1 describes the practical impact of our work at d.light. EC.2 summarizes the main notation. EC.3 presents additional results for a more general contract initiation mechanism offering consumers a menu of downpayments. EC.4 discusses the implementation of the optimal contract. EC.5 analyzes the effects of product costs and cost subsidies. EC.6 considers contracts with a repayment term that may differ from the product's lifetime. EC.7 discusses how three prevalent business models in BoP markets (sales, rent-to-own, and rental) arise as special cases of the optimal contracting framework. EC.8 formally models the approximate contract. EC.9 compares, through numerical analysis, the efficiency of current industry practices with both the optimal and approximate contracts. EC.10 extends the model to allow serially correlated income. Finally, EC.11 provides the proofs of all mathematical results.

EC.1. A Points-Based Incentive Program at d.light Inspired by Our Work

Our theoretical framework inspired a points-based customer loyalty and repayment incentive program at d.light, a pioneer in the PAYGo solar industry. The program has been rolled out in Kenya, Tanzania, Nigeria, Uganda, and Zambia, reaching over 1.5 million active customers.

Before our engagement, d.light's operations closely resembled the model described in Section 5.3. Additionally, customers were encouraged to pay daily, with the flexibility to pay in advance or skip days. For each payment, they received a digital "token" unlocking the solar device for one day of use. Access, ownership transfer, and confiscation were all determined by cumulative payment thresholds. While the company tracked repayment behavior, its incentive structure remained largely passive, relying on repayment rates rather than active reward mechanisms. Our discussions with the firm's Credit & Collections team introduced the concept of the v-score—a dynamic mechanism that could both summarize payment history and generate active incentives. Although the formal v-scoring mechanism was too abstract for immediate deployment, it provided the conceptual foundation for d.light's points-based program, which operationalized the underlying logic in a simpler, more intuitive way for customers and field agents.

Under this program, each (quantum of) payment earned the customer a point, with additional bonus points for payment "streaks" to reward consistency. For example, maintaining daily payments for seven consecutive days earned a 50-point bonus, increasing to 80 points for the next 7-day streak, and so on. Points could be redeemed in two ways: (i) liquidity relief—converted into temporary access tokens when cash-constrained, and (ii) loyalty benefits—discounts on future purchases from d.light. Customers with low points were placed on reduced payment schedules, mirroring the incentive differentiation across v-scores in our model. As per d.light, the program significantly improved repayment regularity and strengthened customer engagement, marking a clear shift from passive to active incentive management.

After we developed the approximate contract (Section 5.2), further refinements emerged. A key limitation in d.light's system was that points only increased and never decreased, leading to an imbalance of "carrots without sticks." This caused excessive point accumulation—a liability on the firm's books, albeit non-cash—and weakened the link between points and repayment discipline. Building on the v-score logic, we proposed not only rewarding customers for their "positive" streaks but also penalizing them for "negative" ones—that is, if a customer fails to pay for seven consecutive days, she loses a certain number of points. This adjustment has multiple benefits: (i) it extends d.light's existing streak logic in an easily communicable way, (ii) prevents uncontrolled point accumulation, (iii) allows points to serve as a summary of repayment history, and (iv) enables contract adjustments and terminations to be tied directly to point tiers. Under this revised framework, customers with persistently low points could face early warnings or confiscation, while those with high points could earn accelerated ownership. d.light has since been experimenting with these refinements. Overall, this case study demonstrates how insights from our theoretical model translated into a large-scale, real-world implementation that aligns customer incentives while strengthening firm performance, how practitioners can adapt the optimal contract's features to fit their own business models, and how these features can be implemented incrementally rather than all at once.

EC.2. Table of Notation

Consumer's consumption utility function	$u(x) = \lambda_1 x + (\lambda_2 - \lambda_1)[x - \kappa]^+$
Consumer's access utility function	$u_a(f) = fA$
Consumer's total utility function	$\mathcal{U}(x, f) = u(x) + u_a(f)$
Consumer's income in period t	Random variable Y_t and realization y_t
Consumer's discount factor	δ
Firm's discount factor	γ
Lifetime of the device	T
Firm's confiscation payoff in period t	$\mathcal{C}_t$
Firm's ownership transfer payoff in period t	$\mathcal{O}_t$
Firm's product cost	$\mathcal{K}$
Access provided to consumer	f_t in period t, and $f^t = \{f_s\}_{s=t+1}^T$
Confiscation probability	c_t in period t, and $c^t = \{f_s\}_{s=t+1}^T$
Ownership probability	o_t in period t, and $o^t = \{o_s\}_{s=t+1}^T$
Incentive-compatible payment of the consumer	p_t in period t, and $p^t = \{p_s\}_{s=t+1}^T$
Continuation contract-strategy pair	(f^t, c^t, o^t, p^t)
Sequential payoff of firm in period t	$\Pi_t(f^t, c^t, o^t, p^t)$, given by (2)
Sequential value of consumer in period t	$V_t(f^t, c^t, o^t, p^t)$, given by (3)
Consumer's expected utility from consuming Y_t	$U_t = \mathbb{E}u(Y_t)$
Consumer's promised future value:	
at the *beginning* of period t	$v_t^b \in [\underline{v}_t^b, \overline{v}_t^b]$, $\underline{v}_t^b = U_t + \underline{v}_t^p$, and $\overline{v}_t^b = U_t + \overline{v}_t^p$
after the *payment* in period t	$v_t^p \in [\underline{v}_t^p, \overline{v}_t^p]$, $\underline{v}_t^p = \underline{v}_t^a$, and $\overline{v}_t^p = A + \overline{v}_t^a$
after providing *access* in period t	$v_t^a \in [\underline{v}_t^a, \overline{v}_t^a]$, $\underline{v}_t^a = \underline{v}_t^e$, and $\overline{v}_t^a = \overline{v}_t^e$
at the *end* of period t	$v_t^e \in [\underline{v}_t^e, \overline{v}_t^e]$, both $\underline{v}_t^e$ and $\overline{v}_t^e$ given by (1)
Firm's recursive payoff:	
at the *beginning* of period t	$\pi_t^b(v_t^b)$, solved in Section 4.3
after the *payment* in period t	$\pi_t^p(v_t^p)$, solved in Section 4.2
after providing *access* in period t	$\pi_t^a(v_t^a)$, solved in Section 4.1
at the *end* of period t	$\pi_t^e(v_t^e) = \gamma \pi_{t+1}^b(v_t^e/\delta)$
Relation between sequential and recursive payoffs	$\pi_t^e(v_t^e) = \max \Pi_t(f^t, c^t, o^t, p^t)$, see Section 3.4
Firm's profit per take-up	π_0, given by (17)
Reward function in period t	$\min\{\alpha_t + \lambda_1 p + (\lambda_2 - \lambda_1)[p - \kappa_t^v]^+, \overline{v}_t^p\}$
Fixed payment-independent part of reward	α_t
Expected payment-dependent part of reward	$\omega_t = v_t^b - \alpha_t$
Kink in (i.e., sharpness of) the reward function	κ_t^v
For contract initiation:	
Consumer's liquid wealth	Random variable W and realization w
Consumer's wealth consumption utility function	$\tilde{u}(x) = \tilde{\lambda}_1 x + (\tilde{\lambda}_2 - \tilde{\lambda}_1)[x - \tilde{\kappa}]^+$
Downpayment for initiation	d
Starting v-score upon initiation	v_1^b

Table EC.1 Important variables, parameters, and functions used in the paper

EC.3. An Alternative Contract Initiation Method

Under this alternative 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. (Hereafter, we use $\pi_0^{(1)}$ to denote the profit under the first initiation method that we discussed in Section 4.5, and accordingly, the profit under this second method is denoted as $\pi_0^{(2)}$.)

$$\pi_0^{(2)} = \max_{d(\cdot)\in\mathcal{D},\, v_1^b(\cdot)\in\mathcal{V}_0,\, \underline{d}\in\bar{\mathbb{R}}_+} \mathbb{E}\big[d(W) + \gamma\pi_1^b(v_1^b(d(W))) - \mathcal{K} \,\big|\, d(W) \geq \underline{d}\big] \times \Pr\big\{d(W) \geq \underline{d}\big\} \tag{EC.1}$$

$$\text{s.t.} \quad d(w) \in \arg\max_{0\leq d\leq w} \begin{cases} \tilde{u}(w-d) + \delta v_1^b(d) & \text{if } d \geq \underline{d}, \\ \tilde{u}(w) + \delta \underline{v}_1^b & \text{otherwise,} \end{cases} \quad \forall\, w \in \mathbb{R}_+. \tag{EC.2}$$

The term in the brackets in (EC.1) 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 and semi-differentiable, and satisfy $d(w) \leq w$; 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 semi-differentiable. The incentive compatibility constraint in (EC.2) 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 (EC.1)–(EC.2) is structurally similar to the problem (9)–(11) 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 EC.1. *For a given $\underline{d}$, there exist $\alpha_0, \kappa_0^v \in \mathbb{R}_+$ such that an optimal reward function is given by*

$$v_1^{b*}(d\,;\, \alpha_0, \kappa_0^v) = \begin{cases} \big(\alpha_0 + \tilde{\lambda}_1 d + \big(\tilde{\lambda}_2 - \tilde{\lambda}_1\big)[d-\kappa_0^v]^+\big)/\delta & \mid \quad 0 \leq d < \hat{d}_0(\alpha_0, \kappa_0^v), \quad \textit{and} \\ \overline{v}_1^b & \mid \quad d \geq \hat{d}_0(\alpha_0, \kappa_0^v), \end{cases} \tag{EC.3}$$

where $\alpha_0 \geq \delta\underline{v}_1^b$, 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 $\underline{w}(\alpha_0, \kappa_0^v, \underline{d}) = \inf\big\{w \in [\underline{d}, \infty) \mid \tilde{u}(w-\underline{d}) + \delta v_1^{b*}(\underline{d}\,;\, \alpha_0, \kappa_0^v) \geq \tilde{u}(w) + \delta\underline{v}_1^b\big\}$, 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

$$d^*(w\,;\, \alpha_0, \kappa_0^v, \underline{d}) = \begin{cases} [w + \underline{d} - \underline{w}(\alpha_0, \kappa_0^v, \underline{d})]^+ & \mid \quad 0 \leq w < \hat{d}_1(\alpha_0, \kappa_0^v, \underline{d}) + \underline{w}(\alpha_0, \kappa_0^v, \underline{d}) - \underline{d}, \\ \hat{d}_1(\alpha_0, \kappa_0^v, \underline{d}) & \mid \quad \hat{d}_1(\alpha_0, \kappa_0^v, \underline{d}) + \underline{w}(\alpha_0, \kappa_0^v, \underline{d}) - \underline{d} \leq w < \hat{d}_1(\alpha_0, \kappa_0^v, \underline{d}) + \tilde{\kappa}, \\ w - \tilde{\kappa} & \mid \quad \hat{d}_1(\alpha_0, \kappa_0^v, \underline{d}) + \tilde{\kappa} \leq w < \hat{d}_2(\alpha_0, \kappa_0^v, \underline{d}) + \tilde{\kappa}, \\ \hat{d}_2(\alpha_0, \kappa_0^v, \underline{d}) & \mid \quad w \geq \hat{d}_2(\alpha_0, \kappa_0^v, \underline{d}) + \tilde{\kappa}, \end{cases}$$

where $\hat{d}_1(\alpha_0, \kappa_0^v, \underline{d}) = \max\big\{\underline{d},\, \min\big\{\kappa_0^v,\, [\delta\Lambda_{1,0} - \alpha_0]^+/\tilde{\lambda}_1,\, \hat{d}_0(\alpha_0, \kappa_0^v)\big\}\big\}$, $\hat{d}_2(\alpha_0, \kappa_0^v, \underline{d}) = \max\big\{\underline{d},\, \big(\delta\Lambda_{2,0} - \alpha_0 - \big(\tilde{\lambda}_1 - \tilde{\lambda}_2\big)\kappa_0^v\big)/\tilde{\lambda}_2,\, \min\big\{\kappa_0^v,\, \hat{d}_0(\alpha_0, \kappa_0^v)\big\}\big\}$, and $\Lambda_{j,0} = \max\big\{\underline{v}_1^b,\, \sup\big\{v \in [\underline{v}_1^b, \overline{v}_1^b] \mid \pi_1^{b\prime}(v) \geq -\delta/\big(\gamma\tilde{\lambda}_j\big)\big\}\big\}$, 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 (low) wealth to make a higher (lower) downpayment and in turn be promised a higher (lower) value at the start of the contract. Proposition EC.1 reduces the problem (EC.1)–(EC.2) into a simpler one with three parameters α_0, κ_0^v, and $\underline{d}$:

$$\begin{aligned} \pi_0^{(2)} = \max_{\alpha_0\in[\delta\underline{v}_1^b, \delta\overline{v}_1^b],\, \kappa_0^v\in\mathbb{R}_+,\, \underline{d}\in\bar{\mathbb{R}}_+} \mathbb{E}\big[\mathbb{1}_{\{W\geq\underline{w}(\alpha_0, \kappa_0^v, \underline{d})\}}\big(&d^*(W\,;\, \alpha_0, \kappa_0^v, \underline{d}) \\ &+ \gamma\pi_1^b(v_1^{b*}(d^*(W\,;\, \alpha_0, \kappa_0^v, \underline{d})\,;\, \alpha_0, \kappa_0^v)) - \mathcal{K}\big)\big]. \end{aligned}$$

PROPOSITION EC.2. *The following statements hold:*

(i) *For every* d^* *that is optimal in* (17) *there exists* $\underline{d}^* \leq d^*$ *that is optimal in* (EC.1)–(EC.2)*; 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:* $\pi_0^{(2)} \geq \pi_0^{(1)}$.

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 (17) 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 EC.2(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. To understand why this happens, notice the trade-off in (17): 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 EC.7.) 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.

EC.4. Implementation of the Optimal Contract

The implementation of contract initiation is straightforward. The consumer is asked to make a downpayment d^*. If she pays, the contract starts with v-score v_1^{b*}. After the contract starts, the necessary details can be communicated through two digital displays $\mathfrak{D}_1$ and $\mathfrak{D}_2$ on consumer's device, see Figure EC.1(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 be updated to $v_t^{(j)}$." Note from the problem formulation in Section 4.3 that 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, which is displayed in $\mathfrak{D}_2$.

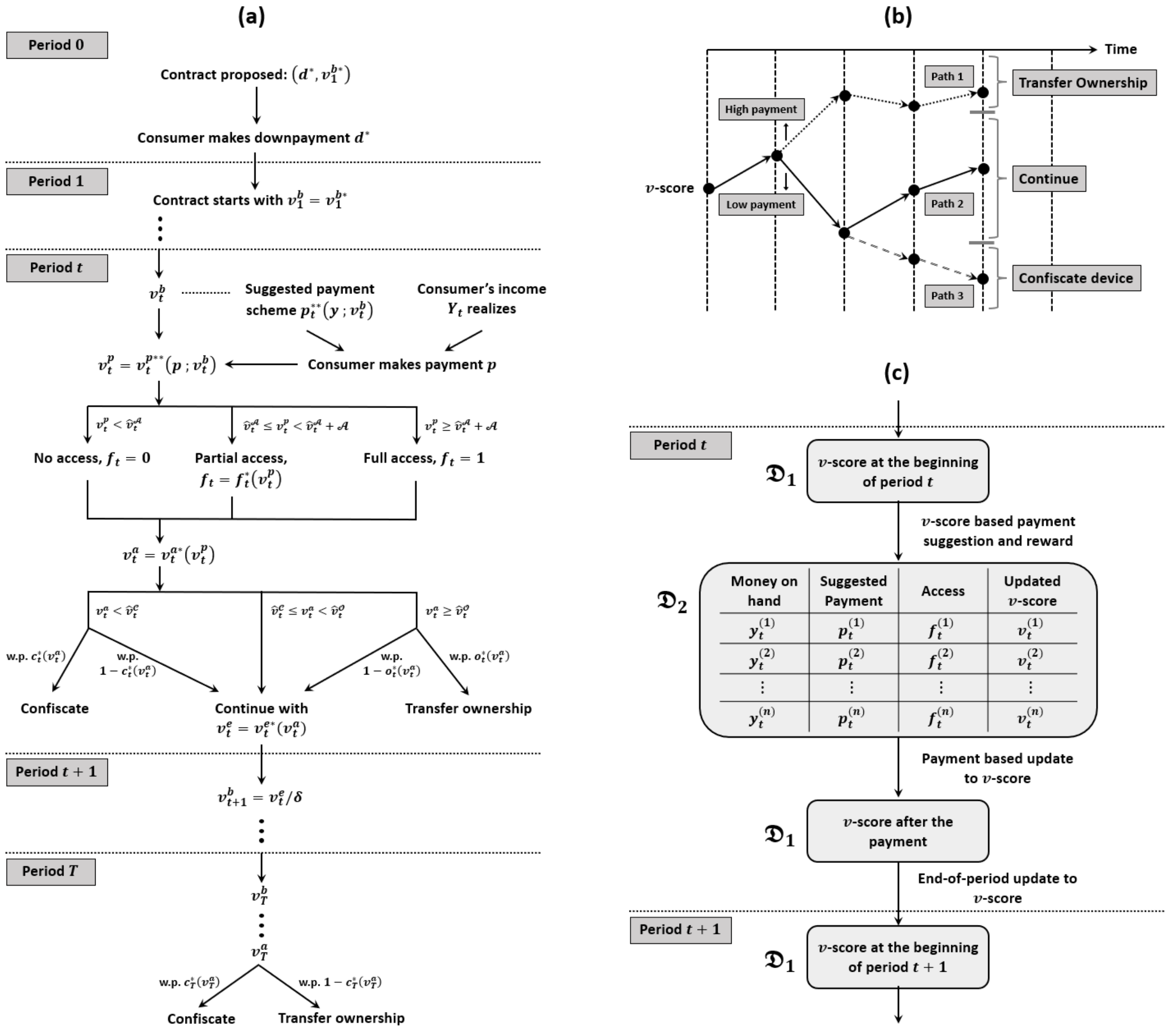

Figure EC.1 (a) Summary of the contract dynamics. (b) Three sample paths of v-scores, starting from the same initial score and ending in ownership, continuation, or confiscation under different payment paths. (c) Implementation of contract using digital displays $\mathfrak{D}_1$ and $\mathfrak{D}_2$.

Figure EC.1(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 information in $\mathfrak{D}_2$ is based on this v-score. After the consumer makes a payment $p_t^{(j)}$, $\mathfrak{D}_1$ displays the updated v-score $v_t^{(j)}$. At the end of the period, if not terminated, the v-score is updated to v_{t+1}^b while moving to the next period. The same update process continues thereafter. All the updates can be done automatically and remotely using IoT technology, and at a frequency that the firm deems appropriate. The confiscation and ownership transfer rules of the v-scoring mechanism 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 to highlight the confiscation threat, the rows that result in continuation can be displayed in green, and blue can be used for the rows that offer ownership. $\mathfrak{D}_1$ too can dynamically change its display color depending on what the current v-score entails.

