# Punish Underperformance with Resting – Optimal Dynamic Contracts in the Presence of Switching Cost

Ping Cao
School of Management, University of Science and Technology of China, Hefei, China, pcao@ustc.edu.cn

Peng Sun
The Fuqua School of Business, Duke University, Durham, NC, USA, psun@duke.edu

Feng Tian
Faculty of Business and Economics, University of Hong Kong, Pofulam Road, Hong Kong, fengtian@hku.hk

This paper studies a dynamic principal-agent setting in which the principal needs to dynamically schedule an agent to work or rest. When the agent is motivated to work, the arrival rate of a Poisson process increases, which increases the principal's payoff. Resting, on the other hand, serves as a threat to the agent by delaying future payments. A key feature of our setting is a switching cost whenever the agent stops resting and starts working. We formulate the problem as an optimal control model with switching, and fully characterize the optimal control policies under different parameter settings. Our analysis shows that when the switching cost is not too high, the optimal contract demonstrates a generalized control-band structure, and may involve randomly switching the agent to rest. The length of each resting episode, on the other hand, is fixed. Overall, the optimal contract is easy to describe, compute, and implement.

*Key words*: dynamic contract; jump process; optimal control; switching cost.

*History*: *This Version:* September 14, 2021.

## 1. Introduction

Designing dynamic contracts to manage incentives is an important and challenging problem. It often involves carefully scheduling "carrots and sticks" over time. In an environment where outcome is stochastically determined by an agent's unobservable effort, it is intuitive that rewards ("carrots"), often in the form of monetary payments, follow good performances. If the agent is cash-constrained or has limited liability, however, the principal cannot charge money from the agent for bad performances. Therefore, it may not be obvious how to design penalties ("sticks") when performance is bad and leverage them to achieve better contracts. Due to analytical challenges, many dynamic contract design models restrict the focus on contracts that induce agents to always exert effort (see, for example Demarzo and Sannikov 2006, Sannikov 2008, Biais et al. 2010, Myerson 2015, Sun and Tian 2018). In these situations, the principal with commitment power can use potential contract termination as a form of penalty. That is, the principal can terminate the agent, which stops all future payment opportunities, if the outcome has been bad for a long enough period of time. The threat of termination helps the principal to induce effort while saving costly rewards. However, termination itself may also be quite costly to the principal, especially in situations where a replacement agent is hard to find. In this paper, we focus on an alternative approach to penalize bad outcomes: asking the agent to take some rest. Note that contract termination is essentially asking the agent to rest forever and never work again, and therefore is a special case of our setting.

In case it is unclear why the principal can use rest as a punishment, here is the intuition. In order to motivate a rational agent to exert effort, which is a private action, the principal needs to pay rent, either in the form of an immediate payment or as a promise to be delivered later. During the period when the agent is directed to rest, on the other hand, the agent loses the rent income. Therefore, the lost income when outcome is bad can be used as a threat, which helps ensure that the agent is willing to exert effort whenever asked to. It is worth noting that asking the agent to rest also means that the principal cannot enjoy good outcome brought by the agent's effort. However, compared with losing the agent forever due to termination, it is often less costly for the principal to endure a short period of time without the agent's effort. Even though the intuition may be clear at this point *why* we may use resting to punish underperformance, deciding *how* to schedule resting episodes in a dynamic environment remains challenging. Therefore, we study the optimal scheduling of payments and resting episodes in a basic dynamic contract design model.

Consider a continuous time optimal contract design problem, in which a principal tries to incentivize an agent to increase the arrival rate of a Poisson process. We may think of the principal as a firm, the agent as a sales representative, and each arrival as a customer. Alternatively, the agent may be the research branch of a firm, and arrivals innovation breakthroughs. Whenever the agent exerts effort, the instantaneous arrival rate increases. However, the effort is costly to the agent and un-observable to the principal. Therefore, frequent arrivals are associated with good performances, and no arrival for a long period of time is bad. In this dynamic setting, when the two players' discount rates are the same, it is optimal for the principal, who has commitment power, to never ask the agent to rest before potential contract termination. Sun and Tian (2018) provide the corresponding optimal solution. If the principal is more patient than the agent, on the other hand, asking the agent to rest from time to time may be beneficial.

It is instructive to consider the intuition behind it. If the principal has to motivate the agent to continuously exert effort, contract termination serves as a threat, which helps to reduce overall payments to motivate effort. In particular, when an arrival does not occur for too long a period of time, the principal has to deliver the threat, despite knowing that the agent has been working hard. Termination is also undesirable to the principal, who will no longer has access to the higher arrival rate. If, instead, the principal can direct the agent to rest, the resting period delays future payments. Payment delay is particularly painful to an agent who is less patient than the principal, and therefore serves as a threat. Although the principal also has to bear a short-term loss due to the lower arrival rate during the resting period, such a short-term loss may still be preferable to permanently losing the agent. Therefore, in this paper, we study optimal contract design that involves scheduling resting episodes as well as payments and contract termination.

Despite the intuitive appeal, contract design with switching between working and resting is generally hard, which explains why many dynamic contract design models do not directly consider resting the agent. A standard approach is to first focus on the class of contracts that only motivate agent to always work, so that the contract design problem can be formulated as a continuous-time optimal control model. After obtaining the optimal contract in this restrictive class, the authors provide a sufficient condition on model parameters under which the optimal contract indeed falls into this restricted class (see, for example Demarzo and Sannikov 2006, Biais et al. 2010).

Zhu (2013) and Grochulski and Zhang (2019) study continuous time optimal contracts allowing shirking. They focus on settings where uncertain outcomes follow a Brownian motion, instead of a jump process as in our paper. Their optimal contract structures involve controls that constantly switch between working and shirking (a "sticky process"). Although these are nice mathematical results, from a managerial point of view, such a control/contract is not practical, because constantly switching between working and shirking must be quite costly in real life. Therefore, we include a fixed cost whenever the principal switches the agent from resting to working. When this switching cost approaches zero, we show that the switching frequency according to our optimal solution also approaches infinity. Therefore, including this switching cost is necessary for the model to be practically relevant, not only in the Browian motion settings, but also in our Poisson arrival setting.

Overall speaking, our analysis reveals that optimal contract structures demonstrate three possibilities depending on model parameters, as illustrated in the three regions of Figure 6 later in the paper. First, if the switching cost is higher than a threshold, it is optimal to never switch the agent from resting to working. In this case, it is simply too costly for the principal to hire the agent to start the work. The second possibility is when the revenue from each arrival is high enough. In this case it is optimal to motivate the agent to always work and not rest. This is also intuitive, because the high revenue per arrival means that principal does not want to let the agent rest and forfeit the higher arrival rate. The third possibility is when neither the switching cost nor the revenue is too high. In this case, the optimal contract demonstrates intricate and rich structures that we will explain next.

In order to explain the optimal contract structures, we describe the optimal contract in the context of the state space of the optimal control model. It is well known, following the dynamic contracting and repeated games literature, that the agent's total future utility, also called the *promised utility*, is a state variable (see, for example, Spear and Srivastava 1987, Abreu et al. 1990).

The optimal policy of many control systems with a fixed cost demonstrates a "control-band" policy structure. For example, the optimal stochastic inventory control policy with a fixed ordering cost has an $(s, S)$-policy structure (Zipkin 2000). Not surprisingly, our optimal contract also involves two thresholds of the promised utility, a lower $\underline{\theta}$ and a higher $\bar{\theta}$. However, compared with the inventory system, in which the state variable (inventory position) always moves in cycles between the two thresholds $s$ and $S$, the dynamics of our state variable is more complex. In particular, the promised utility does not always move between $\underline{\theta}$ and $\bar{\theta}$.

A total of four parameters determines the optimal contract structure. Besides the aforementioned thresholds $\underline{\theta}$ and $\bar{\theta}$, we also need to identify an upper bound $\widehat{w}$ and a lower bound $\breve{w}$ for the promised utility. If the agent has been working, the effort should continue as long as the promised utility is above the lower threshold $\underline{\theta}$. While working, the promised utility takes an upward jump upon each arrival, and continuously decreases between arrivals. An upward jump may bring the promised utility to be above $\bar{\theta}$, or even higher, to the upper bound $\widehat{w}$, which triggers payment. Therefore, frequent arrivals induce upward jumps resulting in rewards (payments). On the flip side, if an arrival does not occur for too long a period of time (bad performance) despite the agent's effort, the promised utility decreases to the lower threshold $\underline{\theta}$. At this point, the principal switches off effort and directs the agent to rest for a fixed period of time. At the end of the resting period, the agent's promised utility is reset to the upper threshold $\bar{\theta}$, when effort is switched on again. Overall, Figure 1 in the paper illustrates the general structure of the promised utility dynamics. The figure also involves a lower bound $\breve{w}$. If $\breve{w} > \underline{\theta}$, when the promised utility decreases to $\breve{w}$ while the agent is working, it stays there for an exponentially distributed random time until dropping to the lower threshold $\underline{\theta}$, unless an arrival triggers an upward jump before the end of the random time period. Regardless of whether $\breve{w}$ is strictly higher or equal to $\underline{\theta}$, the time it takes for the promised utility to increase from $\underline{\theta}$ to $\bar{\theta}$ is fixed. That is, the length of each resting episode is the same. In Sections 4 and 5, we provide complete descriptions on all the contract parameters $\underline{\theta}$, $\bar{\theta}$, $\widehat{w}$ and $\breve{w}$ given any set of model parameters. Furthermore, Appendix A.2 explains how to compute the optimal contract parameters.

If we decrease the switching cost to zero, generally speaking, the gap between thresholds $\underline{\theta}$ and $\bar{\theta}$ diminishes to zero, which implies that the frequency of switching between working and resting increases to infinity. As we have argued earlier, such a control policy is impractical, which further explains the necessity of introducing the positive switching cost in the model. Although a zero switching cost excludes practical contracts, the corresponding optimal value function yields an upper bound on the potential benefit of using resting as a punishment to bad performances. Numerical examples reported in Section 6 shows that this benefit can be non-trivial.

Dynamic moral hazard problem has been a subject of recent management science studies. In particular, Zorc et al. (2019) study a delegated search problem in a discrete-time dynamic environment. A key distinction of that paper is that the agent is risk-averse and can borrow from a bank to pay the principal. In comparison, we assume that the risk-neutral agent is cash-constrained and therefore payment only goes from the principal to the agent. Chen et al. (2020) studies "limited-term" non-monetary rewards contracts in order to induce agents' effort over the long-run. The model contains an adverse-selection component, and is focused on designing near-optimal "limited-term" stationary policies.

Recent decision analysis literature also includes studies of continuous-time games. The stream of papers Kwon et al. (2016), Kwon (2019), and Georgiadis et al. (2020) study continuous-time stochastic games of stopping-time decisions which are based on Brownian motion uncertainties. Continuous-time games studied in Zorc and Tsetlin (2020) and Hu and Tang (2021) do not include Brownian motion uncertainties, but consider richer decision spaces for the players. Unlike our paper, these game-theoretic papers do not focus on dynamic moral hazard issues.

Methodological break-throughs for continuous-time moral hazard problems start from Demarzo and Sannikov (2006) in the finance/economics literature. Earlier studies often use Brownian motion processes to model the underlying dynamics (see, for example, Demarzo and Sannikov 2006, Sannikov 2008, Cvitanić et al. 2016). To our knowledge, Biais et al. (2010) is the first to model underlying uncertainties as a jump process to capture "large risks." Myerson (2015) studies a similar model in a political economy setting with agent replacement.

Compared with contracts for Brownian motion based uncertainties, the optimal contract structure for jump processes is much easier to describe and implement. This is because the promised utility often takes discrete jumps at arrivals, and otherwise changes deterministically. (In contrast, under Brownian motion uncertainties, the promised utility evolves stochastically all the time.) This simplicity in the optimal contract structure makes the model appealing from a practical and managerial perspective. Sun and Tian (2018) and Cao et al. (2021) study optimal contracts that induce effort from an agent to increase the unobservable arrival rate of a point process. In particular, Cao et al. (2021) correctly identify the optimal contract within the restrictive class of contracts that motivate continued effort before termination when the two players discount rates are different. Tian et al. (2021) further extend the model to a two-state setting, where the agent exerts effort to either maintain or repair a machine, depending on which state the machine is in.

Also focusing on a point process, but to decrease the arrival rate, Chen et al. (2020) study optimal schedules to monitor (as well as pay) the agent. The end of that paper points out a connection between monitoring and shirking. That is, with a proper transformation, monitoring episodes in their optimal schedule correspond to shirking episodes in a corresponding model (without monitoring) that allows shirking. We believe that our results also speak to optimal contracts with monitoring for the case of increasing the arrival rate. In comparison, the optimal contracts in our paper demonstrate very different structures compared with those in Chen et al. (2020). We also need to model a fixed cost to be practical, as mentioned earlier. Tackling our problem requires different analysis, for example, variational-inequality-based optimality condition, which do not arise in Chen et al. (2020). Also trying to decrease the arrival rate of a Poisson process in a bank monitoring setting, Hernandez Santibanez et al. (2020) extend Pages (2013) and Pages and Possamai (2014), and study a model that involves both adverse selection and moral hazard while allowing shirking.

Another relevant literature is stochastic optimal control in the presence of switching cost, but not about contract design (see, for example, Brekke and Oksendal 1994, Duckworth and Zervos 2001, Vath and Pham 2007, Vath et al. 2008). Our work has two main differences with this literature. First, we consider a jump process, while the aforementioned papers are all based on diffusion processes. Second, the strategic interactions between the two players make our design and analysis more challenging than standard single-decision-maker control problems.

The remaining of the paper are organized as follows. We first introduce the model and a general description of the optimal contract in Section 2. Then in Section 3 we present the optimality conditions based on variational inequalities. Sections 4 and 5 contain the main results of the paper, which are the optimal contract structures under different model parameters. Next, in Section 6 we let the switching cost approach zero, which allows us to quantify the potential benefit of considering the resting option. We conclude the paper in Section 7. Further discussions, as well as proofs for all the results, are presented in the online appendices.

## 2. Model

Consider a continuous time principal-agent model. The principal faces a Poisson process of arrivals, each of which brings a revenue $R$ to the principal. Without the agent's work, the base arrival rate is $\underline{\mu}$. The agent is able to bring the arrival rate up to $\mu > \underline{\mu}$ if exerting effort,[1] which costs the agent $b$ per unit of time. (For simplicity, we consider binary effort levels, consistent with Biais et al. 2010). Following standard assumptions, the agent has limited liability and is cash-constrained. Therefore, the principal needs to pay the agent the cost $b$ whenever directing the agent to work. If the agent is directed to work but does not exert effort, the agent effectively receives a shirking benefit $b$. Effort is not observable to the principal, who needs to design a contract to motivate the agent's effort.

Whenever directing the agent to work, the principal needs to provide the corresponding work environment, such as offering office spaces, research labs, production equipment, or supporting personnel. There is a fixed cost $K > 0$ for the principal to set up the environment when the agent switches from resting to working. Think about this as the fixed cost related to restarting the lease for office spaces, reopening the lab, resetting production equipments, or recruiting personnel. (We will briefly discuss the case when stop working also incurs a cost in Appendix A.4.) Directing the agent to work may also involve additional costs to the principal, such as rents, maintenance fees, or personnel salaries. We denote $c$ to represent the principal's total cost rate whenever directing the agent to work, including the payment for the agent's effort cost, that is, $c \geqslant b$. Let $\mathcal{E}_t \in \{\mathsf{I}, \emptyset\}$ denote the working/resting state at time $t \geqslant 0$. In particular, state $\mathsf{I}$ ("on") represents the agent is working, while state $\emptyset$ ("off") means that the agent is resting. We use notation $\mathcal{E}_{0-} \in \{\mathsf{I}, \emptyset\}$ to represent the initial state before the dynamic contract starts. It is natural to assume that $\mathcal{E}_{0-} = \emptyset$. That is, at time zero, if the principal decides to hire the agent to start working, the fixed switching cost needs to be paid. We include $\mathcal{E}_{0-} = \mathsf{I}$ for completeness of the analysis.

We define $\Delta\mu = \mu - \underline{\mu}$, and make the following assumption.

Assumption 1. *$R\Delta\mu > c$.*

This is a standard assumption (see, for example, Equation (2) in Sun and Tian 2018), which ensures that exerting effort is socially optimal when the state is $\mathsf{I}$.