The described implementation requires the firm to supplement its devices with digital displays and update them periodically with incentive information. Several of the existing devices already come with digital displays showing other information (e.g., battery condition), which could also be used to pass on the scoring and incentive information. Moreover, the IoT infrastructure required for the dynamic updates already exists and is used by the PAYGo firms. Therefore, we believe implementing the v-scoring mechanism may not be technologically challenging. However, as we mentioned in Section 4.6, a hurdle in its implementation is the complexity of the v-scoring mechanism. The dynamically updating incentive mechanism will result in lookup tables in Figure EC.1(c) that update every period, which can be challenging for consumers to comprehend. This complexity motivated us to design an approximate contract that is easier to implement.

EC.5. Impact of Product Cost and Cost Subsidies

Recall that we defined $\mathcal{K}$ as the cost of getting the technology into the consumer's hands. Therefore, the product cost—even for the same product—could vary depending on multiple factors such as different customs tariffs in different countries, varying logistics costs based on how accessible the rural consumers are, the associated risks of transporting devices there, and the marketing or educational campaigns required to increase the consumers' awareness of the technology. First, note that in our model, $\mathcal{K}$ features only in the contract initiation problem. Therefore, the product cost does not affect the terms of the continuation contract. It only impacts the downpayment that is charged to the consumer and her starting value of v-score.

Second, using our running example, Figure EC.2 shows the impact of increasing product cost on the downpayment, starting v-score, take-up rate, firm's profit, and consumer welfare. Interestingly, for the cost values below a threshold $\hat{\mathcal{K}} = d^* + \gamma\pi_1^b(v_1^{b*})$ (= \$42 in Figure EC.2), the contract retains the same optimal values $(d^*, v_1^{b*}) = (0, 0.6)$ as if there is no impact of product cost, but beyond $\hat{\mathcal{K}}$, the contract starts to charge a downpayment. To make sense of this observation, let us look at the levers that the firm has at its disposal to compensate for the negative impact of increasing $\mathcal{K}$ on the profit per device $d + \gamma\pi_1^b(v_1^b) - \mathcal{K}$. The firm could either (i) increase d, as that would increase profit per device (Figure 5(a)) but that can also sharply drop the take-up rate (Figure 5(c)), or (ii) increase v_1^b (beyond v_1^{b*}) because that increases take-up rate (Figure 5(d)), but doing so also decreases firm's profit (Figure 5(b)). So, both the levers (i) and (ii) have ambiguous effects on the overall profit. The firm finds it optimal not to deploy either of the levers for $\mathcal{K} < \hat{\mathcal{K}}$ and maintain (d^*, v_1^{b*}) at $(0, 0.6)$ as that would maintain the take-up rate at 100% (see Figure EC.2(c)) while resulting in a positive, but linearly decreasing, profit (because $d^* + \gamma\pi_1^b(v_1^{b*}) > \mathcal{K}$), as shown in Figure EC.2(d). But when the cost $\mathcal{K} \geq \hat{\mathcal{K}}$, the profit per device at (d^*, v_1^{b*}) becomes negative, so the firm starts to deploy both the levers: it charges a downpayment and simultaneously also increases v_1^b, as we see in Figures EC.2(a)–(b). The impact of the drop in take-up due to increasing d dominates the opposite effect of increasing v_1^b, so the overall take-up drops beyond $\hat{\mathcal{K}}$ and the firm's profit continues to decrease, as shown in Figures EC.2(c)–(d). Although the starting v-score increases in $\mathcal{K}$, because of the concomitant increase in downpayment, the overall consumer welfare falls beyond $\hat{\mathcal{K}}$, as shown in Figure EC.2(e).

Finally, the above discussion also sheds light on the value of offering cost subsidies to the firm. An external party, such as a donor or a policy organization, could help the firm reduce the costs of getting the devices into consumer's hands by offering a subsidy on every device manufactured or purchased by the firm, waiving off tariffs, partnering in the educational campaigns, or reducing the logistics costs and risks. The overall impact of such actions is that we move leftward on the horizontal axes of Figure EC.2. For relatively low values of $\mathcal{K}$, offering cost subsidies does not

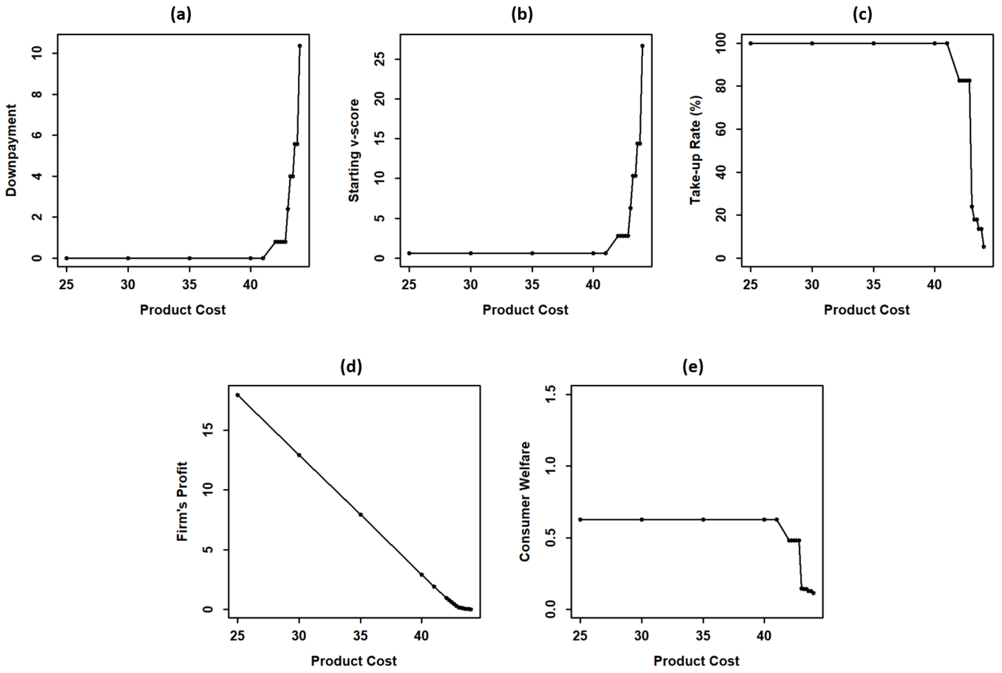

Figure EC.2 Impact of product cost: (a) and (b) plot the optimal downpayment and the optimal starting v-score as functions of $\mathcal{K}$; (c), (d), and (e) plot the take-up rate, firm's profit, and consumer welfare, respectively, as functions of $\mathcal{K}$, evaluated at the corresponding optimal downpayment and starting v-score.

affect the take-up rate and consumer welfare but increases the firm's profit. Thus, for instance, in the markets where the consumers are easily accessible or already aware of the benefits of the technology, and when the technology is not costly, there is no benefit in offering cost subsidies if the focus is to increase take-up or consumer welfare. However, when $\mathcal{K}$ is high, cost subsidies increase the take-up, consumer welfare, and firm profit. So, the cost subsidies can be beneficial in markets that are difficult to access, require substantial educational campaigns, or when the technology is costly. Moreover, the firms may sometimes choose not to enter a market because of low profits; in such cases, the cost subsidies can always help break the barriers to entry.

EC.6. 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 EC.1(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 Proposition 1: 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 (also see Figure EC.3(b) presented in the proof of Proposition 1). 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 EC.3. *If there exists a time period θ such that*

$$\frac{\mathcal{C}_{t-1}}{\mathcal{C}_t} \leq \frac{\gamma}{\delta} \times \frac{\overline{v}^e_{t-1} - \underline{v}^e_{t-1}}{\overline{v}^e_t - \underline{v}^e_t}, \qquad \textit{for } 1 < t \leq \theta, \quad \textit{and} \tag{EC.4}$$

$$\frac{\mathcal{C}_{t-1}}{\mathcal{C}_t} \geq \frac{\gamma}{\delta} \times \frac{\overline{v}^e_{t-1} - \underline{v}^e_{t-1}}{\overline{v}^e_t - \underline{v}^e_t} \times \max\left\{1, \frac{\overline{v}^e_t - \underline{v}^e_t}{\lambda_2 \mathcal{C}_t}\right\} \quad \textit{for } \theta < t \leq T, \tag{EC.5}$$

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 EC.3 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 (EC.4) until a period θ, and thereafter the relative fall is bounded above by a time-varying threshold in (EC.5). It is easy to show that the time-varying thresholds on the right-hand sides of (EC.4) and (EC.5) are increasing in t. Thus, Proposition EC.3 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.

EC.7. 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 EC.6, 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 (this is easy to see in Figure EC.3(a) presented in the proof of Proposition 1). 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 2(b) and 2(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 EC.3. The option of ownership exists in period 0 if $v_1^{b**}(\hat{d}_2^*) = \overline{v}_1^b$, where the threshold $\hat{d}_2^*$ is $\hat{d}_2$ in Proposition EC.1 evaluated at the optimal values of α_0, κ_0^v, 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 (EC.4) and (EC.5). Thus, the structure of the result holds for a repayment term θ^* that is either endogenously determined as per Proposition EC.3, 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 EC.4. *For a firm with the set of confiscation values $\{\mathcal{C}_t\}$ that satisfy $(EC.4)$ and $(EC.5)$ 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 EC.6 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 a high $\mathcal{C}_{\theta^*}$ 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, the firm in this case incurs an operational cost to deliver the service (e.g., it needs to maintain the purification systems or replace 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 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 is indeed the case here. Furthermore, because the single-period payoff of the firm is bounded and discounted, it follows directly from Proposition 1.2.1 of Bertsekas (2012) that the infinite-horizon problem is well-defined in our case and is equivalent to the limit of the finite-horizon version of the model when the number of periods goes to infinity, and the optimal policy can be solved accordingly.

In terms of implementation, the features (OPT-1)–(OPT-7) of the optimal contract carry over to the infinite-horizon case, with a modification to (OPT-6) that there is no ownership option, and hence there will be no ownership threshold in this case. The same applies to the approximate contract as well: a contract to provide clean water or cooking gas service can look like the one in Figure 8, but without the ownership clause. It would be appropriate to specify the access in this case, not in terms of hours, but in terms of quantities (e.g., liters) of water or gas that will be provided.

EC.8. Model of Approximate Contract

As we can see in Figure 8, the approximate contract operates at the level of time blocks, and it does not stipulate the dynamics within a block. Therefore, it makes sense to model its dynamics in terms of time blocks rather than time periods. However, to be able to compare the performance of different contracts, the approximate contract must be formalized in terms of the same model primitives used for the other contracts. This requires us to make a few assumptions for transforming the period-level framework that we have been using to a block-level framework, with the core idea being that the decisions, although now made at the block level, gradually impact all the periods in the block. The details of these assumptions, along with some new notation, are given next.

1. With a slight abuse of notation, we continue to use t to represent time blocks. We denote by b the size of a time block and let B be the total number of blocks in the time duration T. Therefore, $B \approx \lceil T/b \rceil$ and $t \in \{1, \ldots, B\}$.

2. We denote by z_t the credit points that the consumer has at the beginning of block t.
3. Let $\mathcal{T}(z)$ denote the function that returns the tier to which the consumer belongs based on her credit points z. $\mathcal{T}(z)$ is increasing in z.
4. The tier to which the consumer belongs, in turn, determines the expected payment window and the rates at which she gains or loses points. Given a tier $\mathcal{T}$, the lower limit of the payment window is denoted by $\underline{p}_{\mathcal{T}}$, the upper limit by $\overline{p}_{\mathcal{T}}$, the rate at which the consumer gains points is $\lambda^{+}_{\mathcal{T}}$, and the loss rate for points is $\lambda^{-}_{\mathcal{T}}$.
5. The access cost is β: i.e., it costs consumer β points for access to the technology for one full period.
6. As we mentioned above, we need to construct the income random variable at the block level using the income random variables at the period level. We simply define the former as the sum of b period-level income random variables in the block. For example, if the period-level income random variable, as we assumed in our running example, is Exponential with mean μ, then the block-level income random variable is Gamma distributed with shape parameter b and scale parameter μ. With a slight notational abuse, we denote the block-level income random variable as Y_t.
7. We also need to construct consumer's utility function at the block level using the period-level utility function u. Let $u^b(c)$ denote the utility that the consumer derives from consuming c over the course of a time block. We assume that this quantity c is consumed uniformly over all the periods in a block; therefore, the consumer derives utility $u(c/b)$ in each period in a block, and after accounting for discounting within a block, we have the following relationship: $u^b(c) = \sum_{s=0}^{b-1} \delta^s u(c/b)$.
8. As in the period-level model, we assume that the consumer makes her block-level payment decision p_t at the beginning of block t, but this payment is made to the firm uniformly over the periods in the block. The decision p_t in our model is made after the realization y_t of the random variable Y_t, but we do not necessarily need to assume that the consumer is in possession of all the realized income at the beginning of the block. The gradual payment and gradual consumption assumptions that we discussed above imply that the consumer pays p_t/b and consumes $y_t/b - p_t/b$ in each period within the block. In this sense, one could interpret y_t as the total income that the consumer anticipates to realize in the duration of the block (and we assume that these anticipations are correct), and that y_t could realize gradually over the course of the block (but to simplify, we assume that this realization happens uniformly over the periods in the block).
9. With the payment pattern just discussed, the firm's discounted revenue from a block-level payment p_t is equal to $\sum_{s=0}^{b-1} \gamma^s (p_t/b)$.
10. The access policy under the approximate contract is that the consumer is charged β points per period for full access, with also the provision of fractional access. If the points z possessed by the consumer at the beginning of a period exceed βb, then she gets full access for the entire time block. But if $z < \beta b$, then she is uniformly provided a fractional access of $z/(\beta b)$ in each period of the block. After taking into account within-block discounting, the consumer's utility from this access policy is given by $Af_t(z) = \sum_{s=0}^{b-1} \delta^s A \min\{z/(\beta b), 1\}$.
11. The confiscation policy under the approximate contract is $c_t(z) = \mathbb{1}\{z \leq \hat{z}^{\mathcal{C}}_t\}$: the device is confiscated if the consumer's points fall below a confiscation threshold $\hat{z}^{\mathcal{C}}_t$.

12. The ownership policy under the approximate contract is $o_t(z) = \mathbb{1}\{z \geq \hat{z}_t^{\mathcal{O}}\}$, where the ownership threshold is $\hat{z}_t^{\mathcal{O}} = \beta(T-t)$. In other words, the ownership is transferred to the consumer when the points acquired by her exceed the points required to access the device for its remaining lifetime.
13. The approximate contract determines the access that the consumer gets in a time block based on the points at the beginning of the time block (i.e., based on the payments she made so far), and decides whether to continue or terminate the contract based on her updated points at the end of the time block (which also depends on the payment made in the current block).
14. Finally, we denote by $Z(p,z)$ the function that updates the consumer's points based on her block-level payment p and her points z at the beginning of the block. It is given by

$$Z(p,z) = z - \min\{z, \beta b\} + \lambda^-_{\mathcal{T}(z)}\left[p - \underline{p}_{\mathcal{T}(z)}\right]^- + \lambda^+_{\mathcal{T}(z)}\left[\min\{p, \overline{p}_{\mathcal{T}(z)}\} - \underline{p}_{\mathcal{T}(z)}\right]^+.$$

The update has two parts. First, $\min\{z, \beta b\}$ points are deducted for access to the device. Second, the remaining points are updated based on the magnitude of consumer's payment p relative to her payment window $\left[\underline{p}_{\mathcal{T}(z)}, \overline{p}_{\mathcal{T}(z)}\right]$, which depends on her tier $\mathcal{T}(z)$. There is no update to consumer's (remaining) points if she pays exactly the minimum expected payment $\underline{p}_{\mathcal{T}(z)}$. If the consumer pays below the lower limit $\underline{p}_{\mathcal{T}(z)}$ of the payment window, then she loses points proportional to her deviation from the lower limit, at a rate of $\lambda^-_{\mathcal{T}(z)}$ per dollar. If she pays above the lower limit, she gains points proportional to how much she paid above the lower limit, but only up to the upper limit $\overline{p}_{\mathcal{T}(z)}$ of the payment window, at a rate of $\lambda^+_{\mathcal{T}(z)}$ per dollar. The gain and loss rates are dependent on tier $\mathcal{T}(z)$.

With these assumptions and notations in hand, we will now write the consumer's continuation value function V_t^{AP} in block t. It requires the following two state variables: (i) the realization of income y_t in block t, and (ii) the consumer's points z_t at the beginning of block t. The consumer's payment problem is:

$$\begin{aligned} V_t^{AP}(y_t, z_t) = \max_{0 \leq p_t \leq y_t} u^b(y_t - p_t) + Af_t(z_t) + \delta^{b-1}\big[&\underline{v}_t^e c_t(Z(p_t,z_t)) + \overline{v}_t^e o_t(Z(p_t,z_t)) \\ &+ (1 - c_t(Z(p_t,z_t)) - o_t(Z(p_t,z_t)))\delta\,\mathbb{E}V_{t+1}(Y_{b+1}, Z(p_t,z_t))\big]. \end{aligned}$$

The problem is written from the perspective of the beginning of block t, which is why the value-to-go is discounted by δ^{b-1}. The termination payoffs $\underline{v}_t^e$ and $\overline{v}_t^e$ are as defined in (1), but with t here representing blocks instead of periods. Solving the above problem gives us the optimal payment policy of the consumer $p_t^{AP*}(y_t\,;\,z_t)$. Given this payment policy, the firm's expected continuation payoff, Π_t^{AP}, in block t is computed as follows: $\Pi_t^{AP}(z_t) = \mathbb{E}\check{\Pi}_t^{AP}(y_t, z_t)$, where

$$\begin{aligned} \check{\Pi}_t^{AP}(y_t, z_t) = \sum_{s=0}^{b-1}\gamma^s(p_t^{AP*}/b) + \gamma^{b-1}\big[&\mathcal{C}_t c_t(Z(p_t^{AP*}, z_t)) + \mathcal{O}_t o_t(Z(p_t^{AP*}, z_t)) \\ &+ (1 - c_t(Z(p_t^{AP*}, z_t)) - o_t(Z(p_t^{AP*}, z_t)))\gamma\,\Pi_{t+1}^{AP}(Z(p_t^{AP*}, z_t))\big], \end{aligned}$$

and $p_t^{AP*} \equiv p_t^{AP*}(y_t\,;\,z_t)$. As earlier, given a downpayment d, a consumer with wealth W enters the contract only if the following two conditions are satisfied: (i) $W \geq d$, and (ii) $\tilde{u}(W-d) + \delta\,\mathbb{E}V_1(Y_1, z_1) \geq \tilde{u}(W) + \delta\underline{v}_1^b$. The firm's problem then becomes

$$\pi_0^{AP} = \max_{d, z_1}\left(d + \gamma\Pi_1^{AP}(z_1) - \mathcal{K}\right) \times \Pr\left\{W \geq d, \tilde{u}(W-d) + \delta\,\mathbb{E}V_1^{AP}(Y_1, z_1) \geq \tilde{u}(W) + \delta\underline{v}_1^b\right\}. \tag{EC.6}$$

Having established the model, we now describe how we analyze it numerically. Our objective is to identify the profit-maximizing parameters for the approximate contract, but this exercise is overly complicated because the model requires several parameters and functions as inputs: $\mathcal{T}(z)$, $\underline{p}_{\mathcal{T}}$, $\overline{p}_{\mathcal{T}}$, $\lambda^+_{\mathcal{T}}$, $\lambda^-_{\mathcal{T}}$, β, and $\hat{z}^{\mathcal{C}}_t$. To simplify the analysis, as we discussed in Section 5.2, we discretize and restrict the possible number of tiers to 4. We also exogenously set the uppermost tier (i.e., the fourth tier) to be the one where the consumer is not incentivized to pay anything (recall the arguments made toward the end of Section 5.1.1). Therefore, all the parameters related to this tier are set to zero. Because the scale of z does not matter in the model, we arbitrarily set the threshold to enter tier 4 as 200 points. (Choosing a smaller value will scale down other parameters and vice versa.) These restrictions result in a simplified structure for $\mathcal{T}(z)$: given two thresholds $\hat{z}_1$ and $\hat{z}_2$ with $\hat{z}_1 < \hat{z}_2$, the tier $\mathcal{T}(z)$ is equal to 1 if $z \leq \hat{z}_1$, 2 if $\hat{z}_1 < z \leq \hat{z}_2$, 3 if $\hat{z}_2 < z \leq 200$, and 4 if $z > 200$. Instead of being a function of z, $\mathcal{T}(z)$ requires only two parameters, $\hat{z}_1$ and $\hat{z}_2$.