Both the principal and the agent are risk neutral and discount future cash flows. Discount rates are $r$ and $\rho$ for the principal and the agent, respectively, such that $0 < r \leqslant \rho$. That is, the principal is at least as patient as the agent. This paper is mostly focused on the case of $r < \rho$. (In Section A.3 we provide a rigorous proof for the claim in Sun and Tian (2018) that when $r = \rho$ it is optimal to motivate continued effort.)

Denote right-continuous point processes $N := \{N_t\}_{t \geqslant 0}$ and $S := \{S_t\}_{t \geqslant 0}$ to record the total number of arrivals and switchings, respectively, from time 0 to $t$. Define a filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t \geqslant 0}$ to capture all relevant public information up to any time $t$, such that $\mathcal{F}_t = \sigma(\mathcal{E}_{0-}; N_s, S_s : 0 \leqslant s \leqslant t)$. We need to include state switching information in the filtration because we allow randomization in its control. For completeness, we also define $\mathcal{F}_{0-} = \sigma(\mathcal{E}_{0-})$.

The principal has the commitment power to issue a long term *dynamic contract* $\Gamma$, consisting of a tuple $(L, D, q)$, defined as the following.

1. $L = \{L_t\}_{t \geqslant 0}$ is an $\mathcal{F}$-adapted process that tracks the principal's cumulative payment to the agent from time 0 to time $t$. In particular, at any time $t$, the payment can be an instantaneous payment $\Delta L_t$, or a flow with rate $\ell_t$, such that $\mathrm{d}L_t = \Delta L_t + \ell_t \mathrm{d}t$.[2] We assume that the agent is cash-constrained and has limited liability, that is, $\Delta L_t \geqslant 0$ and $\ell_t \geqslant 0$ for all $t \geqslant 0$.
2. $D = \{D_t\}_{t \geqslant 0}$ is an $\mathcal{F}$-adapted counting process which records the total number of switchings between "working" and "resting" up to time $t$. That is, these switchings are "deterministic"

with respect to $\mathcal{F}_t$. In order to have a rich enough class of control policies such that optimal values are attainable, we also need to allow random switchings as well, which comes next.

3. $q = \{q_t\}_{t \geqslant 0}$ is an $\mathcal{F}$-predictable switching intensity process, such that the probability of switching during a short time interval $[t, t+\delta]$ is $q_t\delta + o(\delta)$. Let $Q = \{Q_t\}_{t\geqslant 0}$ be the corresponding counting process that records the cumulative number of all the random switchings up to time $t$. Therefore, the total number of switchings by time $t$ is $S_t = D_t + Q_t$, which, together with the initial state $\mathcal{E}_{0-}$, identifies the state at any time $t \geqslant 0$. In order to establish our optimality results, we need the following technical condition on the switching intensity:

$$\mathbb{E}\left[\int_0^\infty q_t e^{-rt} \mathrm{d}t\right] < \infty. \tag{1}$$

With these notations, we can more rigorously define the $\mathcal{F}$-adapted state process $\mathcal{E} := \{\mathcal{E}_t\}_{t\geqslant 0}$, such that $\mathcal{E}_t \neq \mathcal{E}_{t-}$ if and only if $\mathrm{d}S_t = 1$. Furthermore, state switchings may also include the possible termination of contract, which is the last time that the principal changes the state from working ($\mathsf{I}$) to resting ($\emptyset$), either deterministically or randomly.

Due to limited liability, we need the following constraints for our contracts $\Gamma$, which states that effort cost $b$ needs to be reimbursed in real time,

$$\ell_t \geqslant b\mathbb{1}_{\mathcal{E}_t = \mathsf{I}}, \ \forall t \geqslant 0. \tag{LL}$$

Further denote a right continuous process $\nu = \{\nu_t\}_{t\geqslant 0}$ to represent the agent's effort level over time. In particular, $\nu_t = \mu$ and $\nu_t = \underline{\mu}$ represent that the agent is working and resting at time $t$, respectively. Under a general contract, the agent may not follow the effort process directed by the principal. In fact, it is easy to make sure that the agent follows the direction to rest, by setting $\ell_t = 0$ when $\nu_t = \underline{\mu}$. In this case, the agent cannot afford to work when directed not to. Therefore, any effort process $\nu$ that is *admissible to contract* $\Gamma$ must satisfy $\nu_t = \underline{\mu}$ whenever $\mathcal{E}_t = \emptyset$.

### 2.1. The Agent's Utility and Incentive-Compatible Contracts

Given a dynamic contract $\Gamma = (L, D, q)$ and an effort process $\nu$ starting from state $\mathcal{E}_{0-}$, the expected discounted utility of the agent is

$$u(\Gamma, \nu, \mathcal{E}_{0-}) = \mathbb{E}^{\nu,q}\left[\int_0^\infty e^{-\rho t}(\mathrm{d}L_t - b\mathbb{1}_{\nu_t=\mu}\mathrm{d}t)\middle| \mathcal{E}_{0-}\right], \tag{2}$$

in which $\mathbb{E}^{\nu,q}$ represents expectation taken with respect to the switching intensity process $q$ in $\Gamma$, and arrival rates induced by the effort process $\nu$. For simplicity of notations, when there is no ambiguity, we omit this superscript.

A designed contract needs to induce the agent to follow directions on when to work and rest. Formally, define a "complying effort process" $\bar{\nu}(\Gamma) = \{\bar{\nu}_t\}_{t\geqslant 0}$ for contract $\Gamma$, such that $\bar{\nu}_t = \mu$ if $\mathcal{E}_t = \mathsf{I}$, and $\bar{\nu}_t = \underline{\mu}$ if $\mathcal{E}_t = \emptyset$, at any time $t$. A contract $\Gamma$ is said to be *incentive compatible* (IC) if

$$u(\Gamma, \bar{\nu}(\Gamma), \mathcal{E}_{0-}) \geqslant u(\Gamma, \nu, \mathcal{E}_{0-}) \text{ for any effort process } \nu \text{ admissible to } \Gamma \text{ and initial state } \mathcal{E}_{0-}. \quad (3)$$

That is, under IC contracts, the agent has the incentive to exert effort whenever directed to do so.

Further define the agent's continuation utility at any time $t \in [0, \infty) \cup \{0-\}$ conditional on $\mathcal{F}_t$ as[3]

$$W_t(\Gamma, \nu) = \mathbb{E}\left[\int_{t+}^{\infty} e^{-\rho(s-t)}(\mathrm{d}L_s - b\mathbb{1}_{\nu_s=\mu}\mathrm{d}s) \,\middle|\, \mathcal{F}_t\right]. \quad (4)$$

Therefore, $W_t\big(\Gamma, \bar{\nu}(\Gamma)\big)$ is the agent's continuation utility at time $t$ following the principal's directions, which is often referred to as the *promised utility* (see, for example, Biais et al. 2010). It is convenient to introduce the notation $W_{t-}(\Gamma, \nu) = \lim_{s\uparrow t} W_s(\Gamma, \nu)$ to represent the left-hand limit of the process $W(\Gamma, \nu) = \{W_t(\Gamma, \nu)\}_{t\geqslant 0}$ at $t > 0$. That is, $W_t(\Gamma, \nu)$ is the agent' continuation utility after observing either an arrival or a random switching that occurs at time $t$, while $W_{t-}(\Gamma, \nu)$ is the continuation utility evaluated before obtaining this knowledge. Following these definitions, given the initial state $\mathcal{E}_{0-}$, the $\mathcal{F}_{0-}$-measurable $W_{0-}(\Gamma, \nu)$ takes the value $u(\Gamma, \nu, \mathcal{E}_{0-})$.

Following standard contract theory assumptions, the agent is not required to stay in the contract. Hence, assuming the agent's outside option is normalized to value 0, we impose the following participation (also called the *individual rationality*, IR) constraint

$$W_t(\Gamma, \nu) \geqslant 0, \ \ \forall t \in [0, \infty) \cup \{0-\}. \quad \text{(IR)}$$

Furthermore, we assume that for any contract $\Gamma$ under our consideration, the agent's promised utility $W_t$ is upper bounded. That is, there exists a large enough $\bar{W}$ such that

$$W_t(\Gamma, \nu) \leqslant \bar{W} < \infty, \ \ \forall t \in [0, \infty) \cup \{0-\}. \quad \text{(WU)}$$

This constraint essentially captures the reality that the principal cannot keep delaying payments while pushing the agent's promised utility to infinity. The specific choice of the upper bound $\bar{W}$ is not important, as long as it is high enough such that constraint (WU) is not binding at optimality. Technically, we need this constraint to establish that a process related to $W_t(\Gamma, \nu)$ is a martingale, in the proof of Theorem 1 that comes later in the paper.

The following proposition provides the evolution of the agent's continuation utility process $W_t(\Gamma, \nu)$, which is often called the *promise keeping* (PK) condition in the dynamic contract literature (see, for example, Equation (B.8) of Sun and Tian 2018). The proposition also contains an equivalent recursive representation of incentive compatibility, following Biais et al. (2010).

Proposition 1. (i) *For any contract* $\Gamma$ *and agent's effort process* $\nu$, *there exists* $\mathcal{F}$*-predictable processes* $H(\Gamma,\nu)$ *and* $H^q(\Gamma,\nu)$ *such that*[4]

$$\begin{aligned}\mathrm{d}W_t(\Gamma,\nu) =& [\rho W_{t-}(\Gamma,\nu) + b\mathbb{1}_{\nu_t=\mu} - H_t(\Gamma,\nu)\nu_t + q_t H_t^q(\Gamma,\nu)]\mathrm{d}t \\ & - \mathrm{d}L_t + H_t(\Gamma,\nu)\mathrm{d}N_t - H_t^q(\Gamma,v)\mathrm{d}Q_t, \quad t>0. \end{aligned} \tag{PK}$$

*Furthermore,* (IR) *implies that*

$$H_t(\Gamma,\nu) \geqslant -W_{t-}(\Gamma,\nu) \text{ and } H_t^q(\Gamma,\nu) \leqslant W_{t-}(\Gamma,\nu), \ \forall t>0. \tag{5}$$

(ii) *Define* $\beta := b/\Delta\mu$. *Contract* $\Gamma$ *being incentive compatible is equivalent to*

$$H_t\big(\Gamma,\bar{\nu}(\Gamma)\big) \geqslant \beta \text{ if and only if } \mathcal{E}_t = \mathsf{I}. \tag{IC}$$

For notational convenience, we omit $(\Gamma,\nu)$ from processes $W_t$, $H_t$ and $H_t^q$ when $\nu$ is the complying effort process $\bar{\nu}(\Gamma)$. Part (i) of Proposition 1 specifies the dynamics of the agent's promised utility over time. In particular, $H_t$, if positive, is the magnitude of an *upward* jump at time $t$ if there is an arrival at that time. If it is negative then the jump is downward. In contrast, $H_t^q$, if positive, is the magnitude of a *downward* jump at time $t$ if there is a random switching. Condition (5) ensures that $W_t$ remains nonnegative after all these jumps. The reason why we set $H_t$ to capture upward jumps is because increasing the promised utility with an upward jump after an arrival serves as a reward to induce effort. Although we allow $H_t$ to be negative, later we show that it is always nonnegative under the optimal contract. In comparison, a random switching to resting (or termination) is a punishment, which is associated with a downward jump of promised utility with magnitude $H_t^q$. Finally, the (IC) condition is only required for state $\mathsf{I}$, because in state $\emptyset$ the principal can induce compliance by simply setting payment to zero.

Denote $\mathfrak{C}$ to represent the set of contracts that satisfy (LL), and yield a promised utility process $\{W_t\}_{t\in[0,\infty)\cup\{0-\}}$ that satisfies (PK), (IC), (IR), and (WU). Our contract design problem maximizes the principal's utility over the set $\mathfrak{C}$ of contracts. Therefore, we introduce the principal's utility next.

### 2.2. Principal's Utility

The principal's utility under any contract $\Gamma \in \mathfrak{C}$ is

$$U(\Gamma,\mathcal{E}_{0-}) = \mathbb{E}^{\bar{\nu}(\Gamma)}\left[\int_0^\infty e^{-rt}\left[R\mathrm{d}N_t - \mathrm{d}L_t - (c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}}\mathrm{d}t\right] - \sum_{0\leqslant t\leqslant\infty} e^{-rt}\kappa(\mathcal{E}_{t-},\mathcal{E}_t)\middle|\,\mathcal{E}_{0-}\right],$$

where we introduce notation $\kappa(\mathcal{E}_{t-},\mathcal{E}_t)$ to represent the switching cost when the principal changes the working/resting state from $\mathcal{E}_{t-}$ to $\mathcal{E}_t$, such that $\kappa(\emptyset,\mathsf{I}) = K$, and $\kappa(\mathsf{I},\emptyset) = \kappa(\emptyset,\emptyset) = \kappa(\mathsf{I},\mathsf{I}) = 0$.

Within the integral, the term $R\mathrm{d}N_t$ represents the revenue from arrivals; $\mathrm{d}L_t$ is the payment cost, which satisfies (LL); and $(c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}}$ captures the cost rate of directing the agent to work, in addition to reimbursing the effort cost $b$ already included in the payment term $\mathrm{d}L_t$.

Our optimal contract design problem can be succinctly formulated as the following optimization,

$$\mathcal{Z}(\mathcal{E}_{0-}) := \max_{\Gamma\in\mathfrak{C}} U(\Gamma, \mathcal{E}_{0-}). \tag{6}$$

If the agent is ever terminated, the principal's total expected utility after the termination is

$$\underline{v} := \frac{\mu R}{r}, \tag{7}$$

which is also the base-line total expected revenue that the principal collects without hiring the agent.

As an example of feasible contracts in $\mathfrak{C}$, we first consider a special contract $\bar{\Gamma}$, which directs the agent always to work, and pays the agent $\beta$ for each arrival. It is clear that contract $\bar{\Gamma}$ satisfies all the aforementioned constraints for $\mathfrak{C}$. Under such a contract, the agent's promised utility $W_t$ stays as a constant,

$$\bar{w} := \frac{\beta\mu}{\rho}. \tag{8}$$

Furthermore, it is easy to verify that the principal's utility under contract $\bar{\Gamma}$ starting from state $\mathcal{E}_{0-} = \mathsf{I}$ is

$$U(\bar{\Gamma}, \mathsf{I}) = \frac{R\mu - c - \beta\mu}{r}.$$

We define the corresponding societal utility between the principal and agent as

$$\bar{V} := U(\bar{\Gamma}, \mathsf{I}) + \bar{w} = \frac{R\mu - c - (\rho - r)\bar{w}}{r}. \tag{9}$$

More generally, we define

$$\bar{V}(w) := \frac{R\mu - c - (\rho - r)w}{r}, \tag{10}$$

such that $\bar{V}(\bar{w}) = \bar{V}$. Later in the paper we show that contract $\bar{\Gamma}$ is optimal when the revenue per arrival $R$ is high enough. The value $\bar{w}$ and function $\bar{V}(w)$ are also useful in characterizing the optimal contracts and value functions.

## 2.3. An Overview of General Optimal Contract Structures

We now define a general class of dynamic contracts, which involves a control-band structure and potential randomization. In the rest of the paper, we show that under different model parameters, the optimal contract takes various special cases of this general class of contracts. Specifically, we have the following definition.

DEFINITION 1. For any $w_0, \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w} \in [0, \bar{w}]$, such that $\underline{\theta} \leqslant \breve{w} \leqslant \bar{\theta} \leqslant \widehat{w}$, and any state $\varepsilon_0 \in \{\mathsf{I}, \emptyset\}$, define contract $\Gamma^*(w_0, \varepsilon_0; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w}) = (L^*, D^*, q^*)$ as follows.