With this simplification, the other tier-dependent parameters also get discretized. We already assumed that the access cost is stationary, and we further assume that the confiscation threshold is stationary (inspired by our observations from Section 5.1.4). After these simplifying assumptions, we have the following sets of parameters: (i) the lower limits of payment windows for the three tiers $\{\underline{p}_1, \underline{p}_2, \underline{p}_3\}$, (ii) the corresponding upper limits of payment windows $\{\overline{p}_1, \overline{p}_2, \overline{p}_3\}$, (iii) the gain rates for the three tiers $\{\lambda^+_1, \lambda^+_2, \lambda^+_3\}$, (iv) the respective loss rates $\{\lambda^-_1, \lambda^-_2, \lambda^-_3\}$, (v) the access cost β, (vi) the confiscation threshold $\hat{z}^{\mathcal{C}}$, and (vii) the tier thresholds $\hat{z}_1$ and $\hat{z}_2$. This is a total of 16 parameters, so the grid search for the optimal set of parameters is still computationally prohibitive: even by allowing only five possible values for each parameter, the number of parameter combinations explodes to 5^{14} (around 152 billion). Therefore, we resort to an alternative and simpler—but a suboptimal—approach to identify the parameters. We randomly selected 10,000 parameter combinations from the set of all possible combinations, and chose the parameter set that resulted in the highest profit π^{AP}_0 in (EC.6).

Specifically, for random selection we used the following possible values for the parameters: $\underline{p}_{\mathcal{T}} \in \{0, 0.5, 1, 1.5, 2, 2.5\}$, $\overline{p}_{\mathcal{T}} \in \{5, 10, 15, 20, 25\}$, both $\lambda^+_{\mathcal{T}}$ and $\lambda^-_{\mathcal{T}} \in \{15, 18, 20, 23, 25, 28, 30, 33, 35, 38, 40, 43, 45\}$, both β and $\hat{z}^{\mathcal{C}} \in \{25, 30, 35, 40, 45, 50\}$, $\hat{z}_1 \in \{50, 55, 60, 65, 70\}$, and $\hat{z}_2 \in \{120, 125, 130, 135, 140, 145, 150\}$. Of the 10,000 randomly selected parameter combinations, we found that the following combination—which was also used in Figure 8—resulted in the highest profit: $\hat{z}_1 = 55$, $\hat{z}_2 = 130$, $\underline{p}_1 = 2$, $\underline{p}_2 = 2.5$, $\underline{p}_3 = 1$, $\overline{p}_1 = 20$, $\overline{p}_2 = 10$, $\overline{p}_3 = 5$, $\lambda^+_1 = 35$, $\lambda^+_2 = 35$, $\lambda^+_3 = 38$, $\lambda^-_1 = 20$, $\lambda^-_2 = 38$, $\lambda^-_3 = 35$, $\beta = 30$, and $\hat{z}^{\mathcal{C}} = 30$. By solving (EC.6) with these parameters, we found that $d^* = 0$, $z^*_1 = 35$ points (i.e., the starting credit), and $\pi^{AP}_0 = \$15.64$ per device (as reported in Table EC.2).

Given our optimization procedure, it is important to note that the set of parameters that we have identified need not be optimal. However, interestingly, even with these possibly suboptimal parameters, the approximate contract performs better than the status quo contract (see EC.9), indicating that even the simpler implementations of the optimal contract can potentially improve upon current practices.

EC.9. Model of Current Practices and Comparison of Performance

In this section, we will put the key current practices of PAYGo firms discussed in Section 5.3 into a mathematical framework; we call the resulting model the *status quo contract*. Using our running example, we will demonstrate the differences in the outcomes achieved by this contract vis-à-vis the optimal contract and the approximate contract discussed in Section 5.2 and modeled in EC.8.

We model the status quo contract as a quadruplet (D, P, I, N), where (i) D is the downpayment that a consumer must pay to enter the contract, (ii) P is the total amount that the consumer must pay over the repayment term of the contract to obtain device ownership (P excludes downpayment), (iii) I is the installment that is expected to be paid in each period (accordingly, the suggested repayment term is P/I periods), and (iv) N is the number of non-payment periods allowed after which the technology will be confiscated from the consumer.

The consumer's continuation value function under the status quo contract in period t, denoted by V_t^{SQ}, requires the following three state variables: (i) the realization of her disposable income y_t in period t, (ii) the cumulative amount paid by the consumer until the beginning of period t, which we denote by r_t, and (iii) the number of successive non-payment periods until the beginning of period t, denoted as n_t (note that n_t resets to 0 every time the consumer makes a payment). With this in hand, we can write the consumer's payment problem in period t as follows:

$$\begin{aligned} V_t^{SQ}(y_t, r_t, n_t) = \max_{0 \le p_t \le y_t} \; & u(y_t - p_t) + A f_t(r_t + p_t) + \underline{v}_t^e c_t(n_t + \mathbb{1}\{p_t = 0\}) + \overline{v}_t^e o_t(r_t + p_t) \\ & + \left(1 - c_t(n_t + \mathbb{1}\{p_t = 0\}) - o_t(r_t + p_t)\right) \delta \, \mathbb{E} V_{t+1}(Y_{t+1}, r_t + p_t, \mathbb{1}\{p_t = 0\}(n_t + 1)), \end{aligned}$$

where the access policy f_t, the confiscation policy c_t, and the ownership policy o_t are as per the status quo contract: (1) $f_t(r_t) = \mathbb{1}\{r_t \ge I \times t\}$, i.e., the consumer gets access to the device if and only if her cumulative payment exceeds the cumulative payment expected by period t; (2) $c_t(n_t) = \mathbb{1}\{n_t = N\}$, i.e., the device is confiscated if the number of successive non-payment periods is equal to N; and (3) $o_t(r_t) = \mathbb{1}\{r_t \ge P\}$, i.e., the consumer obtains ownership when her cumulative payment exceeds P.

Solving the above problem gives us the optimal payment policy of the consumer as a function of the state variables: $p_t^{SQ*}(y_t\,;\, r_t, n_t)$. Compare the functions here with their respective versions under the optimal contract: $f_t(v_t^p)$, $c_t(v_t^a)$, $o_t(v_t^a)$, and $p_t^{**}(y_t\,;\, v_t^b)$. The latter set of functions has v-score as the argument, whereas the former set has r_t or n_t as the arguments. So, from the payment history, the status quo contract only uses the information given by r_t and n_t, whereas the optimal contract uses these—and multiple other facets of information provided by payment history—through v-score.

Given the payment policy $p_t^{SQ*}(y_t\,;\, r_t, n_t)$, the firm's expected continuation payoff, Π_t^{SQ}, in period t can be computed as follows:

$$\begin{aligned} \Pi_t^{SQ}(r_t, n_t) &= \mathbb{E}\check{\Pi}_t^{SQ}(y_t, r_t, n_t), \quad \text{where} \\ \check{\Pi}_t^{SQ}(y_t, r_t, n_t) &= p_t^{SQ*} + \mathcal{C}_t c_t(n_t + \mathbb{1}\{p_t^{SQ*} = 0\}) + \mathcal{O}_t o_t(r_t + p_t^{SQ*}) \\ & \quad + \left(1 - c_t(n_t + \mathbb{1}\{p_t^{SQ*} = 0\}) - o_t(r_t + p_t^{SQ*})\right) \gamma \, \Pi_{t+1}^{SQ}(r_t + p_t^{SQ*}, \mathbb{1}\{p_t^{SQ*} = 0\}(n_t + 1)), \end{aligned}$$

and $p_t^{SQ*} \equiv p_t^{SQ*}(y_t\,;\, r_t, n_t)$. As earlier, a consumer with wealth W enters the contract only if the following two conditions are satisfied: (i) $W \ge D$, and (ii) $\tilde{u}(W - D) + \delta\, \mathbb{E} V_1^{SQ}(Y_1, 0, 0) \ge \tilde{u}(W) + \delta \underline{v}_1^b$. The firm's problem is now given by

$$\begin{aligned} \pi_0^{SQ} = \max_{D, P, I, N} \; & \left(D + \gamma \Pi_1^{SQ}(0, 0) - \mathcal{K}\right) \\ & \times \Pr\left\{W \ge D, \tilde{u}(W - D) + \delta\, \mathbb{E} V_1^{SQ}(Y_1, 0, 0) \ge \tilde{u}(W) + \delta \underline{v}_1^b\right\}, \end{aligned} \tag{EC.7}$$

where V_1^{SQ} and Π_1^{SQ} are the functions of parameters P, I, and N.

We solved the problem in (EC.7) by searching for the optimal parameters over a grid of values.[6] We found the following optimal values: $D^* = \$0$, $P^* = \$50$, $I^* = \$1$, and $N^* = 4$ periods. Similar to what we saw in Section 5.1.3, the status quo contract too does not charge a downpayment. The expected installment per week is \$1, and it is expected to be paid over a repayment term of 50 weeks (because the total expected payment is \$50). The value of N—the block of time in which the contract is tolerant to no payment and does not confiscate the device—is equal to 4 periods, which is similar in size to the time block recommended by the optimal contract (i.e., 5 periods).

Table EC.2 presents the performance of the three contracts of interest—status quo contract with the aforementioned parameters, the optimal contract as discussed in Section 5.1, and the approximate contract shown in Figure 8—in terms of the firm's profit, consumer welfare, and confiscation rates. The latter is obtained through simulations. We simulated the model in Sections 3–4 by repeatedly drawing 100 individuals with random initial wealth, offering them the optimal contract, and tracking whether their contracts ended in confiscation or ownership. Each individual's trajectory was driven by the contract terms and their period-by-period random income. This procedure was repeated 1,000 times. Table EC.2 reports the confiscation rates averaged over these runs along with the standard deviations.

The firm's profit under the status quo contract is $\pi_0^{SQ} = \$13.98$ per device, and under the optimal contract is $\pi_0 = \$17.01$ per device, which is a 22% increase on top of the status quo. The profit under the approximate contract, π_0^{AP} (which is formally defined in EC.8), is \$15.64 per device, 12% above the status quo. It is no surprise that the optimal contract outperforms the other two because it is the optimal one in the class of contracts to which the other two also belong. Interestingly, the approximate contract also outperforms the status quo contract. This provides affirmation that even a simpler implementation of the optimal contract has the potential to improve upon the current practices.

	Firm's Profit	Consumer Welfare	Confiscation Rate
Status Quo Contract	13.98	5.0%	28.33 (4.53)
Optimal Contract	17.01	0.6%	20.81 (3.94)
Approximate Contract	15.64	0.6%	24.14 (4.36)

Table EC.2 Performance of three contracts. (The standard deviations are provided for the confiscation rates because they are obtained from simulations.)

As discussed in Section 5.1.3, the optimal contract extracts consumer surplus to the extent possible, so the resultant welfare is 0.6%. The approximate contract also does the same, as seen in Table EC.2, but under the status quo contract, the consumer welfare is 5%. This must not be misconstrued as the latter one being better at consumer welfare; this simply indicates the inefficiency of the status quo contract in incentivizing payments from the consumers, which not only reflects in its lower profit but—when combined with an inefficient confiscation rule—also in the confiscation rates: the confiscation rate under the status quo contract is 7.52%-points higher than the optimal contract and 4.19%-points higher than the approximate contract. Section 5.1.5 discussed how consumer welfare can be increased under the optimal contract while maintaining efficiency. From Figures 7(a)–(b), we note that a welfare weight of $\varpi = 1$ results in a consumer welfare closer to that of the status quo contract (4.8%) but a profit higher than status quo (\$16.83 per device). With $\varpi = 1.1$, the resulting profit (\$13.89 per device) is closer to that of the status quo, but the welfare is 13.5%, much higher than under the status quo contract. These two cases amount to a welfare subsidy of \$1.5 and \$4.5 per device, respectively; see Figure 7(c). (The welfare can be further increased, but that comes at the cost of a lower firm's profit or a higher welfare subsidy.) So, under the optimal contracting framework, if either the firm places an appropriate weight on consumer welfare or an external party incentivizes it through a welfare subsidy, we could increase consumer welfare without compromising contractual efficiency.

EC.10. Model with Serially Correlated Income

One of the assumptions underpinning our model is that the distribution of consumer's income over time—while not necessarily static—evolves in a way that can be perfectly anticipated at contract initiation. However, in reality, unexpected events may happen to the consumer during the contract duration which alter the distribution of their income, e.g., change of jobs or a period of joblessness. With this in mind, we present three extensions of our model that account for correlated income with varying degrees of complexity.

EC.10.1. Income Distribution is Governed by a Publicly Observable Markov Process

We start with the simplest extension that accounts for correlated income but does not add information asymmetries and then proceed with extensions that do. Instead of the consumer income distribution being a deterministic series $\{G_t\}$, here we assume the distribution depends on a vector of *observable* covariates x_t, so that the distribution of consumer income is then $G_t(y_t \mid x_t)$. Here, x_t could include the job status of the consumer, as well as other conditions such as weather (which affects farmers' income). We then assume that (i) these covariates are observable/verifiable, and (ii) the incomes over time are conditionally independent, given the covariate information x_t (a weaker assumption than the unconditional independence assumed in the main model). This setup can be formalized as follows:

ASSUMPTION EC.1. *Let $\{X_t\}$ be a Markov process with discrete time and state, and $\{x_t\}$ its realizations. The state transitions happen at the beginning of every period, before other events. For any time $1 \le i < j$, the consumer's incomes Y_i and Y_j are conditionally independent given X_i. The distribution of consumer's income at time t depends only on the current state of the Markov process: for every x_t in the support of X_t, the distribution of Y_t is given by the cdf $G_t(y_t \mid x_t)$, which is continuous and has a decreasing hazard rate.*

With this assumption, the state of the chain will also become a state variable that the contract tracks. In practical terms, the contract will offer terms that depend on, for example, current job status and weather. This unfortunately complicates the structure of the optimal contract and potentially distances it from practice. However, with this formulation, all of the key mathematical results in our paper will still hold, which we formalize in the following result.

PROPOSITION EC.5. *Let Assumption EC.1 hold, and the transition probabilities and state of $\{X_t\}$ be public knowledge. Then, Propositions 1–6 in the paper still hold, but with the following modifications:*

(i) *The consumer's single-period utility of autarchy is state-dependent: $U_t(x_t) = \mathbb{E}[u(Y_t)|x_t]$.*

(ii) *As an immediate consequence of (i), the bounds on consumer utility are also state-dependent:*

$$\overline{v}_t(x_t) = \overline{v}_t^e(x_t) = \overline{v}_t^a(x_t) = \delta\left(\sum_{s=t+1}^{\infty} \delta^{s-t-1}\mathbb{E}[U_s(X_s)|x_t] + \frac{A(1-\delta^{T-t})}{1-\delta} + \delta^{T-t}R\right),$$

$$\underline{v}_t(x_t) = \underline{v}_t^e(x_t) = \underline{v}_t^a(x_t) = \underline{v}_t^p(x_t) = \delta\sum_{s=t+1}^{\infty} \delta^{s-t-1}\mathbb{E}[U_s(X_s)|x_t],$$

$$\overline{v}_t^p(x_t) = \overline{v}_t^a(x_t) + A,$$

$$\overline{v}_t^b(x_t) = \overline{v}_t^p(x_t) + U_t(x_t),$$

$$\underline{v}_t^b(x_t) = \underline{v}_t^p(x_t) + U_t(x_t).$$

(iii) *All of the composites in Propositions 1-6 that depend on Y_t, U_t, or the utility bounds listed in (ii) above also require x_t as an additional argument. $($We do not list them explicitly for brevity, as they are numerous.$)$*

Additionally, the inter-period updating of the v-score also needs to take state transitions into account. Specifically:

$$\pi^e_{t-1}(v^e_{t-1}, x_{t-1}) = \max_{v^b_t(x_t)\in[\underline{v}^b_t(x_t),\overline{v}^b_t(x_t)],\forall x_t\in supp X_t} \gamma\mathbb{E}\left[\pi^b_t(v^b_t(X_t), X_t)\middle| x_{t-1}\right], \tag{EC.8}$$

$$s.t.\ v^e_{t-1} = \delta\mathbb{E}[v^b_t(X_t)\big| x_{t-1}]. \tag{EC.9}$$

The resulting function $\pi^e_{t-1}(v^e_{t-1}, x_{t-1})$ is concave in v^e_{t-1}.

In addition, this setting enables us to ask how exactly the contract's terms change depending on the current state. For instance, one may expect the contract to offer easier terms (such as a higher tolerance for non-payment) to consumers who are in an unfavorable state, such as joblessness. Note that a state can be unfavorable for two reasons: (a) the distribution of income under the state can be bad (cash is tight when one is jobless), or (b) the probability of transitioning to another unfavorable state may be high (finding a new job might be tough). To see how the state affects the contract, we adopt the following simplifying assumption.