(i) The dynamics of the agent's promised utility $W_t$ follows $W_0 = w_0 \wedge \widehat{w}$, and

$$\begin{aligned} \mathrm{d}W_t = &\left\{-\rho\,(\bar{w} - W_{t-})\,\mathrm{d}t\mathbb{1}_{W_{t-}\in(\breve{w},\widehat{w}]} - (\breve{w} - \underline{\theta})\,\mathrm{d}Q_t + [(\widehat{w} - W_{t-}) \wedge \beta]\,\mathrm{d}N_t\right\}\mathbb{1}_{\mathcal{E}_{t-}=\mathsf{I}} \\ &+ \rho W_{t-}\mathrm{d}t\mathbb{1}_{\mathcal{E}_{t-}=\emptyset}, \end{aligned} \tag{11}$$

in which we use notation $a \wedge b$ to represent $\min\{a, b\}$ for any $a, b \in \mathbb{R}$, and the point process $\{Q_t\}_{t\geqslant 0}$ represents the number of random switchings, following intensity

$$q_t^* = \frac{\rho(\bar{w} - \breve{w})}{\breve{w} - \underline{\theta}}\mathbb{1}_{W_{t-}=\breve{w}}, \text{ if } \breve{w} > \underline{\theta}. \tag{12}$$

(ii) The payment to the agent follows $\mathrm{d}L_0^* = (w_0 - \widehat{w})^+$ and, for $t > 0$,

$$\mathrm{d}L_t^* = [(W_{t-} + \beta - \widehat{w})^+\mathrm{d}N_t + b\mathrm{d}t]\,\mathbb{1}_{\mathcal{E}_{t-}=\mathsf{I}}. \tag{13}$$

(iii) The "determinisitc" switching $D^*$ follows $\mathrm{d}D_0^* = 1$ (switching) if and only if $\mathcal{E}_{0-} \neq \varepsilon_0$, and, for $t > 0$,

$$\mathrm{d}D_t^* = \mathbb{1}_{W_{t-}=\bar{\theta} \text{ and } \mathcal{E}_{t-}=\emptyset} + \mathbb{1}_{W_{t-}=\underline{\theta} \text{ and } \mathcal{E}_{t-}=\mathsf{I}}. \tag{14}$$

Following Definition 1, all the components $L^*$, $D^*$, and $q^*$ of the contract $\Gamma^*$, as well as the promised utility process $\{W_t\}_{t\geqslant 0}$, are completely determined by parameters $(w_0, \varepsilon_0, \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w})$. In particular, (11) indicates that the promised utility generally decreases with a slope $\rho(\bar{w} - W_{t-})$ when $W_{t-} \in (\breve{w}, \widehat{w}]$ and the agent is directed to work, except when there is an arrival ($\mathrm{d}N_t = 1$), which triggers an upward jump of magnitude $(\widehat{w} - W_{t-}) \wedge \beta$. This implies that the promised utility is never above $\widehat{w}$. When the promised utility decreases to $\breve{w}$, it stays at that level until either an arrival ($\mathrm{d}N_t = 1$) or a random switching of state ($\mathrm{d}Q_t = 1$) occurs. According to (12), the random switching only occurs if $\breve{w} > \underline{\theta}$ and when the promised utility is at $\breve{w}$. When random switching happens, (11) further indicates that the promised utility takes a downward jump from $\breve{w}$ to $\underline{\theta}$. Furthermore, the last term of (11) indicates that when the agent is directed to rest, the promised utility keeps increasing at rate $\rho W_{t-}$ regardless of whether there are arrivals. The increasing rate corresponds to accrued interests if we consider the promised utility as a bank account balance.

According to (13), payment only occurs when the agent is directed to work. Besides reimbursing the effort cost ($b\mathrm{d}t$), the principal only pays the agent when an arrival occurs and the current promised utility is within $\beta$ below $\widehat{w}$. The instantaneous payment, $W_{t-} + \beta - \widehat{w}$, plus the corresponding upward jump in (11), $\widehat{w} - W_{t-}$, is exactly $\beta$.

Finally, (14) implies that the principal directs the agent to stop working when the promised utility decreases to $\underline{\theta}$, and start working again when the promised utility increases to $\bar{\theta}$. Therefore,

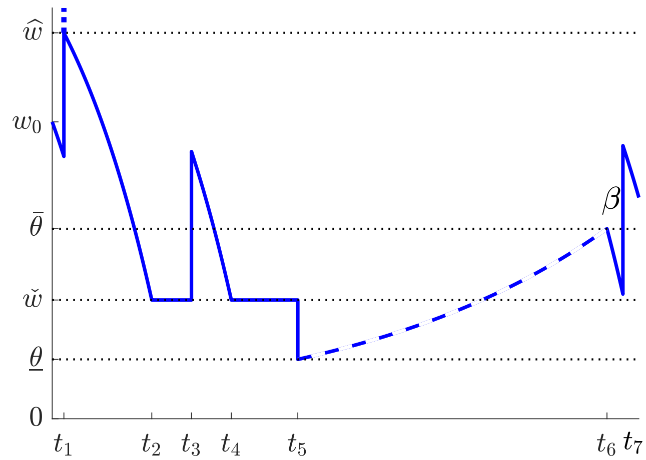



**Figure 1 A sample trajectory for the agent's promised utility according to $\Gamma^*(w_0, \varepsilon_0; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w})$ with $\rho = 0.5$, $r = 0.2$, $R = 2$, $\mu = 2$, $\Delta\mu = 1.2$, $c = b = 0.3$ where we set $w_0 = 0.5$, $\varepsilon_0 = \mathsf{I}$, $\underline{\theta} = 0.1$, $\breve{w} = 0.2$, $\bar{\theta} = 0.32$, and $\hat{w} = 0.65$. The dotted line depicts the payment.**

if $\underline{\theta} = \breve{w}$, there is no random switching, and the switching policy is similar to the traditional "control band policy" between the two thresholds $\underline{\theta}$ and $\bar{\theta}$.

Figure 1 depicts the dynamics of the promised utility following a general contract of Definition 1. As we can see, the agent is working at time 0, and the promised utility starts at $w_0$ and gradually decreases until the first arrival at time $t_1$. At this point in time, an upward jump of $\beta$ would take the promised utility above $\widehat{w}$. Therefore, the promised utility instead jumps to $\widehat{w}$, and the principal pays the agent $W_{t_1-} + \beta - \widehat{w}$ (depicted by the dotted line at $t_1$). No further arrival occurs until time $t_2$, when the promised utility reaches $\breve{w}$. From this point on, the promised utility stays the same at $\breve{w}$, while a random switching occurs with rate $q^* = \rho(\bar{w} - \breve{w})/\breve{w}$. At time $t_3$, there is an arrival before a random switching occurs, which causes the promised utility to jump up by $\beta$. The promised utility decreases to $\breve{w}$ again at $t_4$ and the random switching occurs at time $t_5$, which brings the promised utilty to $\underline{\theta}$. At this point the agent is directed to rest, until the promised utility increases to $\bar{\theta}$ at $t_6$. There may be arrivals between $t_5$ and $t_6$ if $\underline{\mu} > 0$, but these arrivals do not affect the dynamics of the contract. At time $t_6$, the principal (deterministically) switches the agent to working again. Time epoch $t_7$ sees another arrival, which triggers the promised utility to jump up by $\beta$.

Remark 1. Note that directing the agent to rest serves as a type of punishment. Before the promised utility decreases to the threshold $\underline{\theta}$, any arrival brings an upward jump in the promised utility, which makes the agent closer to getting paid, if not already being paid. However, as soon as the agent is directed to rest, it takes a *fixed period of time* with length $(\ln \bar{\theta} - \ln \underline{\theta})/\rho$ for work to resume. Because the effort cost is reimbused, from the agent's point of view, the only difference between working and resting is that working brings potential rent payment, while resting delays future rent payments for a period of time. Therefore, resting serves as a threat to the agent, who is less patient than the principal ($\rho > r$). If the lower threshold $\underline{\theta}$ is 0, the length of the resting time period becomes infinity. That is, directing the agent to rest is equivalent to terminating the contract. □

Following Definition 1, if $\widehat{w} = \bar{w}$, upon reaching $\bar{w}$, the promised utility does not decrease any more, and the agent is paid $\beta$ for each future arrival. (Figure 1, on the other hand, depicts the case that $\widehat{w} < \bar{w}$, and $\breve{w} > \underline{\theta} > 0$.) Therefore, after reaching $\bar{w}$ the contract becomes the aforementioned $\bar{\Gamma}$. In fact, contract $\bar{\Gamma}$ can be expressed as a special case of $\Gamma^*(w_0, \varepsilon_0; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w})$, such that

$$\bar{\Gamma} = \Gamma^*(\bar{w}, \mathsf{I}; 0, 0, \bar{w}, \bar{w}). \tag{15}$$

That is, $w_0 = \bar{\theta} = \widehat{w} = \bar{w}$, and $\underline{\theta} = \breve{w} = 0$.

If $\underline{\theta} = \breve{w} > 0$, there is no randomized switching, and contract $\Gamma^*(w_0, \varepsilon_0; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w})$ demonstrates a "control band" structure, where the promised utility is moving between $\underline{\theta}$ and $\bar{\theta}$, unless an arrival triggers an upward jump to carry the promised utility to $(\bar{\theta}, \widehat{w}]$. In this case the agent is never terminated, as long as $w_0 > 0$.

If $\underline{\theta} = 0$, then following contract $\Gamma^*(w_0, \varepsilon_0; 0, \breve{w}, \bar{\theta}, \widehat{w})$, whenever the state switches to $\emptyset$, the promised utility must have hit $\underline{\theta} = 0$. At this point the contract is terminated.

Another special case is not to hire the agent from the beginning, or,

$$\underline{\Gamma} := \Gamma^*(0, \emptyset; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w}). \tag{16}$$

In this case, the agent's promised utility starts at $w_0 = 0$, and never climbs to be positive according to (11). Therefore, the specific values of $\underline{\theta}$, $\breve{w}$, $\bar{\theta}$, and $\widehat{w}$ do not matter.

Later in the paper we see that contracts $\bar{\Gamma}$, $\underline{\Gamma}$, and other special cases of the general contract structure $\Gamma^*(w_0, \varepsilon_0; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w})$ could be optimal under different model parameter settings.

Before we close this section, we have the following result which implies that if the contract $\Gamma^*$ starts the continuation utility at $w_0$, then it delivers the agent a total utility $w_0$.

LEMMA 1. *For any $\varepsilon_0 \in \{\mathsf{I}, \emptyset\}$, $\mathcal{E}_{0-} \in \{\mathsf{I}, \emptyset\}$, and $\underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w}$ such that $0 \leqslant \underline{\theta} \leqslant \breve{w} \leqslant \bar{\theta} \leqslant \widehat{w} \leqslant \bar{w}$, we have*

$$u\left(\Gamma^*(w_0, \varepsilon_0; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w}), \bar{\nu}, \mathcal{E}_{0-}\right) = w_0, \ \forall w_0 \geqslant 0. \tag{17}$$

In order to specify the optimal contract, we need to identify the initial promised utility $w_0$.

## 3. Optimality Conditions

In this section, we first provide an optimality condition, in the form of quasi-variational inequalities, and show that any function that satisfies these conditions must yield an upper bound of the optimal value $\mathcal{Z}(\mathcal{E}_{0-})$ defined in (6). In the next two sections we provide dynamic contracts that achieve these upper bounds, which implies that they are optimal.

We claim that the optimal value function is concave, although it may not be differentiable. Therefore, denote $\mathbb{CA}$ to be the set of all continuous concave functions defined on $\mathbb{R}_+$. It is worth noting that any continuous concave function is differentiable except on a countable set of points. If a function $f \in \mathbb{CA}$ is not differentiable at point $w \geqslant 0$, we abuse notation and use $f'(w)$ to represent its left-derivative at $w$. Define operators $\mathcal{A}_{\mathsf{I}}$ and $\mathcal{A}_{\emptyset}$ that map a function $f \in \mathbb{CA}$ to functions $\mathcal{A}_{\mathsf{I}} f$ and $\mathcal{A}_{\emptyset} f$, respectively, such that for all $w \geqslant 0$,

$$(\mathcal{A}_{\mathsf{I}} f)(w) := (\mu + r) f(w) - \mu f(w+\beta) + \rho(\bar{w} - w) f'(w) - (\mu R - c) + (\rho - r) w, \text{ and} \tag{18}$$

$$(\mathcal{A}_{\emptyset} f)(w) := r f(w) - \rho w f'(w) + (\rho - r) w - R\underline{\mu}. \tag{19}$$

Equipped with these notations, we are ready to present the following Verification Theorem.

THEOREM 1. *Suppose there exists a pair of nondecreasing functions $V_{\mathsf{I}}$ and $V_{\emptyset}$ in $\mathbb{CA}$, such that*

$$(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) \geqslant 0, \ (\mathcal{A}_{\emptyset} V_{\emptyset})(w) \geqslant 0, \tag{20}$$

$$0 \leqslant V_{\mathsf{I}}(w) - V_{\emptyset}(w) \leqslant K, \ \textit{and} \tag{21}$$

$$V_{\mathsf{I}}(0) \geqslant \underline{v}, \ V_{\emptyset}(0) \geqslant \underline{v}, \tag{22}$$

*for any $w \in \mathbb{R}_+$. Then, for any contract $\Gamma \in \mathfrak{C}$, value $w \in [0,\infty)$ and initial state $\mathcal{E}_{0-} \in \{\mathsf{I}, \emptyset\}$, such that $u(\Gamma, \bar{\nu}(\Gamma), \mathcal{E}_{0-}) = w$, we have*

$$U(\Gamma, \mathcal{E}_{0-}) \leqslant V_{\mathcal{E}_{0-}}(w) - w.$$

*Therefore, we have*

$$\max_{w \in [0,\infty)} \left\{ V_{\mathcal{E}_{0-}}(w) - w \right\} \geqslant \mathcal{Z}(\mathcal{E}_{0-}).$$

Theorem 1 indicates that $V_{\mathsf{I}}(w) - w$ and $V_{\emptyset}(w) - w$ are upper bounds for the principal's utility under any contract in $\mathfrak{C}$ that yields an agent's utility $w$ when the initial state is $\mathsf{I}$ and $\emptyset$, respectively. Therefore, we can interpret functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ as societal value functions that contain both the principal and the agent's utilities, as long as they are attainable according to certain contracts. The quasi-variational inequality based optimality condition (20)–(22) may not appear intuitive. Therefore, in Appendix A.1, we provide a heuristic derivation, which reveals how we obtain these conditions. In a nutshell, the condition (20) describes the shape of the value functions; the condition

(21) reflects that switching from resting to working costs $K$; and the condition (22) captures the intuition that without the agent (the agent's promised utility $w=0$), the societal value is $\underline{v}$ as defined in (7).

If both functions $V_{\mathsf{l}}$ and $V_{\emptyset}$ are differentiable on $\mathbb{R}_+$, then this result is a classic verification theorem, which is extensively used in the optimal control literature. In fact, a typical method of obtaining an optimal control policy, which is called the "guess-and-verify" approach, consists of two main steps. In the first step, we guess a function and use the verification theorem to establish that this function is an upper bound of the optimal value function. In the second step, we propose a control policy that yields an objective value reaching this upper bound, which implies that the proposed control policy is optimal. It is worth mentioning that the control space in our dynamic contract problem includes potentially randomized control, which is rich enough to achieve the corresponding upper bound.

In order to identify optimal value functions, it is worth considering functions that satisfy conditions (20) and (22) with equality. First consider the resting state $\emptyset$. A value function $V$ that satisfies $V(0)=\underline{v}$ and the ordinary differential equation $(\mathcal{A}_{\emptyset}V)(w)=0$ must have the following form

$$\mathsf{V}_c(w)=\underline{v}+w+cw^{r/\rho}, \tag{23}$$

for some constant $c$. Later in the paper we show that the resting state's value function indeed takes this form under certain model parameter settings and for certain $w$ values.

Next, we consider the working state $\mathsf{l}$. In particular, consider a generic function $V_{\widetilde{w}}$ that is differentiable on $[0,\widetilde{w}]$ for some $\widetilde{w}\leqslant\bar{w}$, takes a constant value for $w\geqslant\widetilde{w}$, and satisfies (20) with $(\mathcal{A}_{\mathsf{l}}V)(w)=0$. That is, $V_{\widetilde{w}}$ satisfies the following differential equation,