ASSUMPTION EC.2. *The process $\{X_t\}$ has only two possible states $(x_t \in \{1,2\})$. The distribution of income under state 2 dominates the distribution of income under state 1 in the sense of first-order stochastic dominance $(G_t(y_t \mid 2) \le G(y_t \mid 1), \forall y_t \in \mathbb{R})$. The probability of staying in a state exceeds the probability of transitioning out of that state $($denoting by p_{ijt} the probability of transitioning from state i to state j at time t, this condition is $p_{11t} > p_{12t}$ and $p_{22t} > p_{21t})$.*

For example, income with an exponential distribution $G_t(y_t \mid x_t) = 1 - \exp\{-\lambda(x_t)y_t\}$, where $\lambda(x_t)$ is increasing, satisfies both Assumptions EC.1 and EC.2.

PROPOSITION EC.6. *Let Assumptions EC.1 and EC.2 hold and the state and transition probabilities of $\{X_t\}$ be public knowledge.*

(i) *Denote by $\alpha^*_t(x_t, \kappa^\nu_t)$ the α_t that solves (25)–(26) of the main paper for a certain κ^ν_t, when the state of the chain is x_t. Then, $\alpha^*_t(1, \kappa^\nu_t) - U_t(1) \ge \alpha^*_t(2, \kappa^\nu_t) - U_t(2)$.*

(ii) *Denote by $v^{b*}_t(x_t; v^e_{t-1}, x_{t-1})$ the value of $v^b_t(x_t)$ that solves the maximization problem in Proposition EC.5 for a given v^e_{t-1} and x_{t-1}. Then, $\frac{\partial\pi^b_t(v,1)}{\partial v}\Big|_{v=v^{b*}_t(1;v^e_{t-1},x_{t-1})} = \frac{\partial\pi^b_t(v,1)}{\partial v}\Big|_{v=v^{b*}_t(2;v^e_{t-2},x_{t-1})}$, $\forall x_{t-1} \in \{1,2\}$, $v^e_{t-1} \in [\underline{v}^e_{t-1}(x_{t-1}), \overline{v}^e_{t-1}(x_{t-1})]$.*

(iii) *Let us denote the Arrow-Pratt coefficient of absolute risk aversion for a function $f(\cdot)$ at point x by $ARA_{f(\cdot)}(x)$. If $ARA_{\pi^b_t(\cdot,1)}(v - \underline{v}^b_t(1)) < ARA_{\pi^b_t(\cdot,2)}(v - \underline{v}^b_t(2))$, $\forall v$, then $v^{b*}_t(1; v^e_{t-1}, x_{t-1}) - \underline{v}^b_t(1) \ge v^{b*}_t(2; v^e_{t-1}, x_{t-1}) - \underline{v}^b_t(2)$.*

Recall that in the absence of payment (or for very low payments), the v-score of the consumer drifts downward. The first part of Proposition EC.6 establishes that this downward drift will be smaller for a consumer who is in the unfavorable income state. Essentially, the penalty for non-payment (or very small payment) will be smaller for a consumer in an unfavorable state. This is consistent with the aforementioned intuition, that the firm will be willing to give somewhat easier contract terms to a consumer in an unfavorable state.

The second part of the proposition speaks of the way the firm treats the consumer differently depending on the state transitions. The continuation value of the consumer at any point in time can be decomposed into two additive terms: (a) the value of autarchy (how much utility the consumer could get in the absence of the contract, which is now state dependent), and (b) the additional value generated to the consumer through the contract. This additional value just after state transition can be expressed as $v_t^{b*}(x_t; v_{t-1}^e, x_{t-1}) - \underline{v}_t^b(x_t)$. If the firm does not discriminate based on state transitions, we would have $v_t^{b*}(1; v_{t-1}^e, x_{t-1}) - \underline{v}_t^b(1) = v_t^{b*}(2; v_{t-1}^e, x_{t-1}) - \underline{v}_t^b(2)$. Intuitively, one may expect that the firm will provide a higher additional value to a consumer in an unfavorable state and a lower additional value to a consumer in a favorable state (so, $v_t^{b*}(1; v_{t-1}^e, x_{t-1}) - \underline{v}_t^b(1) > v_t^{b*}(2; v_{t-1}^e, x_{t-1}) - \underline{v}_t^b(2)$). Such a measure would mitigate the consumer's risk and enable their consumption smoothing. However, that is not necessarily true. Since the consumer's continuation utilities are already in the units of utility (not consumption), the consumer is actually indifferent about how her continuation utilities depend on the state transition; it is only the firm that benefits from it, and the way the firm discriminates between state transitions depends solely on the curvature of its own continuation utilities, as exemplified by the $\frac{\partial \pi_t^b(v,1)}{\partial v}\Big|_{v=v_t^{b*}(1; v_{t-1}^e, x_{t-1})} = \frac{\partial \pi_t^b(v,1)}{\partial v}\Big|_{v=v_t^{b*}(2; v_{t-2}^e, x_{t-1})}$ condition. Essentially, the firm gives the consumer continuation utilities such that its own marginal continuation utility at v_t^{b*} is the same under both state transitions.

The last part of the proposition states that if $ARA_{\pi_t^b(\cdot,1)}(v - \underline{v}_t^b(1)) \leq ARA_{\pi_t^b(\cdot,2)}(v - \underline{v}_t^b(2))$, then the firm provides more value to the consumer in an unfavorable state. This happens because the firm's continuation function—shifted so that it can be expressed as a function of the additional value provided to the consumer—is more concave in the favorable state if the above condition is satisfied. This is not an unreasonable condition: a more favorable state can also be more volatile due to a greater possibility for higher payments but also a greater penalty for low payments (because of the $\alpha_t^*(1, \kappa_t^\nu) - U_t(1) \geq \alpha_t^*(2, \kappa_t^\nu) - U_t(2)$ established in part (i) of the proposition).

EC.10.2. The Markov Process is the Consumer's Private Knowledge

It is quite conceivable that in practice, one may run into situations where $\{X_t\}$ is private knowledge of the consumer. For example, many jobs—particularly part-time ones in the developing markets—are ill-documented, causing difficulties for the firm in verifying the consumer's employment status. Similarly, illness may negatively impact the distribution of the consumer's income but may not be verifiable by the firm. Thus, here we will consider a variant of our model under Assumption EC.1, but where the current state of the Markov chain is the consumer's private knowledge.

Such situations will cause persistent information asymmetries, due to which the firm will be uncertain about the consumer's continuation utility under any particular contract. This, in turn, will disable the promised future utility approach because its key assumption is the common knowledge of the consumer's continuation utilities. Thus, tracking just the promised future utility will no longer be sufficient. This is a general property that is not unique to our model: Pavan et al. (2014) showed that whenever there is persistent private information, the optimal mechanism must depend on the entire belief process rather than any single state variable like continuation utility.

By following Fernandes and Phelan (2000), we can still recursively solve the problem by including additional so-called *threat-keeping* constraints. The essence of that work is as follows. When the current state is the consumer's private knowledge, the firm will ask the consumer to reveal their state. To keep the consumer honest, the firm will need to think not only about incentive compatibility in the current period but also about inter-period incentive compatibility:

that at no point can the consumer benefit from having lied in the past. This will complicate the state space, as the firm will now have to think not only about the consumer's continuation utility if the consumer has reported the current state honestly (which results in a promise-keeping constraint) but also about counterfactual continuation utilities of consumers who have dishonestly reported the state in the past (which result in threat-keeping constraints).

While this change in assumptions will affect much of our model only superficially, we highlight the two places where private information will make a stark difference.

The first of these occurs at the beginning of every period. There, in our main model, the consumer's v-score evolves according to a relatively simple deterministic equation: $\pi^e_{t-1}(v^e_{t-1}) = \gamma\pi^b_t(v^e_{t-1}/\delta)$. However, with private information, the consumer will need to report their state and based on that report may get different promised continuation utilities. Without loss of generality, we can enumerate the possible states of the Markov chain so that $x_t \in \{1, 2, ..., n_t\}$. We also denote $\mathbf{v}^e_t = (v^e_t(1), v^e_t(2), ..., v^e_t(n_t))$ the n_t-dimensional vector of all possible continuation utilities for the consumer, depending on the true state. Analogously, $\mathbf{v}^b_t(x_t) = (v^b_t(x_t, 1), v^b_t(x_t, 2), ..., v^e_t(x_t, n_t))$ is the vector of beginning-of-period continuation values of the consumer, when the consumer's report of the current state (but not necessarily the true state) was x_t. Then, the beginning-of-period problem can be written as

$$\pi^e_{t-1}(\mathbf{v}^e_{t-1}, x_{t-1}) = \max_{\mathbf{v}^b_t(x_t)} \gamma\mathbb{E}\left[\pi^b_t(\mathbf{v}^b_t(X_t), X_t)\middle| x_{t-1}\right], \tag{OBJ}$$

$$\text{s.t. } x_t \in \arg\max_{\hat{x}_t} v^b_t(\hat{x}_t, x_t),\ \forall x_t \in \{1, ..., n_t\}, \tag{IC}$$

$$v^e_{t-1}(x_{t-1}) = \delta\mathbb{E}[v^b_t(X_t, X_t)\big| x_{t-1}], \tag{PK}$$

$$v^e_{t-1}(\hat{x}_{t-1}) = \delta\mathbb{E}[v^b_t(X_t, X_t)\big| \hat{x}_{t-1}],\ \forall \hat{x}_{t-1} \in \{1, 2, ...n_{t-1}\}\backslash\{x_{t-1}\}. \tag{TK}$$

The firm needs to choose the continuation utilities for every possible report in order to maximize its continuation payoff. In addition to the usual incentive compatibility and promise-keeping constraints (IC and PK), we can see the aforementioned threat-keeping constraints (TK) here. The added complexity of the problem also manifests in the expanded state space because now the firm also needs to think about what the consumer's utility will be under each possible lie about the current state. This issue will propagate to the problem of determining the optimal suggested payment and reward function—the second modification to our problem, which now becomes

$$\pi^b_t(\mathbf{v}^b_t, x_t) = \max_{\mathbf{p}_t(\cdot)\in\mathcal{P}^{n_t}_t,\, \mathbf{v}^p_t(\cdot)\in\mathcal{V}^{n_t}_t} \quad \mathbb{E}\big[p_t(Y_t, x_t) + \pi^p_t(\mathbf{v}^p_t(\mathbf{p}_t(Y_t)), x_t)\big| x_t\big] \tag{OBJ2}$$

$$\text{s.t.} \quad p_t(y_t, \hat{x}_t) \in \underset{0\le p\le y_t}{\text{argmax}}\ u(y_t - p) + v^p_t(p, \hat{x}_t), \quad \forall y_t \in \mathcal{Y}_t, \hat{x}_t \in \{1, 2, ..., n_t\} \tag{IC2}$$

$$v^b_t(x_t, \hat{x}_t) = \mathbb{E}\big[u(Y_t - p_t(Y_t, x_t)) + v^p_t(p_t(Y_t, x_t), x_t)\big], \forall \hat{x}_t \in \{1, 2, ...n_t\}. \tag{PK2/TK2}$$

The firm cannot forget the possibility that the consumer might have misreported the income distribution at the beginning of the period (that $\hat{x}_t$ might not equal x_t), thus needs to know which payment decision is incentive compatible in each of those cases and which reward function to use. The result is that the firm now needs to decide on n_t different payment and reward functions in each period. This may seem surprising, because at first glance, it may seem that a single reward function suffices; after all, because the firm does not know the hidden distribution, it needs to give the consumer the same reward irrespective of her true type. However, the rewards are in terms of future reward functions (access to technology and ownership). While all the consumer types benefit equally from ownership and technology usage, they may get very different utility from the reward function. To illustrate why, consider a linear reward function that rewards the consumer at a high rate ($> \lambda_1$). The expected utility from having access to such a reward function is higher for a consumer who has a better distribution of income (in terms of first-order stochastic dominance).

So, to a certain extent, this approach works as it enables a recursive representation of the problem, which can then be solved numerically. However, the techniques used to solve (11) the analog of (OBJ2)–(TK2) above do not extend to (OBJ2)–(TK2). But with some simplifications that we discuss in our next extension, we can restore analytical tractability.

EC.10.3. The Markov Process is Private Information but Contains Only Employment Status

Arguably, the most important piece of information contained in the Markov process is the consumers' employment status, as it determines whether they can be reasonably expected to have any disposable income. A consumer in a state of joblessness will not have much disposable income, and we assume it to be zero in the model. Further, if the jobs are ill-documented, it may be difficult for the firm to verify the consumer's employment status, rendering it her private information. Similarly to the model of Fu and Krishna (2019), we consider a setup that focuses only on this piece of private information. For the same, we adopt Assumption EC.1 as well as the following one.

ASSUMPTION EC.3. *The process $\{X_t\}$ has only two possible states ($x_t \in \{0,1\}$). The state of the process is the consumer's private information. The consumer has no disposable income under state 0 (joblessness), i.e., Y_t conditional on $x_t = 0$ is 0 almost surely. A consumer in state 1 (employment) has a strictly positive income. The probability of staying in a state exceeds the probability of transitioning out of that state (denoting by p_{ijt} the probability of transitioning from state i to state j at time t, this condition is $p_{00t} > p_{01t}$ and $p_{11t} > p_{10t}$).*

The state transition part of the assumption can be thought of as a consumer who is employed is more likely to stay employed in the next period as well, and a consumer who is unemployed is likely to stay unemployed. Note that under this assumption, only one temptation to misreport exists: a consumer who is employed may want to report that she is unemployed (perhaps to take advantage of some tolerance for non-payment that is extended to jobless consumers). The opposite temptation does not exist: a jobless consumer who reports being employed will immediately get caught in the lie when she is unable to make any payment. With this in mind, under assumptions EC.1 and EC.3, the system (OBJ)–(TK) reduces to

$$\pi^e_{t-1}(\mathbf{v}^e_{t-1}, x_{t-1}) = \max_{\mathbf{v}^b_t(x_t)} \gamma \mathbb{E}\left[\pi^b_t(\mathbf{v}^b_t(X_t), X_t) \middle| x_{t-1}\right], \tag{EC.10}$$

$$\text{s.t. } v^b_t(1,1) \geq v^b_t(0,1), \tag{EC.11}$$

$$v^e_{t-1}(x_{t-1}) = \delta \mathbb{E}[v^b_t(X_t, X_t) \big| x_{t-1}], \tag{EC.12}$$

$$v^e_{t-1}(\hat{x}_{t-1}) = \delta \mathbb{E}[v^b_t(X_t, X_t) \big| \hat{x}_{t-1}], \text{ for } \hat{x}_{t-1} = |x_{t-1} - 1|, \tag{EC.13}$$

The system (OBJ2)–(TK2) reduces to the problem of designing payment and reward functions only for the consumer in state 1 (who has income) and is given by

$$\pi^b_t(\mathbf{v}^b_t, 1) = \max_{p_t(\cdot)\in\mathcal{P}_t,\, v^p_t(\cdot)\in\mathcal{V}_t} \quad \mathbb{E}\left[p_t(Y_t) + \pi^p_t((\underline{v}^p_t(0), v^p_t(p_t(Y_t))), 1) \mid x_t = 1\right] \tag{EC.14}$$

$$\text{s.t.} \quad p_t(y_t) \in \operatorname*{argmax}_{0 \leq p \leq y_t} u(y_t - p) + v^p_t(p), \quad \forall\, y_t \in \mathcal{Y}_t, \tag{EC.15}$$

$$v^b_t(x_t) = \mathbb{E}\left[u(Y_t - p_t(Y_t, x_t)) + v^p_t(p_t(Y_t)) \mid x_t = 1\right]. \tag{EC.16}$$

Note that, unlike (OBJ2)–(TK2), this problem is similar to (9)–(11) from the paper. Thus, the novelty of this setting compared to the main paper can be contained to the problem (EC.10)–(EC.13) and the consequences stemming from it. This results in a tractable setting, for which we can condense the key findings in the following proposition.

PROPOSITION EC.7. *Let Assumptions EC.1 and EC.3 hold and the transition probabilities of $\{X_t\}$ be public knowledge. Then:*

(i) *The utility of autarchy and the bounds for the promised future utility are state-dependent and are given in parts (i) and (ii) of Proposition EC.5. The promised threat utilities have the same bounds.*

(ii) *The constraints* (EC.11)–(EC.13) *characterize a unique point.*

(iii) *All of the firm's continuation utility functions are concave in promised future and threat utilities.*

(iv) *The contract with the firm never provides more value to an unemployed consumer* (*beyond that of autarchy*) *than to an employed one.*

The first part of the proposition ensures that the problem has well-defined bounds. The second part of the proposition shows that the firm does not really have a choice in the problem (EC.10)–(EC.13), as the constraints reduce the feasible set to just a point. This happens for two reasons. First, the choice of $v_t^b(0,0)$ uniquely determines $v_t^b(0,1)$. Intuitively, this happens because a consumer without a job and a consumer with a job but who lied differ only in that the latter will get additional utility from consuming income and more favorable transition probabilities in the next period (as she is more likely to be employed in next period as well). Both these are transient benefits that will give her some utility now but will no longer be relevant after the next state transition occurs. Second, the promise-keeping and threat-keeping constraints are two line equations of $v_t^b(0,1)$ and $v_t^b(1,1)$ with different slopes, thus they uniquely determine the pair $(v_t^b(0,1), v_t^b(1,1))$. The third part of the proposition uses the results of (ii) to ensure that the objective functions are well-behaved; as a result, the key methods used in the paper, which rely on the concavity of continuation payoff functions being inherited through the periods, also apply here. The last part of the proposition stands in contrast to Proposition EC.6, parts (ii) and (iii). Unlike there, in this setting, the firm does not offer any benefit to an unemployed consumer. The reason for this is straightforward: such benefits would be exploitable by an employed but dishonest consumer.

EC.11. Proofs of Mathematical Results in the Main Paper and the Electronic Companion

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 $\pi(x)-\pi(v)\leq k(x-v)$ 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 the 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 $F^{-1*}(y):=\sup\{x\in\mathrm{dom}F|F(x)\leq y\}$.
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 EC.1–EC.2.

Proof of Proposition 1. The proposition follows from the geometric construction shown in Figure EC.3. The firm in our model has the ability to terminate probabilistically. By randomizing between confiscation (transferring ownership) and continuation, the firm can obtain all payoffs that lie on any line segment between $(\underline{v}_t^a,\mathcal{C}_t)$ $((\overline{v}_t^a,\mathcal{O}_t))$ and any point on the graph of $\pi_t^e(\cdot)$, and by randomizing between confiscation and ownership 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 EC.3(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 dotted line 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 (7), and illustrated in Figure EC.3(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 EC.3(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 $\pi_t^e(\cdot)$. The concavity of the payoff function π_t^a in v_t^a follows by construction. Moreover, simple inspection of Figure EC.3(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.