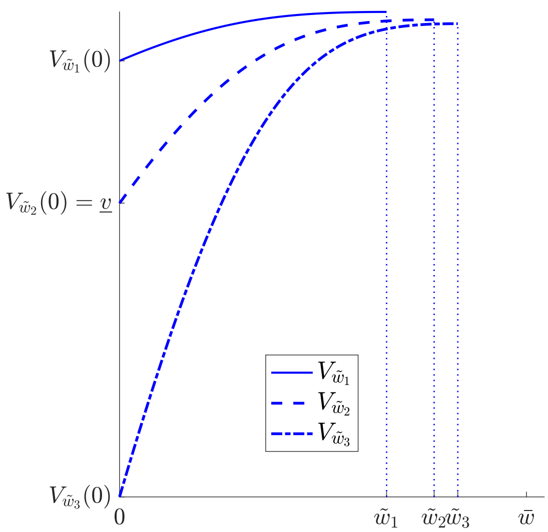
$$(\mu+r)V_{\widetilde{w}}(w)-\mu V_{\widetilde{w}}\big((w+\beta)\wedge\widetilde{w}\big)+\rho(\bar{w}-w)V'_{\widetilde{w}}(w)-(\mu R-c)+(\rho-r)w=0, \tag{24}$$

with boundary condition,

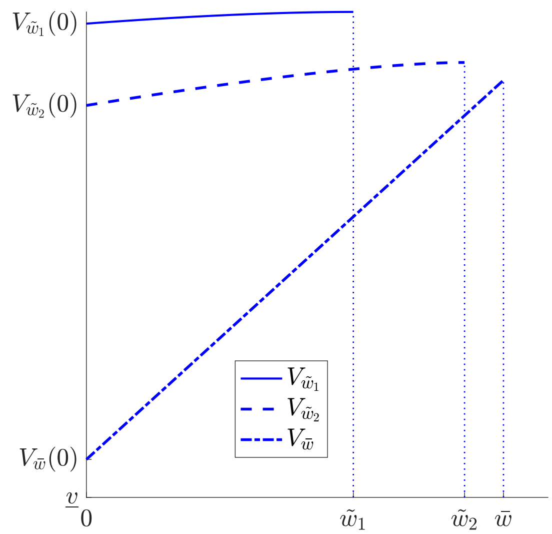
$$V_{\widetilde{w}}(w)=\bar{V}(\widetilde{w}),\ \forall w\geqslant\widetilde{w}, \tag{25}$$

in which $\bar{V}(\cdot)$ is defined in (10). Lemma 6 in Appendix B.2 establishes the existence and uniqueness of function $V_{\widetilde{w}}(w)$, and summarizes its key properties.

We also desire the boundary condition $V_{\widetilde{w}}(0)=\underline{v}$. Figure 2(a) demonstrates an example in which we can find a $\widetilde{w}$ value such that this boundary condition holds. As we can see, if we increase $\widetilde{w}$ to take three different values, $\tilde{w}_1$, $\tilde{w}_2$, and $\tilde{w}_3$, the entire function decreases as $\widetilde{w}$ increases, consistent with Lemma 6. This implies that at $w=0$, we have $V_{\tilde{w}_1}(0)>V_{\tilde{w}_2}(0)>V_{\tilde{w}_3}(0)$. In particular, we can identify a particular $\widetilde{w}=\tilde{w}_2$ such that $V_{\tilde{w}_2}(0)=\underline{v}$. Following Lemma 6, this situation corresponds to model parameters that satisfy the following condition.




(a) Value functions under Condition 1




(b) Value functions under Condition 2

**Figure 2** **(a): Value functions with** $r=0.2$, $\rho=0.5$, $c=0.2$, $R=2$, $\Delta\mu=0.7$, **and** $\mu=2$. **In this case,** $\bar{w}=1.14$ **and we let** $\tilde{w}_1=0.6$, $\tilde{w}_2=0.9$, **and** $\tilde{w}_3=1$. **(b): Value functions with** $r=0.2$, $\rho=1.2$, $c=b=1$, $R=5$, $\Delta\mu=0.8$, **and** $\mu=0.9$. **In this case,** $\bar{w}=0.94$, **and we let** $\tilde{w}_1=0.6$, $\tilde{w}_2=0.85$, **and** $\tilde{w}_3=\bar{w}$.

Condition 1.

$$\mu \geqslant \rho - r \text{ or } R < \hat{R} := \left[\frac{c}{b} + \frac{(\rho-r)(\rho-\mu)\mu}{\Delta\mu(\rho-r-\mu)\rho}\right]\beta.$$

However, in general we may not be able to find a value $\tilde{w} \leqslant \bar{w}$ to satisfy the boundary condition $V_{\tilde{w}}(0)=\underline{v}$. Figure 2(b), for example, depicts another model parameter setting such that as we increase $\tilde{w}$ to approach $\bar{w}$, the corresponding limiting value $V_{\bar{w}}(0)$ is always higher than $\underline{v}$. In this case we cannot use (24)–(25) to determine the optimal value function. This situation corresponds to model parameters that follow the next condition, opposite to Condition 1.[5]

Condition 2.

$$\mu < \rho - r \text{ and } R \geqslant \hat{R}.$$

Generally speaking, the solution $V_{\tilde{w}}$ that satisfies (24) with boundary condition (25) may not be concave. Therefore, we may need to construct a concave value function according to the following result.

Lemma 2. *For any* $\tilde{w} \in (0, \bar{w})$, *consider the function* $V_{\tilde{w}}$ *that uniquely solves* (24)–(25).

(i) *There exists a* $\tilde{w}$*-dependent threshold* $\breve{w}(\tilde{w}) \in [0, \tilde{w})$, *such that* $V''_{\tilde{w}}(w) < 0$ *over* $w \in (\breve{w}(\tilde{w}), \tilde{w})$ *and* $V''_{\tilde{w}}(w) > 0$ *over* $w \in [0, \breve{w}(\tilde{w}))$. *Moreover, we have* $\breve{w}(\tilde{w}) \leqslant (1 - r/\rho)\beta$.

(ii) *Define function*

$$\mathcal{V}_{\tilde{w}}(w) := \begin{cases} V_{\tilde{w}}(\breve{w}(\tilde{w})) + V'_{\tilde{w}}(\breve{w}(\tilde{w})) \cdot (w - \breve{w}(\tilde{w})), & w \in [0, \breve{w}(\tilde{w})), \\ V_{\tilde{w}}(w \wedge \tilde{w}), & w \in [\breve{w}(\tilde{w}), \infty). \end{cases}$$

*Function $\mathcal{V}_{\tilde{w}}(w)$ is increasing and concave in $w$ on $[0,\infty)$.*

(iii) *Fixing any $\tilde{w}_1$ and $\tilde{w}_2$ with $0 < \tilde{w}_1 < \tilde{w}_2 < \bar{w}$, we have*

$$\mathcal{V}_{\tilde{w}_1}(w) > \mathcal{V}_{\tilde{w}_2}(w), \text{ and } \mathcal{V}'_{\tilde{w}_1}(w) < \mathcal{V}'_{\tilde{w}_2}(w), \forall w \in [0, \tilde{w}_1).$$

Therefore, if function $V_{\tilde{w}}$ is not concave, we construct a concave function $\mathcal{V}_{\tilde{w}}$ by attaching a linear piece on $\big[0, \breve{w}(\tilde{w})\big)$ to the concave part of function $V_{\tilde{w}}(w)$ for $w \geqslant \breve{w}(\tilde{w})$. This function is closely related to the optimal value function when the state is $\mathsf{I}$, as we show in the next two sections.

In the next section, we first study optimal contract structures under Condition 1, and leave the situation when Condition 2 holds to the following section.

## 4. Optimal Contract under Condition 1

It is intuitive that the optimal contract structure depends on the switching cost $K$. For example, if $K$ is very high, the principal may not want to ever switch the agent from resting to working. On the contrary, if $K$ is very small, the principal may not mind frequently switching the state between working and resting. Therefore, in this section, we start with high $K$ values in Subsection 4.1, and continue with medium and low $K$ values in Subsections 4.2 and 4.3, respectively.

In this section, the optimal value function for state $\mathsf{I}$ relies on the following result.

Lemma 3. *Under Condition 1, there exists a unique $\hat{\mathrm{w}}$ in $[0, \bar{w})$ such that $\mathcal{V}_{\hat{\mathrm{w}}}(0) = \underline{v}$, in which the concave function $\mathcal{V}_{\hat{\mathrm{w}}}$ is defined in Lemma 2 with $\hat{\mathrm{w}}$ replacing $\tilde{w}$. Furthermore, if $\breve{w}(\hat{\mathrm{w}}) > 0$, then $\mathcal{V}'_{\hat{\mathrm{w}}}(0) = \mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\hat{\mathrm{w}})) > 1$.*

Lemma 3 allows us to uniquely identify an upper bound $\hat{\mathrm{w}}$ and a function $\mathcal{V}_{\hat{\mathrm{w}}}$, which is independent of the switching cost $K$, along with a lower bound $\breve{w}(\hat{\mathrm{w}})$, also defined in Lemma 2. If $\breve{w}(\hat{\mathrm{w}}) > 0$, the value function $\mathcal{V}_{\hat{\mathrm{w}}}$ is linear over $w \in [0, \breve{w}(\hat{\mathrm{w}}))$, which is associated with randomized control. The slope of function $\mathcal{V}_{\hat{\mathrm{w}}}$ on this linear piece is larger than 1, which implies that if the societal value function is $\mathcal{V}_{\hat{\mathrm{w}}}$, the principal's utility function $\mathcal{V}_{\hat{\mathrm{w}}}(w) - w$ has a positive maximizer. Overall, the function $\mathcal{V}_{\hat{\mathrm{w}}}$ is concave.