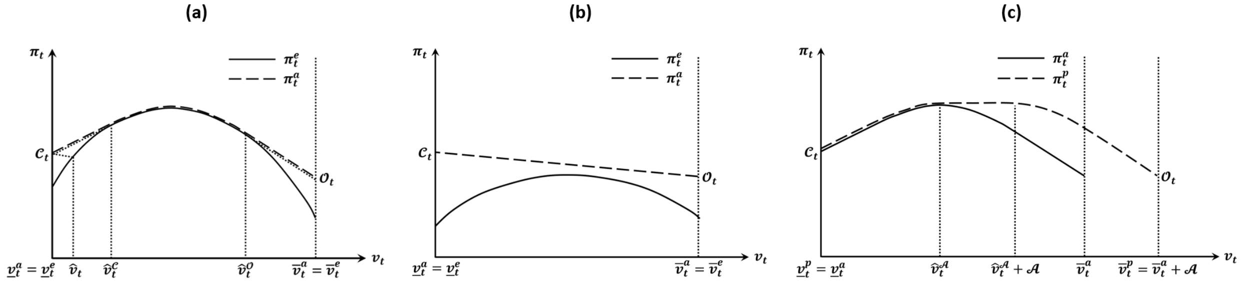

Figure EC.3 **Construction of π_t^a is shown in (a) when (7) is not satisfied, and in (b) when (7) is satisfied; and construction of π_t^p is shown in (c).**

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 $(\hat{v}_t^{\mathcal{C}}, v_t^e(\hat{v}_t^{\mathcal{C}}))$ (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

$$\pi_t^a(v_t^a) = \begin{cases} \mathcal{C}_t + s_t^{\mathcal{C}}(v_t^a - \underline{v}_t^a) & \text{if } \underline{v}_t^a \le v_t^a < \hat{v}_t^{\mathcal{C}}, \\ \pi_t^e(v_t^a) & \text{if } \hat{v}_t^{\mathcal{C}} \le v_t^a \le \hat{v}_t^{\mathcal{O}}, \\ \pi_t^e(\hat{v}_t^{\mathcal{O}}) + s_t^{\mathcal{O}}(v_t^a - \hat{v}_t^{\mathcal{O}}) & \text{if } \hat{v}_t^{\mathcal{O}} < v_t^a \le \overline{v}_t^a. \end{cases} \qquad \square \tag{EC.17}$$

Proof of Proposition 2. Substituting $v_t^a = -f_t A + v_t^p$ into (8) reduces it to $\max_{f_t \in [0,1]} \pi_t^a(v_t^p - f_t A)$. Geometrically, as shown in Figure EC.3(c), we obtain π_t^p by laterally shifting π_t^a by A units starting from its maximum value (achieved at $\hat{v}_t^{\mathcal{A}}$). Using the property established in Proposition 1 that π_t^a is a concave function gives the expression for f_t^* in the proposition statement. Inserting the expression for f_t^* into $v_t^a = -f_t A + v_t^p$ yields the optimal continuation value v_t^{a*} in the proposition statement. Lastly, inserting v_t^{a*} into the objective function of (8) yields $\pi_t^p(v_t^p) = \pi_t^a(v_t^p - \min\left\{ [v_t^p - \hat{v}_t^{\mathcal{A}}]^+, A \right\})$, 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 EC.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 $\mathbb{E}[h(X)] \ge \mathbb{E}[g(X)]$. If g and h are right-differentiable and $h'^+(x) \le g'^+(x)$, $\forall x \in \mathcal{X} \setminus \{\sup \mathcal{X}\}$ 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 $h'^-(x) \le g'^-(x)$, $\forall x \in \mathcal{X} \setminus \{\inf \mathcal{X}\}$ 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 $F_{h(X)}(x) = P(h(X) \le x) = \int_\xi^{h^{-1*}(x)} dF(u) = F(h^{-1*}(x))$, and analogously $F_{g(X)}(x) = F(g^{-1*}(x))$. Then, $(h(X) \succeq_{SSD} g(X)) \Leftrightarrow (\int_{-\infty}^x F_{h(X)}(u) - F_{g(X)}(u) du \le 0, \forall x \in \mathbb{R})$. Notice that $\bar{g}(\xi) \le \bar{h}(\xi)$, otherwise $h'^+(x) \le g'^+(x)$ or $h'^-(x) \le g'^-(x)$ would imply $\mathbb{E}[h(X)] > \mathbb{E}[g(X)]$. If $\bar{g}(\psi) \le \bar{h}(\psi)$ or $\bar{g}(\xi) = \bar{h}(\xi)$, then as $h'^+(x) \le g'^+(x)$ or $h'^-(x) \le g'^-(x)$ we have $g(x) \le h(x), \forall x \in \mathcal{X}$, thus $F_{h(X)}(x) \le F_{g(X)}(x), \forall x \in \mathcal{X}$, 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 $h'^{+}(x) \leq g'^{+}(x)$ or $h'^{-}(x) \leq g'^{-}(x)$, the set of all points on which these functions are equal is a closed interval ($\exists \xi^*, \psi^*$ such that $\xi < \xi^* \leq \psi^* < \psi$, where $g(x) = h(x)$, $\forall x \in [\xi^*, \psi^*]$, and $\bar{g}(x) \neq \bar{h}(x), \forall x \in [\xi, \xi^*) \cup (\psi^*, \psi]$). Thus, the cdf-s $F_{h(X)}$ and $F_{g(X)}$ have the single-crossing property where $F(h^{-1*}(x)) \leq F(g^{-1*}(x)), \forall x < g(\xi^*)$ and $F(h^{-1*}(x)) \geq F(g^{-1*}(x)), \forall x \geq g(\xi^*)$. 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 $u \leq g(\xi^*)$ and weakly positive for $u \geq g(\xi^*)$. Hence, $I(x)$ is non-increasing up to $x = g(\xi^*)$ and non-decreasing afterwards, which combined with $I(\bar{g}(\xi)) = 0$ and $I(\bar{g}(\psi)) \leq 0$ implies that $I(x) \leq 0, \forall x \in \{g(y) | y \in \mathcal{X}\}$, completing the proof. $\square$

Geometric intuition for Proposition 3. Before we provide a formal proof for Proposition 3, we will pictorially demonstrate why the shapes of the optimal reward and payment functions are as stated in Proposition 3. We can see from (10) 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 (12). Figure EC.4(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 EC.4(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 EC.4(b). 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 $y \geq \hat{p}_{0,t} + \kappa$. The space between the two functions $\underline{p}_t(y)$ and $\overline{p}_t(y)$ 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 EC.4(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, $\pi_t^{p'}(v_t^{p*}(p))$ is decreasing in p (plotted on the left-hand side of Figure EC.4(b)), and so the derivative of the payoff function, $1 + \pi_t^{p'}(\alpha_t + \lambda_1 p)\lambda_1$, is also decreasing in p. Therefore, the optimal $p_t^*(y)$, as demonstrated in Figure EC.4(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 EC.4(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 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 EC.4(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 (11), we have $\mathbb{E}\check{v}_t^p(p_t^*(y)) = \mathbb{E}v_t^{p*}(p_t^*(y))$. Figure EC.4(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

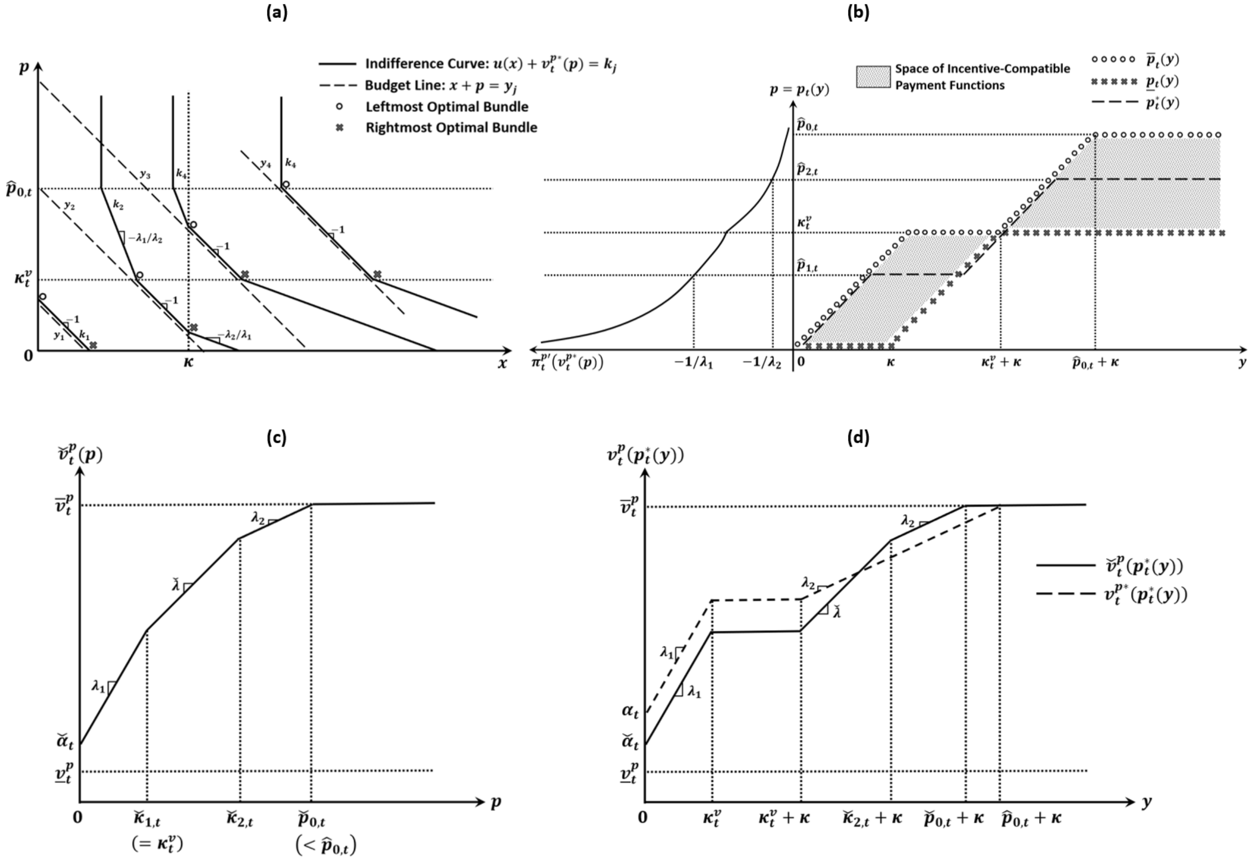

Figure EC.4 Construction of optimal payment and reward functions: (a) Indifference curves $u(x) + v_t^{p*}(p) = k_j$, along with the tangential budget lines $x + p = y_j$, shown in the (x, p) plane for $j \in \{1, \dots, 4\}$; (b) on the right-hand side is the space of incentive-compatible payment functions 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.

prefers the reward function that results in a lower payoff variance (see (9)), which in this case is v_t^{p*}. This logic continues to hold more generally, as will be evident in the proof next. 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.

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 (9)–(11) can be be restricted to $\mathcal{P}_t \times \bar{\mathcal{V}}_t$ without loss of optimality, where $\bar{\mathcal{V}}_t := \{v_t^{p*}(p; \alpha_t, \kappa_t^v) | \alpha_t \geq \underline{v}_t^p, \kappa_t^v \in \bar{\mathbb{R}}_+\}$, and $v_t^{p*}(p; \alpha_t, \kappa_t^v)$ is given by (12). The proof is based on showing that for any pair $(p_t(\cdot), v_t^p(\cdot))$ that solves (9)–(11), we can construct $v_t^{p*}(\cdot) \in \bar{\mathcal{V}}_t$ such that $(p_t(\cdot), v_t^{p*}(\cdot))$ also solves (9)–(11). Assume $(p_t(\cdot), v_t^p(\cdot))$ solves (9)–(11). Then, from (10) we have $p_t(y) \in \underset{0 \leq p \leq y}{\operatorname{argmax}}\ u(y-p) + v_t^p(p), \forall y \in \mathcal{Y}_t$, which implies (since $p_t(y)$ is a local maximum)

$$\frac{d}{d^- p}\left(v_t^p(p) + u(y-p)\right)\Big|_{p = p_t(y)} \geq 0, \quad \forall y \in \mathcal{Y}_t \text{ such that } p_t(y) \neq 0, \tag{EC.18}$$

$$\frac{d}{d^+ p}\left(v_t^p(p) + u(y-p)\right)\Big|_{p = p_t(y)} \leq 0, \quad \forall y \in \mathcal{Y}_t \text{ such that } p_t(y) \neq y. \tag{EC.19}$$

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 $\kappa_t^{v*} := \inf\{p | 0 \leq p \leq \hat{\hat{p}}_t, \exists r_p > p \text{ such that } v_t^p(x) < v_t^p(p) + \lambda_1(x - p), \forall x \in (p, r_p)\}$. Since both one-sided derivatives of $u(y)$ are either

λ_1 or λ_2, from (EC.18) it follows that $\frac{d}{d^-p}v_t^p(p) \geq \lambda_2$, for all p such that $0 < p \leq \hat{\hat{p}}_t$. 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 ($v_t^p(y) \leq \bar{v}_t^p, \forall y \in \mathcal{Y}_t$) but consumption utility is not ($\lim_{y\to\infty} u(y) = \infty$).

First, since from the definition of κ_t^{v*} we have $\frac{d}{d^+p}v_t^p(p) \geq \lambda_1, \forall p \in [0, \kappa_t^{v*})$, for incomes below κ_t^{v*} any incentive compatible payment policy will leave at most κ for consumption. Thus, $\min\{\hat{\hat{p}}_t, [y-\kappa]^+\} \leq p_t(y) \leq \min\{\hat{\hat{p}}_t, y\}, \forall y \leq \kappa_t^{v*}$. 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 (EC.19) implies that $\frac{d}{d^+p}v_t^p(p) \leq \lambda_1$ 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 $\frac{d}{d^+p}v_t^p(p) \leq \lambda_1$ for all $p > \kappa_t^{v*}$, so $\min\{\hat{\hat{p}}_t, [y-\kappa]^+\} \leq p_t(y) \leq \min\{\kappa_t^{v*}, \hat{\hat{p}}_t\}], \forall y \in (\kappa_t^{v*}, \kappa_t^{v*}+\kappa]$, 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 $\frac{d}{d^+p}v_t^p(p) \leq \lambda_1$ 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 $p_t(y) \in [\min\{\kappa_t^{v*}, \hat{\hat{p}}_t\}, \min\{\hat{\hat{p}}_t, y-\kappa\}], \forall y \geq \kappa_t^{v*}$. Noticing that $\hat{\hat{p}}_t \geq \kappa_t^{v*}$ 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 (12). 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 (10). (This region is the space of incentive-compatible payment functions illustrated Figure EC.4(b), when $\kappa_t^v = \kappa_t^{v*}$; it can also be written using the notation of footnote 5 as $\underline{p}_t(y) \leq \acute{p}(y) \leq \overline{p}_t(y)$.) 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 $\forall y \in \mathcal{Y}_t, p \in [\min\{[y-\kappa]^+, \kappa_t^{v*}\}, \max\{y-\kappa, \kappa_t^{v*}\}]:$

$$v_t^{p*}(p;\alpha_t,\kappa_t^{v*}) + u(y-p) = \begin{cases} \lambda_1 y & \text{if } y \leq \kappa + \kappa_t^{v*}, \\ \lambda_1(\kappa_t^{v*}+\kappa) + \lambda_2(y-\kappa_t^{v*}-\kappa) & \text{otherwise.} \end{cases} \tag{EC.20}$$

Taking a left derivative of (EC.20) with respect to p yields that (EC.18) holds for $v_t^p(\cdot) = v_t^{p*}(\cdot;\alpha_t,\kappa_t^{v*})$ with equality. Hence, from (EC.18) 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 (11) 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 EC.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 (9), while both pairs satisfy the constraints (10)–(11), 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 (9)–(11), 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, analogously to Part 2 of Proposition EC.1. The intuition 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 (10)–(11). 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 (9)–(11). 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 $y \le \kappa + \kappa_t^v$ 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. Recall from the proposition statement that

$$\hat{p}_{1,t}(\alpha_t,\kappa_t^v) = \min\left\{[\Lambda_{1,t} - \alpha_t]^+/\lambda_1, \kappa_t^v, \hat{p}_{0,t}(\alpha_t,\kappa_t^v)\right\}, \tag{EC.21}$$

$$\hat{p}_{2,t}(\alpha_t,\kappa_t^v) = \max\left\{(\Lambda_{2,t} - \alpha_t - (\lambda_1 - \lambda_2)\kappa_t^v)/\lambda_2, \min\left\{\kappa_t^v, \hat{p}_{0,t}(\alpha_t,\kappa_t^v)\right\}\right\}, \tag{EC.22}$$

and $\Lambda_{j,t} = \max\left\{\underline{v}_t^p, \sup\left\{v \in [\underline{v}_t^p, \overline{v}_t^p] \mid \pi_t^{p'}(v) \ge -1/\lambda_j\right\}\right\}$, for $j \in \{1,2\}$. Let $\hat{p}_{2,t}$ be as given above, and define

$$p_t^\dagger(y;\alpha_t;\kappa_t^v) := \begin{cases} p_t(y) & \text{if } y \le \kappa + \kappa_t^v, \\ y - \kappa & \text{if } \kappa_t^v + \kappa \le y \le \hat{p}_{2,t}(\alpha_t,\kappa_t^v) + \kappa, \\ \hat{p}_{2,t}(\alpha_t,\kappa_t^v) & \text{if } y \ge \hat{p}_{2,t}(\alpha_t,\kappa_t^v) + \kappa. \end{cases} \tag{EC.23}$$

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 (10)–(11).

Now, to show 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 (10). Now, we shall establish that for any $y \in \mathcal{Y}_t$,

$$\pi_t^p(v_t^{p*}(p_t^\dagger(y;\alpha_t;\kappa_t^v);\alpha_t,\kappa_t^v)) + p_t^\dagger(y;\alpha_t;\kappa_t^v) \ge \pi_t^p(v_t^{p*}(p_t(y);\alpha_t,\kappa_t^v)) + p_t(y). \tag{EC.24}$$

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 (EC.23), so (EC.24) 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 (EC.23) we have $y - \kappa = p_t^\dagger(y;\alpha_t;\kappa_t^v) \ge p_t(y) \ge \kappa_t^v$, from which (EC.24) 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 $y \in [\hat{p}_{2,t}(\alpha_t,\kappa_t^v) + \kappa, \infty) \cap \mathcal{Y}_t$, (EC.24) 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 (EC.24) 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 (9).

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 (9)–(11). So, let $p_t^*(y;\alpha_t,\kappa_t^v)$ be the one given by (13). 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 (10)–(11) (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 $v_t^{p*}(p;\alpha_t;\kappa_t)$. $\square$

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 $v_t^b - \omega_t + \lambda_1 p \in [\underline{v}_t^p, \bar{v}_t^p]$.

Part (i): if $v_t^b \leq \hat{v}_t^b$ 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 (14)–(15) 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 (9)–(11), 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 (11) still binds). Then, combining incentive compatibility (10) with (EC.20) 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 (13) 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 (10). 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 (11) 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 $\frac{d}{d^+p}v_t^{p\dagger\dagger}(p;\alpha_t,\kappa_t^v) \leq \frac{d}{d^+p}v^{p*}(p;\alpha_t,\kappa_t^v), \forall y \in \mathcal{Y}_t$ (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, $\frac{d}{d^+y}\left(v_t^{p\dagger\dagger}(p_t^*(y;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)\right) \leq \frac{d}{d^+y}\left(v_t^{p*}(p_t^*(y;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)\right), \forall y \in \mathcal{Y}_t$, thus from Lemma EC.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) $\Leftrightarrow$ ($\mathbb{E}[q(X)] \geq \mathbb{E}[q(Y)], \forall$ concave $q(\cdot)$) (Rothschild and Stiglitz 1970), from concavity of $\pi_t^p(v_t^p)$ we have $\mathbb{E}[\pi_t^p(v_t^{p\dagger\dagger}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)) + p_t^*(Y_t;\alpha_t,\kappa_t)] \geq \mathbb{E}[\pi_t^p(v_t^{p\dagger\dagger}(p_t^*(Y_t;\alpha_t,\kappa_t^v);\alpha_t,\kappa_t^v)) + p_t^*(Y_t;\alpha_t,\kappa_t^v)]$, 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 (9), while both pairs satisfy the constraints (10)–(11), 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

$$\pi_t^b(v_t^b)=\max_{\omega_t\in[\underline{\omega}_t,\bar{\omega}_t]}\tilde{\pi}_t^b(v_t^b,\omega_t),\quad\text{where}\tag{EC.25}$$

$$\tilde{\pi}_t^b(v_t^b,\omega_t)=\mathbb{E}\big[p_t^*(Y_t\,;\,v_t^b-\omega_t,\,\kappa_t^v(\omega_t))+\pi_t^p(v_t^{p*}(p_t^*(Y_t\,;\,v_t^b-\omega_t,\,\kappa_t^v(\omega_t))\,;\,v_t^b-\omega_t,\,\kappa_t^v(\omega_t)))\big]\tag{EC.26}$$

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 (EC.25)–(EC.26), $\kappa_t^v=\hat{p}_{1,t}(\alpha_t,\kappa_t^v)$ 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

$$p_t^{\kappa*}(y;\alpha_t,\kappa_t):=\begin{cases}y & \text{for}\quad 0\le y<\kappa_t^v,\\ \kappa_t^v & \text{for}\quad \kappa_t^v\le y<\kappa_t^v+\kappa,\\ y-\kappa & \text{for}\quad \kappa_t^v+\kappa\le y<\hat{p}_{2,t}(\alpha_t,\kappa_t^v)+\kappa,\\ \hat{p}_{2,t}(\alpha_t,\kappa_t^v) & \text{for}\quad y\ge\hat{p}_{2,t}(\alpha_t,\kappa_t^v)+\kappa.\end{cases}\tag{EC.27}$$

Putting these changes together, we will solve the following modified version of (EC.25)–(EC.26):

$$\pi_t^{b*}(v_t^b)=\max_{\omega_t\in[\underline{\omega}_t,\bar{\omega}_t]}\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t),\quad\text{where}\tag{EC.28}$$

$$\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)=\mathbb{E}\big[p_t^{\kappa*}(Y_t\,;\,v_t^b-\omega_t,\,\kappa_t^v(\omega_t))+\pi_t^{p\epsilon}(v_t^{p*}(p_t^{\kappa*}(Y_t\,;\,v_t^b-\omega_t,\,\kappa_t^v(\omega_t))\,;\,v_t^b-\omega_t,\,\kappa_t^v(\omega_t)))\big].\tag{EC.29}$$

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 (EC.22) 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 (EC.29) we obtain

$$\begin{aligned}\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)=&\int_0^{\kappa_t^v(\omega_t)}\left[\pi_t^{p\epsilon}(v_t^b-\omega_t+\lambda_1 y)+y\right]dG_t(y)\\&+\int_{\kappa_t^v(\omega_t)}^{\kappa_t^v(\omega_t)+\kappa}\left[\pi_t^{p\epsilon}(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))+\kappa_t^v(\omega_t)\right]dG_t(y)\\&+\int_{\kappa_t^v(\omega_t)+\kappa}^{p_{2,t}(v_t^b,\omega_t)+\kappa}\left[\pi_t^{p\epsilon}(v_t^b-\omega_t+(\lambda_1-\lambda_2)\kappa_t^v(\omega_t)+\lambda_2(y-\kappa))+y-\kappa\right]dG_t(y)\\&+\int_{p_{2,t}(v_t^b,\omega_t)+\kappa}^{\infty}\left[\pi_t^{p\epsilon}(v_t^b-\omega_t+(\lambda_1-\lambda_2)\kappa_t^v(\omega_t)+\lambda_2 p_{2,t}(v_t^b,\omega_t))+p_{2,t}(v_t^b,\omega_t)\right]dG_t(y).\end{aligned}\tag{EC.30}$$

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

$$\begin{aligned}\frac{\partial\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial v_t^b}=&\int_0^{\kappa_t^v(\omega_t)}(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1 y)dG_t(y)\\&+(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))]\\&+\int_{\kappa_t^v(\omega_t)+\kappa}^{p_{2,t}(v_t^b,\omega_t)+\kappa}(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2(y-\kappa))dG_t-\frac{1}{\lambda_2}\bar{G}_t(p_{2,t}(v_t^b,\omega_t)+\kappa),\end{aligned}\tag{EC.31}$$

$$\frac{\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial(v_t^b)^2}=\int_0^{\kappa_t^v(\omega_t)}(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1 y)dG_t(y)\tag{EC.32}$$

$$+(\pi_t^{p\epsilon})''(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))]\tag{EC.33}$$

$$+\int_{\kappa_t^v(\omega_t)+\kappa}^{p_{2,t}(v_t^b,\omega_t)+\kappa}(\pi_t^{p\epsilon})''(v_t^b-h_t(\omega_t)+\lambda_2(y-\kappa))dG_t(y)\tag{EC.34}$$

$$+\frac{1}{\lambda_2}g_t(p_{2,t}(v_t^b,\omega_t)+\kappa)\frac{\partial p_{2,t}(v_t^b,\omega_t)}{\partial v_t^b}\tag{EC.35}$$

$$+(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2 p_{2,t}(v_t^b,\omega_t))g_t(p_{2,t}(v_t^b,\omega_t)+\kappa)\frac{\partial p_{2,t}(v_t^b,\omega_t)}{\partial v_t^b}\leq 0.\tag{EC.36}$$

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

$$\begin{aligned}\frac{\partial\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial\omega_t}=&-\int_0^{\kappa_t^v(\omega_t)}[(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1 y)-(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))]dG_t(y)\\&+[\lambda_1(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))+1]\kappa_t^{v\prime}(\omega_t)[G_t(\kappa+\kappa_t^v(\omega_t))-G_t(\kappa_t^v(\omega_t))]\\&+h_t'(\omega_t)\Bigg[(\pi_t^{p\epsilon})'(v_t^b-\omega_t+\lambda_1\kappa_t^v(\omega_t))\bar{G}_t(\kappa+\kappa_t^v(\omega_t))+\frac{1}{\lambda_2}\bar{G}_t(p_{2,t}(v_t^b,\omega_t)+\kappa)\\&-\int_{\kappa+\kappa_t^v(\omega_t)}^{p_{2,t}(v_t^b,\omega_t)+\kappa}(\pi_t^{p\epsilon})'(v_t^b-h_t(\omega_t)+\lambda_2(y-\kappa))dG_t(y)\Bigg].\end{aligned}\tag{EC.37}$$

Before we delve into determining the rest of the second-order partial derivatives, we introduce some additional notation for recurring expressions. First, denote the term in square brackets in the last two rows of (EC.37)—the one that multiples $h_t'(\omega_t)$—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. Second, denote

$M_t(z) := (\bar{G}_t(z) - \bar{G}_t(z+\kappa))/((\lambda_1 - \lambda_2)\bar{G}_t(z+\kappa))$. Note that our assumption that $G_t(\cdot)$ has a decreasing hazard rate implies that $M_t(\cdot)$ is decreasing. Finally, denote

$$\begin{aligned}
K_1 :=& \int_0^{\kappa_t^v(\omega_t)} (\pi_t^{p\epsilon})''(v_t^b - \omega_t + \lambda_1 y)dG_t(y),\\
K_2 :=&[\lambda_1(\pi_t^{p\epsilon})'(v_t^b - \omega_t + \lambda_1\kappa_t^v(\omega_t)) + 1]\frac{\partial}{\partial\omega_t}M_t(\kappa_t^v(\omega_t)),\\
K_3 :=& \int_{\kappa+\kappa_t^v(\omega_t)}^{p_{2,t}(v_t^b,\omega_t)+\kappa} (\pi_t^{p\epsilon})''(v_t^b - h_t(\omega_t) + \lambda_2(y-\kappa))dG_t(y),\\
K_4 :=&h_t''(\omega_t)\Upsilon,\\
K_5 :=&(\pi_t^{p\epsilon})''(v_t^b - \omega_t + \lambda_1\kappa_t^v(\omega_t))[G_t(\kappa+\kappa_t^v(\omega_t)) - G_t(\kappa_t^v(\omega_t))],\\
K_6 :=& \left(\frac{1}{\lambda_2} + (\pi_t^{p\epsilon})'(v_t^b - h_t(\omega_t) + \lambda_2 p_{2,t}(v_t^b,\omega_t))\right) g_t(p_{2,t}(v_t^b,\omega_t)+\kappa)\frac{\partial p_{2,t}(v_t^b,\omega_t)}{\partial v_t^b}.
\end{aligned} \tag{EC.38}$$

These expressions contain some notation that is not ubiquitous in the paper; so, for clarity, recall that Υ and M_t are the auxiliary objects that were just defined in the last paragraph, that $p_{2,t}(v_t^b,\omega_t) = (\Lambda_{2,t} - v_t^b + \omega_t - (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t))/\lambda_2$ (as was introduced in (EC.30) and defined immediately preceding that equation), and that g_t is the pdf of the consumer's disposable income distribution.

We proceed to show that $K_i \leq 0, \forall i \in \{1,2,\ldots,6\}$. For K_1 and K_3 this follows directly from the concavity of $\pi_t^{p\epsilon}$ which ensures the integrand in the expressions is negative. Looking at the expression for K_2, as $(\pi_t^{p\epsilon})'(v_t^b - \omega_t + \lambda_1\kappa_t^v(\omega_t)) \geq -1/\lambda_1$ (which follows from (EC.27), specifically, how $\hat{p}_{2,t}(\alpha_t,\kappa_t^v)$ is defined via the slope of π_t^p in Proposition 3), the term in square brackets is positive, while $\partial M_t(\kappa_t^v(\omega_t))/\partial\omega_t$ is negative as M_t and κ_t^v are decreasing and increasing functions respectively. Thus, $K_2 \leq 0$. $K_4 \leq 0$ follows from $\Upsilon \geq 0$ (established in the paragraph preceding (EC.38)) and concavity of $h_t(\omega_t)$ (from $h_t'(\omega_t) = -G(\kappa+\kappa_t^v(\omega_t))/\bar{G}(\kappa+\kappa_t^v(\omega_t))$, we have $h_t''(\omega_t) = -(\kappa_t^v)'(\omega_t)G'(\kappa+\kappa_t^v(\omega_t))/(\bar{G}(\kappa+\kappa_t^v(\omega_t)))^2 < 0$). As $\pi_t^{p\epsilon}$ is concave and G_t is an increasing function, we have that $K_5 \leq 0$. Lastly, K_6 is a term that already appeared in (EC.36), and as was the casd there, $K_6 \leq 0$ follows from $(\pi_t^{p\epsilon})'(v_t^b - h_t(\omega_t) + \lambda_2 p_{2,t}(v_t^b,\omega_t)) \geq -1/\lambda_2$ and $\partial p_{2,t}(v_t^b,\omega_t)/\partial v_t^b = -1/\lambda_2$.

Using this notation, we can write the second partial derivatives of $\tilde{\pi}_t^{b\epsilon}$ as

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

We will use these derivatives to establish joint concavity by examining the Hessian matrix. 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, $\frac{\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial\omega_t^2} \leq 0$ and $\frac{\partial^2\tilde{\pi}_t^{b\epsilon}(v_t^b,\omega_t)}{\partial(v_t^b)^2} \leq 0$, which is already shown above), while the determinant of the Hessian is weakly positive. Calculating the determinant yields

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

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 $v_t^b \in [\hat{v}_t^b, \bar{v}_t^b]$. 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 $v_t^b \in [\hat{v}_t^b, \bar{v}_t^b]$:

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

From (EC.22) 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 $\hat{\hat{v}}_t^b \in [\hat{v}_t^b, \bar{v}_t^b]$ 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 $v_t^b \in [\hat{\hat{v}}_t^b, \bar{v}_t^b]$. Thus, on $v_t^b \in (\hat{v}_t^b, \hat{\hat{v}}_t^b)$,

$$\begin{aligned}
\frac{d\pi_t^{b*}(v_t^b)}{dv_t^b} = & (\pi_t^{p\epsilon})'(v_t^b - \underline{\omega}_t) G_t(\kappa) + \int_\kappa^{(\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa} (\pi_t^{p\epsilon})'(v_t^b - \underline{\omega}_t + \lambda_2(y - \kappa))\, dG_t(y) \\
& - \frac{1}{\lambda_2} \bar{G}_t((\Lambda_{2,t} - v_t^b + \underline{\omega}_t)/\lambda_2 + \kappa),
\end{aligned} \tag{EC.39}$$

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

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 $[\hat{\hat{v}}_t^b, \bar{v}_t^b]$, 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 $[\hat{v}_t^b, \bar{v}_t^b]$ follows from

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

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 (EC.31) we have that

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

while from (EC.39) it follows that

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

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 $\tilde{\pi}_t^{b\epsilon X}$ 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

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

where all K_j-s are given by (EC.38). Since $\alpha_t^*(v_t^b) = v_t^b - \omega_t^*(v_t^b)$, at those points we also have

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

From $h_t(\omega_t) = \omega_t - (\lambda_1 - \lambda_2)\kappa_t^v(\omega_t)$, using that $\kappa_t^v(\omega_t)$ is defined as an inverse of the right-hand side of (16) as a function of κ_t^v and applying the Inverse Function Theorem yields $h_t'(\omega_t) = 1 - 1/\bar{G}_t(\kappa + \kappa_t^v(\omega_t)) < 0$. Then, since all K_j-s are weakly negative, $h_t'(\omega_t)$ is negative 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 $\omega_t^*(v_t^b)$ 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). $\hat{p}_{1,t}^*(v_t^b) = \kappa_t^{v*}(v_t^b)$ is established in the proof of Proposition 4, part (i), after which $\hat{p}_{1,t}^*(v_t^b)$ exhibiting a $\cap$–pattern in v_t^b follows from part (ii) of this proposition.

Part (iv). 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

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

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 $d\hat{p}_{2,t}^*(v_t^b)/dv_t^b \leq 0$.

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 (v). 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 $\cap$-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 $p_{2,t}^*(v_t^b) - p_{1,t}^*(v_t^b)$. $\square$

Proof of Proposition 6. The result follows directly from noting that $\pi_T^a(v_T^a)$ is linear, thus also concave, and applying Propositions 2–4. $\square$

Proof of Proposition 7. *Part 1: problem reduction.* For the problem reduction, it is sufficient to focus on the following part of (18):

$$\max_{f^0, c^0, o^0, p^0} \Pi_0(f^0, c^0, o^0, p^0) + \varpi V_0(f^0, c^0, o^0, p^0), \tag{EC.40}$$

subject to the incentive compatibility constraint $V_t(f^t, c^t, o^t, p^t) = v_t^e = \max_{\check{p}^t} V_t(f^t, c^t, o^t, \check{p}^t)$, $\forall\, \check{p}^t$ such that $(f^t, c^t, o^t, \check{p}^t) \in \Gamma_t$. Using the definitions of Π_t and V_t from (2) and (3):

$$\begin{aligned} \Pi_t(f^t, c^t, o^t, p^t) + \varpi V_t(f^t, c^t, o^t, p^t) &= \mathbb{E}\Big[\sum_{t<s\le\tau} \gamma^{s-t} p_s + \varpi \sum_{t<s\le\tau} \delta^{s-t} \mathcal{U}(Y_s - p_s, f_s) \\ &\quad + \gamma^{\tau-t}(\rho \mathcal{C}_\tau + (1-\rho)\mathcal{O}_\tau) + \varpi \delta^{\tau-t}(\rho \underline{v}_\tau + (1-\rho)\overline{v}_\tau) \,\Big|\, f^t, c^t, o^t, p^t\Big] \\ &= \mathbb{E}\Big[\sum_{t<s\le\tau} \gamma^{s-t}\big(p_s + \varpi_{s-t}\, \mathcal{U}(Y_s - p_s, f_s)\big) \\ &\quad + \gamma^{\tau-t}\big(\rho \mathcal{C}_\tau + (1-\rho)\mathcal{O}_\tau + \varpi_{\tau-t}(\rho \underline{v}_\tau + (1-\rho)\overline{v}_\tau)\big) \,\Big|\, f^t, c^t, o^t, p^t\Big]. \end{aligned}$$

In other words, we can express the weighted sum $\Pi_t(f^t, c^t, o^t, p^t) + \varpi V_t(f^t, c^t, o^t, p^t)$ similarly to the expression of Π_t, but with p_s in period s replaced with weighted sum $p_s + \varpi_{s-t}\, \mathcal{U}(Y_s - p_s, f_s)$, where the weight ϖ_{s-t} is placed on consumer utility and is given by $\varpi_t = (\delta/\gamma)^t \varpi$ for any given t. Similar replacement holds for the termination part as well.

Now, following the recursive formulation logic described in Section 3.4, we can define the following end-of-period continuation payoff function:

$$\begin{aligned} \zeta_t^e(v_t^e) = \max_{(f^t, c^t, o^t, p^t)} \quad & \Pi_t(f^t, c^t, o^t, p^t) + \varpi V_t(f^t, c^t, o^t, p^t) \\ \text{s.t.} \quad & V_t(f^t, c^t, o^t, p^t) = v_t^e = \max_{\check{p}^t} V_t(f^t, c^t, o^t, \check{p}^t). \end{aligned}$$

As in Section 4, we can define continuation functions ζ_t^b, ζ_t^p, and ζ_t^a at different stages within period t. The problem for ζ_t^b is given by

$$\begin{aligned} \zeta_t^b(v_t^b) = \max_{p_t, v_t^p} \quad & \mathbb{E}\big[p_t(Y_t) + \varpi_t u(Y_t - p_t(Y_t)) + \zeta_t^p(v_t^p(p_t(Y_t)))\big] \\ \text{s.t.} \quad & p_t(y_t) \in \operatorname*{argmax}_{0 \le p \le y_t} u(y_t - p) + v_t^p(p), \quad \forall\, y_t \in \mathcal{Y}_t, \quad \text{and} \\ & v_t^b = \mathbb{E}\big[u(Y_t - p_t(Y_t)) + v_t^p(p_t(Y_t))\big]. \end{aligned}$$

Compare this with the problem in (9)–(11): in the objective, as we just discussed above, instead of only payment p_t, we additionally have consumer's consumption utility $u(Y_t - p_t)$ with a weight ϖ_t. Now notice that using the promise-keeping constraint above, we can rewrite this problem as follows:

$$\begin{aligned}\zeta_t^b(v_t^b) - \varpi_t v_t^b = \max_{p_t,\, v_t^p} \quad & \mathbb{E}\big[p_t(Y_t) + \zeta_t^p(v_t^p(p_t(Y_t))) - \varpi_t v_t^p(p_t(Y_t))\big] \\ \text{s.t.} \quad & p_t(y_t) \in \operatorname*{argmax}_{0\le p\le y_t} u(y_t - p) + v_t^p(p), \quad \forall y_t \in \mathcal{Y}_t, \quad \text{and} \\ & v_t^b = \mathbb{E}\big[u(Y_t - p_t(Y_t)) + v_t^p(p_t(Y_t))\big],\end{aligned} \tag{EC.41}$$

which is a problem that solves for $\zeta_t^b(v_t^b) - \varpi_t v_t^b$ recursively in terms of $\zeta_t^p(v_t^p) - \varpi_t v_t^p$. We will now see the same pattern with the other problems. For ζ_t^p, we solve

$$\begin{aligned}\zeta_t^p(v_t^p) = \max_{f_t,\, v_t^a} \quad & \varpi_t f_t A + \zeta_t^a(v_t^a) \\ \text{s.t.} \quad & v_t^p = f_t A + v_t^a,\end{aligned}$$

where the objective includes consumer's access utility with weight ϖ_t. This problem can be re-expressed as

$$\begin{aligned}\zeta_t^p(v_t^p) - \varpi_t v_t^p = \max_{f_t,\, v_t^a} \quad & \zeta_t^a(v_t^a) - \varpi_t v_t^a \\ \text{s.t.} \quad & v_t^p = f_t A + v_t^a,\end{aligned} \tag{EC.42}$$

i.e., $\zeta_t^p(v_t^p) - \varpi_t v_t^p$ is expressed as a recursive function of $\zeta_t^a(v_t^a) - \varpi_t v_t^a$. In a similar vein:

$$\begin{aligned}\zeta_t^a(v_t^a) = \max_{c_t,\, o_t,\, v_t^e} \quad & c_t\mathcal{C}_t + o_t\mathcal{O}_t + \varpi_t(c_t\mathcal{C}_t + o_t\mathcal{O}_t) + (1 - c_t - o_t)\zeta_t^e(v_t^e) \\ \text{s.t.} \quad & v_t^a = c_t\underline{v}_t^e + o_t\overline{v}_t^e + (1 - c_t - o_t)v_t^e,\end{aligned}$$

which can be transformed as

$$\begin{aligned}\zeta_t^a(v_t^a) - \varpi_t v_t^a = \max_{c_t,\, o_t,\, v_t^e} \quad & c_t\mathcal{C}_t + o_t\mathcal{O}_t + (1 - c_t - o_t)(\zeta_t^e(v_t^e) - \varpi_t v_t^e) \\ \text{s.t.} \quad & v_t^a = c_t\underline{v}_t^e + o_t\overline{v}_t^e + (1 - c_t - o_t)v_t^e,\end{aligned} \tag{EC.43}$$

in which again $\zeta_t^a(v_t^a) - \varpi_t v_t^a$ is a recursive function of $\zeta_t^e(v_t^e) - \varpi_t v_t^e$. Finally, $\zeta_t^e(v_t^e) = \gamma\zeta_{t+1}^b(v_t^e/\delta)$, and it can also be rewritten as

$$\zeta_t^e(v_t^e) - \varpi_t v_t^e = \gamma(\zeta_{t+1}^b(v_t^e/\delta) - \varpi_{t+1}(v_t^e/\delta), \tag{EC.44}$$

because $\varpi_t = \gamma\varpi_{t+1}/\delta$, and that gives us $\zeta_t^e(v_t^e) - \varpi_t v_t^e$ expressed recursively in terms of $\zeta_{t+1}^b(v_{t+1}^b) - \varpi_{t+1}v_{t+1}^b$. Now we can see from (EC.41), (EC.42), (EC.43), and (EC.44) that all the above transformed problems are the same as in Sections 4.1 to 4.3 if we substitute $\pi_t^b = \zeta_t^b(v_t^b) - \varpi_t v_t^b$, $\pi_t^p = \zeta_t^p(v_t^p) - \varpi_t v_t^p$, $\pi_t^a = \zeta_t^a(v_t^a) - \varpi_t v_t^a$, and $\pi_t^e = \zeta_t^e(v_t^e) - \varpi_t v_t^e$. Thus, instead of solving for $\zeta_t^j(v_t^j)$, we could solve for $\pi_t^j(v_t^j)$, which would result in the same continuation contract as in Section 4, and we can get back ζ_t^j using $\zeta_t^j(v_t^j) = \pi_t^j(v_t^j) + \varpi_t v_t^j$ for $j \in \{b, p, a, e\}$. As a result, the problem (EC.40) that we started with—which is same as $\zeta_0^e(v_0^e)$—can be written as $\zeta_0^e(v_0^e) = \pi_0^e(v_0^e) + \varpi v_0^e = \gamma\pi_1^b(v_1^b) + \varpi\delta v_1^b$, as we see in (19).

Part 2: $d^*(\varpi)$ is weakly decreasing. We first introduce several new objects for the analysis and derive their properties. Note that since $v_1^b \in [\underline{v}_1^b, \overline{v}_1^b]$, the uptake probability in (19) will be zero for any possible choice of v_1^b if $\tilde{u}(W -$

$d) + \delta \bar{v}_1^b < \tilde{u}(W) + \delta \underline{v}_1^b$. Since $\tilde{u}(W) - \tilde{u}(W-d) \geq \tilde{\lambda}_2 d$, an immediate implication is that the uptake probability will be zero if $d > \delta(\bar{v}_1^b - \underline{v}_1^b)/\tilde{\lambda}_2$. Thus, without loss of optimality, we can constrain the choice set for d in (19) to $D := \left[0, \delta(\bar{v}_1^b - \underline{v}_1^b)/\tilde{\lambda}_2\right]$. Similarly, from the $\tilde{u}(W-d) + \delta \bar{v}_1^b \geq \tilde{u}(W) + \delta \underline{v}_1^b$ condition for the consumer to take the contract in (19), and the shape of $\tilde{u}$, it follows that the wealth cutoff point (the lowest wealth at which the consumer will take up the contract) in (19) is

$$\underline{w}(d, v_1^b) \;:=\; \max\left\{ d,\; \tilde{\kappa} + \frac{\tilde{\lambda}_1 d - \delta\left(v_1^b - \underline{v}_1^b\right)}{\tilde{\lambda}_1 - \tilde{\lambda}_2} \right\}.$$

Note that $\underline{w}(d, v_1^b)$ is continuous and piecewise affine in (d, v_1^b). Its right partial derivatives are $\partial \underline{w}/\partial^+ d \in \left\{1, \tilde{\lambda}_1/(\tilde{\lambda}_1 - \tilde{\lambda}_2)\right\}$ and $\partial \underline{w}/\partial^+ v_1^b \in \left\{0,\; -\delta/(\tilde{\lambda}_1 - \tilde{\lambda}_2)\right\}$, so $\partial \underline{w}/\partial^+ d \geq 1$ and $\partial \underline{w}/\partial^+ v_1^b \leq 0$. Since $D \times [\underline{v}_1^b, \bar{v}_1^b]$ is compact and $\underline{w}$ is continuous, the image of $\underline{w}$ is compact.

Denote by $\mathcal{G}$ the welfare kernel (the expected consumer utility from taking up the contract), which is

$$\mathcal{G}(d, v_1^b, w) \;:=\; \tilde{u}(w-d) + \delta\, v_1^b - \tilde{u}(w) - \delta\, \underline{v}_1^b.$$

The right partial derivatives are $\partial \mathcal{G}/\partial^+ v_1^b = \delta > 0$ and $\partial \mathcal{G}/\partial^+ d = -\tilde{u}'(w-d) \leq 0$. We also have $\mathcal{G}\big(d, v_1^b, \underline{w}(d, v_1^b)\big) \geq 0$ by construction.

Lastly, with H as the cdf of W, we can write the objective function of (19) as

$$\tilde{\pi}_0(d, v_1^b, \varpi) := \int_{\underline{w}(d, v_1^b)}^{\infty} \Big[d + \gamma\, \pi_1^b(v_1^b) - \mathcal{K} + \varpi\, \mathcal{G}(d, v_1^b, w)\Big]\, dH(w),$$

which can be decomposed into profit and welfare components

$$\tilde{\mathcal{F}}(d, v_1^b) := [d + \gamma\, \pi_1^b(v_1^b) - \mathcal{K}]\,(1 - H(\underline{w}(d, v_1^b))), \qquad \tilde{\mathcal{W}}(d, v_1^b) := \int_{\underline{w}(d, v_1^b)}^{\infty} \mathcal{G}(d, v_1^b, w)\, dH(w),$$

so that $\tilde{\pi}_0(d, v_1^b, \varpi) = \tilde{\mathcal{F}}(d, v_1^b) + \varpi\, \tilde{\mathcal{W}}(d, v_1^b)$. Note that all of the following functions are semi-differentiable: $\mathcal{G}$ and $\underline{w}$ (as they are piecewise-linear), H (as it is differentiable), and π_1^b (as it is concave). Thus, $\tilde{\mathcal{F}}$, $\tilde{\mathcal{W}}$ and $\tilde{\pi}_0$ are also semi-differentiable. From Leibniz rule we have

$$\frac{\partial \tilde{\mathcal{W}}(d, v_1^b)}{\partial^+ d} = \int_{\underline{w}(d, v_1^b)}^{\infty} \frac{\partial \mathcal{G}(d, v_1^b, w)}{\partial^+ d}\, dH(w) - \frac{\partial \underline{w}(d, v_1^b)}{\partial^+ d} \mathcal{G}(d, v_1^b, \underline{w}(d, v_1^b)) h(\underline{w}(d, v_1^b)) < 0. \tag{EC.45}$$

With this notation, we can write the optimal decisions as

$$(d^*(\varpi), v_1^{b*}(\varpi)) \in \underset{(d, v_1^b) \in D \times [\underline{v}_1^b, \bar{v}_1^b]}{\arg\max}\; \tilde{\pi}_0(d, v_1^b, \varpi), \tag{EC.46}$$

so that the overall payoff $\pi_0(\varpi) = \tilde{\pi}_0(d^*(\varpi), v_1^{b*}(\varpi), \varpi)$, expected firm profit $\mathcal{F}(\varpi) = \tilde{\mathcal{F}}(d^*(\varpi), v_1^{b*}(\varpi))$, and expected consumer welfare $\mathcal{W}(\varpi) = \tilde{\mathcal{W}}(d^*(\varpi), v_1^{b*}(\varpi))$. The problem (EC.46) can be reduced to a univariate one through introduction of the reduced value function

$$\phi(d, \varpi) := \max_{v_1^b \in [\underline{v}_1^b, \bar{v}_1^b]} \tilde{\pi}_0(d, v_1^b, \varpi).$$

From compactness of the choice set and continuity of the objective function, it follows that $\phi(d, \varpi)$ is continuous and finite by Berge's Theorem of the Maximum (Berge 1963). The original problem $\max_{d, v_1^b} \tilde{\pi}_0(d, v_1^b, \varpi)$ is equivalent to $\max_d \phi(d, \varpi)$.

For a fixed (d,ϖ), denote the set of optimal v_1^b-s by $\mathscr{V}(d,\varpi) := \arg\max_{v_1^b} \tilde{\pi}_0(d,v_1^b,\varpi)$. By Danskin's Theorem (Danskin 1967), $\partial(d,\varpi)\phi/\partial^+ d = \max_{v_1^b \in \mathscr{V}(d,\varpi)} \partial\tilde{\pi}_0(d,v_1^b,\varpi)/\partial^+ d$. Since

$$\frac{\partial\tilde{\pi}_0(d,v_1^b,\varpi)}{\partial^+ d} = \frac{\partial\tilde{\mathcal{F}}(d,v_1^b)}{\partial^+ d} + \varpi\frac{\partial\tilde{\mathcal{W}}(d,v_1^b)}{\partial^+ d},$$

and from (EC.45) we have $\partial\tilde{\mathcal{W}}(d,v_1^b)/\partial^+ d < 0$, the map $\varpi \mapsto \partial\tilde{\pi}_0(d,v_1^b,\varpi)/\partial^+ d$ is affine and weakly decreasing in ϖ for each (d,v_1^b). Taking the max over $v_1^b \in \mathscr{V}(d,\varpi)$ preserves this monotonicity, so for each fixed d, the map $\varpi \mapsto \partial\phi(d,\varpi)/\partial^+ d$ is weakly decreasing. So, for any $\varpi_2 > \varpi_1$, we have that $\partial\phi(d,\varpi_2)/\partial^+ d - \partial\phi(d,\varpi_1)/\partial^+ d \le 0$, thus $\phi(d,\varpi)$ is submodular. Since D is a compact interval and ϕ is submodular, Topkis's Theorem (Topkis 1998) implies that the argmax correspondence

$$\mathscr{D}(\varpi) = \arg\max_{d\in D} \phi(d,\varpi)$$

is nonincreasing in ϖ. Therefore, there exists a selection $d^*(\varpi) \in \mathscr{D}(\varpi)$ that is weakly decreasing.

Part 3: $v_1^{b*}(\varpi)$ is weakly increasing. For any fixed d, The cross-effect of ϖ on the marginal value of v_1^b is

$$\frac{\partial^2\tilde{\pi}_0}{\partial^+ v_1^b\,\partial^+\varpi}(d,v_1^b,\varpi) = \frac{\partial}{\partial^+ v_1^b}\,\tilde{\mathcal{W}}(d,v_1^b) = \underbrace{\int_{\underline{w}(d,v_1^b)}^{\infty} \frac{\partial\mathcal{G}}{\partial^+ v_1^b}\,dH(w)}_{=\,\delta\,\overline{H}(\underline{w})\,\ge 0} \;-\; \underbrace{\mathcal{G}(d,v_1^b,\underline{w})}_{\ge 0}\,\underbrace{\frac{\partial\underline{w}}{\partial^+ v_1^b}}_{\le 0} \;\ge\; 0. \tag{EC.47}$$

Here we used $\partial\mathcal{G}/\partial^+ v_1^b = \delta > 0$, $\mathcal{G}(d,v_1^b,\underline{w}) \ge 0$, and $\partial\underline{w}/\partial^+ v_1^b \le 0$. Since $\tilde{\pi}_0$ is continuous (shown in Part 2 of this proof), it follows from (EC.47) that for each fixed d, $\tilde{\pi}_0(d,\cdot,\cdot)$ has *increasing differences* in (v_1^b,ϖ).

Define $\hat{\phi}(v_1^b,\varpi) := \sup_d \tilde{\pi}_0(d,v_1^b,\varpi)$. The pointwise supremum over d of functions that have increasing differences continues to have increasing differences (Topkis 1998). Therefore, the argmax in v_1^b is increasing in ϖ; that is, there exists a selection $v_1^{b*}(\varpi)$ that is weakly increasing.

Part 4: $\pi_0(\varpi)$ is weakly increasing (and convex). For any fixed (d,v_1^b), $\tilde{\pi}_0(d,v_1^b,\varpi) = \tilde{\mathcal{F}}(d,v_1^b) + \varpi\,\tilde{\mathcal{W}}(d,v_1^b)$ is affine in ϖ with slope $\tilde{\mathcal{W}}(d,v_1^b)$. We first show $\tilde{\mathcal{W}}(d,v_1^b) \ge 0$. By construction, $\mathcal{G}(d,v_1^b,\underline{w}(d,v_1^b)) \ge 0$. Moreover, $\mathcal{G}$ is increasing in w because $\partial\mathcal{G}/\partial^+ w = \tilde{u}'^{+}(w-d) - \tilde{u}'^{+}(w) \;\ge\; 0$, since $w - d \le w$ and $\tilde{u}'^{+}$ is weakly decreasing (concavity). Therefore, for all $w \ge \underline{w}(d,v_1^b)$, $\mathcal{G}(d,v_1^b,w) \ge 0$, and

$$\tilde{\mathcal{W}}(d,v_1^b) = \int_{\underline{w}(d,v_1^b)}^{\infty} \mathcal{G}(d,v_1^b,w)\,dH(w) \;\ge\; 0.$$

Thus $\tilde{\pi}_0(d,v_1^b,\varpi)$ is affine in ϖ with nonnegative slope. The pointwise supremum of affine functions with nonnegative slopes is weakly increasing and convex, so $\pi_0(\varpi) = \max_{d,v_1^b} \tilde{\pi}_0(d,v_1^b,\varpi)$ inherits both properties.

Part 5: $\mathcal{W}(\varpi)$ is weakly increasing. Consider $\mathcal{W}(\varpi) = \tilde{\mathcal{W}}\big(d^*(\varpi), v_1^{b*}(\varpi)\big)$. From (EC.45) we have that $\tilde{\mathcal{W}}(d,v_1^b)$ is decreasing in d and using the Leibniz rule yields

$$\frac{\partial\tilde{\mathcal{W}}(d,v_1^b)}{\partial^+ v_1^b} = \int_{\underline{w}}^{\infty} \frac{\partial\mathcal{G}(d,v_1^b,w)}{\partial^+ v_1^b}\,dH(w) - \mathcal{G}(d,v_1^b,\underline{w})\,h(\underline{w})\,\frac{\partial\underline{w}}{\partial^+ v_1^b} = \delta\,\overline{H}(\underline{w}) - \mathcal{G}(d,v_1^b,\underline{w})\,h(\underline{w})\,\frac{\partial\underline{w}}{\partial^+ v_1^b} \ge 0,$$

so $\tilde{\mathcal{W}}(d,v_1^b)$ is weakly increasing in v_1^b. Since $d^*(\varpi)$ is weakly decreasing (part 2 of this proposition), and $v_1^{b*}(\varpi)$ is weakly increasing (part 3 of this proposition), it follows that the composition $\mathcal{W}(\varpi) = \tilde{\mathcal{W}}\big(d^*(\varpi), v_1^{b*}(\varpi)\big)$ is weakly increasing in ϖ.

Part 6: $\mathcal{F}(\varpi)$ is weakly decreasing. For any $\varpi_2 \ge \varpi_1 \ge 0$, from optimality of d^* and v^{b*} we have

$$\tilde{\pi}_0(d^*(\varpi_1), v_1^{b*}(\varpi_1), \varpi_1) = \mathcal{F}(\varpi_1) + \varpi_1\,\mathcal{W}(\varpi_1) \ge \mathcal{F}(\varpi_2) + \varpi_1\,\mathcal{W}(\varpi_2) = \tilde{\pi}_0(d^*(\varpi_1), v_1^{b*}(\varpi_1), \varpi_2)$$

Rearranging the inequality gives $\mathcal{F}(\varpi_2)-\mathcal{F}(\varpi_1)\leq\varpi_1\big(\mathcal{W}(\varpi_1)-\mathcal{W}(\varpi_2)\big)\leq 0$, because $\mathcal{W}(\varpi)$ weakly increasing (from part 5 of this proposition). Hence, $\mathcal{F}(\varpi_2)\leq\mathcal{F}(\varpi_1)$. $\square$

Proof of Proposition EC.1. *Part 1: optimality of* $v_1^{b*}(d;\alpha_0,\kappa_0^v)$. First, we note an intuitive property: (EC.2) implies that without loss of optimality we can restrict attention in (EC.1)–(EC.2) 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 (EC.2), 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 (EC.1)–(EC.2), this problem can be formulated as

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

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

$$\max_{v_1^b\in\mathcal{V}_0}\quad\mathbb{E}\big[d(W_{d,\underline{d}})+\gamma\pi_1^b(v_1^b(d(W_{d,\underline{d}})))\big]\tag{EC.48}$$

$$\text{s.t.}\quad d(w)\in\underset{\underline{d}\leq d<w}{\operatorname{argmax}}\ \tilde{u}(w-d)+\delta v_1^b(d),\quad\forall w\in\operatorname{supp}W_{d,\underline{d}},\tag{EC.49}$$

$$\tilde{u}(w-d(w))+\delta v_1^b(d(w))\geq\tilde{u}(w)+\delta\underline{v}_1^b,\quad\forall w\in\operatorname{supp}W_{d,\underline{d}},\tag{EC.50}$$

$$v_{d,\underline{d}}^{\dagger\dagger}=\mathbb{E}\big[\tilde{u}(W_{d,\underline{d}}-d(W_{d,\underline{d}}))+\delta v_1^b(d(W_{d,\underline{d}}))\big].\tag{EC.51}$$

Now, consider the closely related problem

$$\max_{v_1^b\in\mathcal{V}_0}\quad\mathbb{E}\big[d(W_{d,\underline{d}})+\gamma\pi_1^b(v_1^b(d(W_{d,\underline{d}})))\big]\tag{EC.52}$$

$$\text{s.t.}\quad d(w)\in\underset{0\leq d<w}{\operatorname{argmax}}\ \tilde{u}(w-d)+\delta v_1^b(d),\quad\forall w\in\operatorname{supp}W_{d,\underline{d}},\tag{EC.53}$$

$$v_{d,\underline{d}}^{\dagger\dagger}=\mathbb{E}\big[\tilde{u}(W_{d,\underline{d}}-d(W_{d,\underline{d}}))+\delta v_1^b(d(W_{d,\underline{d}}))\big].\tag{EC.54}$$

This is useful as (EC.52)–(EC.54) is analogous to (9)–(11) 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 (EC.3), such that it satisfies (EC.53)–(EC.54) and performs as well as $\hat{v}_1^b$ or better in (EC.52). It is easily verifiable that v_1^{b*} also satisfies (EC.50) and—because (EC.53) is tighter than (EC.49)—also satisfies (EC.49). Thus, v_1^{b*} also performs as well or better than $\hat{v}_1^b$ in (EC.48)–(EC.51). Consequently, in (EC.1)–(EC.2), we can restrict attention to reward functions that satisfy the functional form in (EC.3) 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 (EC.1)–(EC.2), and afterword verify that this solution really belongs to $\mathcal{D}$. To do that, we first introduce some notation needed to rewrite (EC.2) 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 optimality, in (EC.1)–(EC.2) 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 (EC.2)) 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 (EC.1)–(EC.2) for each $w \geq \underline{w}$ by

$$\max_{d(w)} \quad d(w) + \gamma \pi_1^b(v_1^{b*}(d(w); \alpha_0, \kappa_0^v)) - \mathcal{K} \tag{EC.55}$$

$$\text{s.t.} \quad d^\circ(w; \alpha_0, \kappa_0^v, \underline{d}) \leq d(w) \leq d^\times(w; \alpha_0, \kappa_0^v, \underline{d}). \tag{EC.56}$$

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 (EC.55)–(EC.56) 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 (EC.55)–(EC.56), 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, note that $d^*(w; \alpha_0, \kappa_0^v, \underline{d})$ is piecewise-linear (thus also semi-differentiable) as well as continuous, so $d^*(w; \alpha_0, \kappa_0^v, \underline{d}) \in \mathcal{D}$, completing the proof. ☐

Proof of Proposition EC.2. *Part (i).* To avoid notational confusion, we denote by $d^\dagger, v_1^{b\dagger}$ the solution of (17), and assume the opposite: there exist $\alpha_0^*, \kappa_0^{v*}, \underline{d}^*$, such that $d^{**}, v_1^{b**}, \underline{d}^*$ solve (EC.1)–(EC.2) and $d^\dagger < \underline{d}^*$. (As introduced in the paragraph after Proposition EC.1, d^{**}, v_1^{b**} are the functions d^*, v_1^{b*} given by Proposition EC.1, evaluated the optimal parameters $\alpha_0^*, \kappa_0^{v*}, \underline{d}^*$.) Then, denoting the cdf of W by H, from (EC.1) we have that the expected profit of the firm under $d^{**}, v_1^{b**}, \underline{d}^*$ is

$$\begin{aligned} &(\underline{d}^* + \gamma\pi_1^b(v_1^{b**}(\underline{d}^*)) - \mathcal{K}) \times \Pr\{W \geq \underline{w}(\alpha_0^*, \kappa_0^{v*}, \underline{d}^*)\} \\ &\qquad + \int_{\inf\{x \in \mathbb{R}_+ | d^{**}(x) > \underline{d}^*\}}^{\infty} \left(\pi_1^b(v_1^{b**}(d^{**}(w))) + d^{**}(w) - \pi_1^b(v_1^{b**}(\underline{d}^*)) - d^{**}(\underline{d}^*)\right) dH(w). \end{aligned} \tag{EC.57}$$

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

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

Here, the first line is no larger than the first line of (EC.57) (because of the optimality of $d^\dagger, v_1^{b\dagger}$ in (17)), the second line is the same as in (EC.57), 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 $\Pr\left\{W \geq \underline{d},\, \tilde{u}(W - \underline{d}) + \delta v_1^b \geq \tilde{u}(W) + \delta \underline{v}_1^b\right\} = 1 - H(\underline{w}(\alpha_0, \kappa_0^v, \underline{d}))$ 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 (17) being the same optimization problem as (EC.1)–(EC.2), just over a smaller parameter space. More precisely, restricting the firm's choice in (EC.1)–(EC.2) to reward functions that satisfy $v_1^b(d) = v_1^b, \forall d \geq \underline{d}$ retrieves (17). ☐

Proof of Proposition EC.3. *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 $(\pi_t^p)'(\bar{v}_t^p) \geq -1/\lambda_2$, or equivalently $(\pi_t^a)'(\bar{v}_t^a) \geq -1/\lambda_2$. 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 $\hat{v}_t^b < \bar{v}_t^b$ then $\pi_t^b(v_t^b) = \pi_t^p(v_t^b - \underline{\omega}_t)$, so from $(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) = \gamma/\delta(\pi_t^b)'(\bar{v}_{t-1}^e/\delta)$ in this case we have $(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) = \gamma/\delta(\pi_t^a)'(\bar{v}_t^a)$. On the other hand, if $\hat{v}_t^b = \bar{v}_t^b$, by inserting $\omega_t^*(v_t^b)$ into (EC.39) we obtain $(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) = \gamma/\delta\left[(\pi_t^a)'(\bar{v}_t^a)G_t(\kappa) - \bar{G}_t(\kappa)/\lambda_2\right]$.

Applying Proposition 1, the situations where the principal desires to terminate irrespective of payment arise if and only if (7) is satisfied, which here becomes equivalent to $-\mathcal{C}_t/(\bar{v}_t^e - \underline{v}_t^e) \leq (\pi_t^e)'(\bar{v}_t^e)$. 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

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

If $t > \theta$, then (EC.5) holds, so

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

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

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

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 EC.4. From the proof of Proposition EC.3 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

$$(\pi_{t-1}^e)'(\bar{v}_{t-1}^e) = \begin{cases} \frac{\gamma}{\delta}(\pi_t^e)'(\bar{v}_t^e) & \text{if } (\pi_t^e)'(\bar{v}_t^e) < -\frac{1}{\lambda_2}, \\ -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_t}{\bar{v}_t^e - \underline{v}_t^e}G_t(\kappa) + \frac{1}{\lambda_2}\bar{G}_t(\kappa)\right] & \text{if } (\pi_t^e)'(\bar{v}_t^e) \geq -\frac{1}{\lambda_2}. \end{cases} \tag{EC.58}$$

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 (EC.58)), 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 EC.1 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,

$$(\pi^e_{\theta^*-1})'(\bar{v}^e_{\theta^*-1}) = -\frac{\gamma}{\delta}\left[\frac{\mathcal{C}_{\theta^*}}{\bar{v}^e_{\theta^*} - \underline{v}^e_{\theta^*}} G_{\theta^*}(\kappa) + \frac{1}{\lambda_2}\bar{G}_{\theta^*}(\kappa)\right], \tag{EC.59}$$

which could be either smaller or greater than $-1/\lambda_2$. Note that (EC.58) 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 (EC.59) 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 (EC.59), the statement of parts (i) and (ii) of the proposition then follow by setting

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

Proof of Proposition EC.5. The state dependency of autarchy (part (i) of the proposition) follows directly from Assumption EC.1. Proofs of Propositions 1-6 for this setting are analogous to ones in the main text, except this state dependency needs to be accounted for in all model composites that depend on the utility of autarchy (notably, which promised future utilities are feasible). Parts (ii) and (iii) of the proposition collect the resulting changes.

The key divergence from the main model comes in the way promised utility is updated between periods. In the main model, this is deterministic, as the only way for the firm to ensure the the promise of v^e_{t-1} is met at this point was to appreciate the v-score at the consumer's discount rate, resulting in $v^e_{t-1} = v^b_t/\delta$. Under Assumption EC.1, this is no longer the case, as all updates that satisfy (EC.9) ensure the firm's promise is kept. Thus, the firm faces an additional optimization problem: which of such updates maximizes its own interest; this problem is given in (EC.8)–(EC.9).

It remains to show concavity of $\pi^e_{t-1}(v^e_{t-1}, x_{t-1})$. Note that this property is needed to ensure the inductive argument introduced in the first paragraph of Section 4 still holds in this setting. From Proposition 4, for any $x_t \in \operatorname{supp}X_t$, $\pi^b_t(v^b_t(x_t), x_t)$ is concave in $v^b_t(x_t)$ and does not depend on $v^b_t(y_t)$ for any $y_t \in \operatorname{supp}X_t \setminus \{x_t\}$. Thus, for any $x_t \in \operatorname{supp}X_t$, $\pi^b_t(v^b_t(x_t), x_t)$ is jointly concave in $\mathbf{v}^b_t := (v^b_t(x_1), v^b_t(x_2), ..., v^b_t(x_{|\operatorname{supp}X_t|}))$. Since the objective function of (EC.8) is a non-negative weighted sum of functions that are jointly concave it $\mathbf{v}^b_t$, it is also jointly concave in $\mathbf{v}^b_t$. Finally, since (EC.9) defines a convex set (it is the intersection of a hyperrectangle and a hyperplane, both of which are convex), concavity of $\pi^e_{t-1}(v^e_{t-1}, x_{t-1})$ follows from the results on preservation of concavity under maximization for jointly concave functions (Boyd and Vandenberghe 2004, p.p. 87-88). $\square$

Proof of Proposition EC.6. *Part (i).* Expanding the integral (expectation) in (15), analogously to how the expression for ω_t in (16) was derived, yields that

$$\begin{aligned}\alpha^*_t(x_t, \kappa^\nu_t) &= v^b_t - \int_0^{\kappa^v_t+\kappa} \lambda_1 y \, dG_t(y|x_t) - \int_{\kappa^v_t+\kappa}^{\infty} \left[(\lambda_1 - \lambda_2)(\kappa^v_t + \kappa) + \lambda_2 y\right] dG_t(y|x_t), \\ &= v^b_t - U_t(x_t) - \int_\kappa^\infty (\lambda_1 - \lambda_2)\min\{y - \kappa, \kappa^\nu_t\} dG_t(y|x_t).\end{aligned}$$

Since the integrand of the last integral is a non-decreasing function of y, first-order stochastic dominance implies the integral here is larger when $x_t = 2$. Thus, $\alpha^*_t(1, \kappa^\nu_t) - U_t(1) \ge \alpha^*_t(2, \kappa^\nu_t) - U_t(2)$.

Part (ii). From Karush-Kuhn-Tucker conditions of the problem (EC.8)–(EC.9) it follows that $\frac{\partial \pi^b_t(v,1)}{\partial v}\Big|_{v=v^{b*}_t(1; v^e_{t-1}, x_{t-1})} = \frac{\partial \pi^b_t(v,1)}{\partial v}\Big|_{v=v^{b*}_t(2; v^e_{t-2}, x_{t-1})}$ is a necessary condition for optimality. (As a technical detail, note that we actually do not have a guarantee the π_t^b is differentiable; however, this can be rigorously resolved in the same way as it was done in the proof of Proposition 4, by replacing π_t^b with its smooth approximation, which can—due to its concavity—be made arbitrarily precise by following Azagra 2013.)

Part (iii). Assume $ARA_{\pi_t^b(\cdot,1)}(v - \underline{v}_t^b(1)) < ARA_{\pi_t^b(\cdot,2)}(v - \underline{v}_t^b(2))$, $\forall v$ and introduce substitution $z_i^* := v_t^{b*}(i; v_{t-1}^e, x_{t-1}) - \underline{v}_t^b(i)$ for $i \in \{1,2\}$. For contradiction assume that $z_1^* < z_2^*$. We can define the firm's shifted continuation functions as $\tilde{\pi}_t^b(z,i) := \pi_t^b(v - \underline{v}_t^b(i), i)$. The necessary optimality condition then gives $(\tilde{\pi}_t^b)'(z_1^*,1) = (\tilde{\pi}_t^b)'(z_2^*,2)$. By concavity, $(\tilde{\pi}_t^b)'(z,1)$ is decreasing, so $(\tilde{\pi}_t^b)'(z_2^*,1) \le (\tilde{\pi}_t^b)'(z_1^*,1) = (\tilde{\pi}_t^b)'(z_2^*,2)$, hence $\log(\tilde{\pi}_t^b)'(z_2^*,1) \le \log(\tilde{\pi}_t^b)'(z_2^*,2)$. The derivative of this log gap is

$$\frac{d}{dz}\log(\tilde{\pi}_t^b)'(z,1) - \frac{d}{dz}\log(\tilde{\pi}_t^b)'(z,2) = -\mathrm{ARA}_{\pi_t^b(\cdot,1)}(z + \underline{v}_t^b(1)) + \mathrm{ARA}_{\pi_t^b(\cdot,1)}(z + \underline{v}_t^b(1)) < 0.$$

So, the log gap is strictly decreasing, contradicting that it was zero at z_1^* and smaller or equal to zero at z_2^*. Hence $z_1^* \ge z_2^*$. □

Proof of Proposition EC.7. *Part (i).* follows directly from state-dependence of consumer utility in absence of the firm, analogously to parts (i) and (ii) of Proposition EC.7.

Part (ii). follows from the choice of $v_t^b(0,0)$ uniquely determining $v_t^b(0,1)$ as described in the text, and from constraints (EC.12)–(EC.13) being two linear equations of $v_t^b(0,1)$ and $v_t^b(1,1)$ with different slopes (due to state-dependent transition probabilities), thus they uniquely determine the pair $(v_t^b(0,1), v_t^b(1,1))$.

Part (iii). follows analogously to Propositions 1-6 after observing (and using part (ii) of this proposition) that in all sub-periods of the problem, any fixed choice of promised future utility also uniquely defines the promised threat utility.

Part (iv). Assume the opposite: there exist $v_t^b(0,0)$ and $v_t^b(1,1)$ that are part of the solution for problem (EC.10)–(EC.13) such that $v_t^b(0,0) - \underline{v}_t^b(0) > v_t^b(1,1) - \underline{v}_t^b(1)$ (the firm provides greater utility to an unemployed consumer). But then $v_t^b(0,1) \ge v_t^b(0,0) - \underline{v}_t^b(0) + \underline{v}_t^b(1)$. Since the utility of autarchy is higher when employed as is the probability of also being employed in the next period, we have $\underline{v}_t^b(1) > \underline{v}_t^b(0)$ (in absence of the firm, an employed consumer is always better off than an unemployed one). Thus, $v_t^b(0,0) - \underline{v}_t^b(0) + \underline{v}_t^b(1) > v_t^b(0,0)$, so $v_t^b(0,1) > v_t^b(0,0)$, which is a contraction as it violates the incentive compatibility constraint (EC.11). □

References

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

Berge, C. 1963. *Topological Spaces*. Macmillan, New York.

Bertsekas, D. 2012. *Dynamic programming and optimal control: Volume II*. Athena scientific.

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

Danskin, J.M. 1967. *The theory of max-min and its application to weapons allocation problems*. Springer Science & Business Media.

Fernandes, Ana, Christopher Phelan. 2000. A recursive formulation for repeated agency with history dependence. *Journal of Economic Theory* **91**(2) 223–247.

Fu, Shiming, R Vijay Krishna. 2019. Dynamic financial contracting with persistent private information. *The RAND Journal of Economics* **50**(2) 418–452.

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.

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

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.

Pavan, A., I. Segal, J. Toikka. 2014. Dynamic mechanism design: A myersonian approach. *Econometrica* **82**(2) 601–653.

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

Topkis, D.M. 1998. *Supermodularity and Complementarity*. Princeton University Press.