### 4.1. High Switching Cost $K$

We consider high switching cost to be the case such that whenever the state is $\emptyset$, the principal would rather terminate the contract than switching the state to $\mathsf{I}$. In other words, whenever in state $\mathsf{I}$, the principal directs the agent to work until the agent's promised utility reaches 0, at which point the contract is terminated.

The condition for high switching cost $K$ is,

$$K \geqslant \bar{K}_1, \text{ in which } \bar{K}_1 := \max_{w\in[0,\bar{w}]} \mathcal{V}_{\hat{\mathrm{w}}}(w) - \underline{v} = \bar{V}(\hat{\mathrm{w}}) - \underline{v}, \tag{H1}$$

in which the upper bound $\hat{\mathrm{w}}$ is defined according to Lemma 3 and function $\bar{V}(w)$ is defined in (10). The next result implies the optimality of contract $\Gamma^*(w,\mathsf{I};0,\breve{w}(\hat{\mathrm{w}}),\hat{\mathrm{w}},\hat{\mathrm{w}})$ in state $\mathsf{I}$, in which the lower bound $\breve{w}(\hat{\mathrm{w}})$ is also defined in Lemma 3. Following Definition 1, this contract means that in state $\mathsf{I}$, the agent is asked to rest only when the promised utility decreases to 0, which implies contract termination.

Proposition 2. *Under Conditions 1 and (H1), we have*

$$U\Big(\Gamma^*(w,\mathsf{I};0,\breve{w}(\hat{\mathrm{w}}),\hat{\mathrm{w}},\hat{\mathrm{w}}),\mathsf{I}\Big) = \mathcal{V}_{\hat{\mathrm{w}}}(w) - w, \ \forall w \in [0,\hat{\mathrm{w}}], \ \textit{and} \tag{26}$$

$$U\Big(\underline{\Gamma},\emptyset\Big) = \underline{v}. \tag{27}$$

*in which contract $\underline{\Gamma}$ is defined in (16).*

*Furthermore, functions $V_{\mathsf{I}}(w) = \mathcal{V}_{\hat{\mathrm{w}}}(w)$ and $V_{\emptyset}(w) = \underline{v}$ satisfy the optimality condition (20)–(22).*

Following (26), for any contract $\Gamma \in \mathfrak{C}$ that yields a promised utility $w$ for the initial state $\mathsf{I}$, we have

$$U\Big(\Gamma^*(w,\mathsf{I};0,\breve{w}(\hat{\mathrm{w}}),\hat{\mathrm{w}},\hat{\mathrm{w}}),\mathsf{I}\Big) = \mathcal{V}_{\hat{\mathrm{w}}}(w) - w \geqslant U(\Gamma,\mathsf{I})$$

where the inequality follows Theorem 1 and Lemma 1. That is, the principal's utility under contract $\Gamma^*(w,\mathsf{I};0,\breve{w}(\hat{\mathrm{w}}),\hat{\mathrm{w}},\hat{\mathrm{w}})$ is the highest possible that delivers utility $w$ to the agent starting from state $\mathsf{I}$. Similarly, (27) implies that the optimality of contract $\underline{\Gamma}$ defined in (16) is optimal for state $\emptyset$, because

$$U(\Gamma,\emptyset) \leqslant \underline{v} - w \leqslant \underline{v} = U\Big(\underline{\Gamma},\emptyset\Big).$$

Therefore, we have the following result on the optimal contract.

Theorem 2. *Under Conditions 1 and (H1), when the initial state is $\mathcal{E}_{0-} = \mathsf{I}$, it is optimal to implement contract $\Gamma^*(w_0^*,\mathsf{I};0,\breve{w}(\hat{\mathrm{w}}),\hat{\mathrm{w}},\hat{\mathrm{w}})$, in which $w_0^* \in [0,\hat{\mathrm{w}}]$ is a maximizer of function $\mathcal{V}_{\hat{\mathrm{w}}}(w) - w$; when the initial state is $\mathcal{E}_{0-} = \emptyset$, contract $\underline{\Gamma}$ is optimal.*

Remark 2. According to Theorem 2, the principal should not hire the agent to start working ($\mathcal{E}_{0-} = \emptyset$). Furthermore, if the principal is forced to deliver a utility $w > 0$ to the agent, then it is optimal to immediately pay off $\mathrm{d}L_0 = w$ without hiring the agent. It is clear that the corresponding principal's utility under such a contract is $\underline{v} - w$. If we consider the setting in which the agent has been working in the very beginning ($\mathcal{E}_{0-} = \mathsf{I}$), the principal should keep the agent working until contract termination, because whenever the agent rests, it is not worth switching on the work due to the high cost. □

### 4.2. Medium Switching Cost $K$

Even if the switching cost $K$ is lower than $\bar{K}_1$, it may still be high enough for the agent to never start working from resting. Following (H1), when $K < \bar{K}_1$ there must exist value $w$ such that $\mathcal{V}_{\hat{w}}(w) > \underline{v} + K$. Given that function $\mathcal{V}_{\hat{w}}$ is concave and continuously differentiable, this further implies the following result.

Lemma 4. *Under Condition 1 and $K < \bar{K}_1$, there exists $K$-dependent values $\bar{\theta}^K \in [\breve{w}(\hat{w}), \hat{w}]$ and $m^K \in [0, \mathcal{V}'_{\hat{w}}(0)]$ such that*

$$\mathcal{V}_{\hat{w}}(\bar{\theta}^K) = m^K \bar{\theta}^K + K + \underline{v}, \text{ and } \mathcal{V}'_{\hat{w}}(\bar{\theta}^K) = m^K. \tag{28}$$

*Furthermore, we have $\bar{\theta}^K$ is increasing in $K$, $m^K$ is decreasing in $K$, and $\lim_{K \downarrow 0} \bar{\theta}^K = \breve{w}(\hat{w})$.*

With the help of $m^K$, we can define the interval for "medium level" switching cost $K$ as

$$\underline{K}_1 \leqslant K < \bar{K}_1, \text{ in which } \underline{K}_1 := \inf\left\{K \in \left(0, \bar{K}_1\right] \mid m^K < 1\right\}. \tag{M1}$$

Geometrically, Lemma 4 implies that the line $m^K w + \underline{v}$ is tangent to the curve $\mathcal{V}_{\hat{w}}(w) - K$ at $w = \bar{\theta}^K$. Therefore, we define the following societal value function for the resting state,

$$V_{\emptyset}(w) = \begin{cases} m^K w + \underline{v}, & w \in \left[0, \bar{\theta}^K\right], \\ \mathcal{V}_{\hat{w}}(w) - K, & w \in \left[\bar{\theta}^K, \hat{w}\right], \end{cases} \tag{29}$$

which is clearly concave, and linear on $\left[0, \bar{\theta}^K\right]$ with slope $m^K$. Function $V_{\emptyset}(w)$ is shown as the dashed curve in Figure 3. Because the slope $m^K$ is less than or equal to 1 under condition (M1), the corresponding principal's utility function, $V_{\emptyset}(w) - w$, is monotonically non-increasing. Note that if $\underline{K}_1 > 0$, then $m^{\underline{K}_1} = 1$.

Similar to Proposition 2, we have the following result.

Proposition 3. *Under Conditions 1 and (M1), Equations (26) and (27) still hold. Furthermore, functions $V_{\mathsf{I}}(w) = \mathcal{V}_{\hat{w}}(w)$ and $V_{\emptyset}(w)$ as defined in (29) satisfy the optimality condition (20)–(22).*

Therefore, the only difference between Propositions 2 and 3 is the value function for state $\emptyset$. The optimal contract structure is the same under (H1) and (M1). Figure 3 shows an example of the societal value functions. It is clear that $V_{\emptyset}(w)$ is linear over the interval $\left[0, \bar{\theta}^K\right]$. Furthermore, functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ are "parallel" with a difference of $K$ for $w \geqslant \bar{\theta}^K$. At time $t$, if the promised utility $W_t > \bar{\theta}^K$, and state $\mathcal{E}_t = \emptyset$, it is optimal to switch the agent to work, which explains the difference $K$ between the value functions. Finally, as mentioned earlier, under condition (M1), the slope of the function $V_{\emptyset}$ is always less than 1, which implies that the slope of the principal's

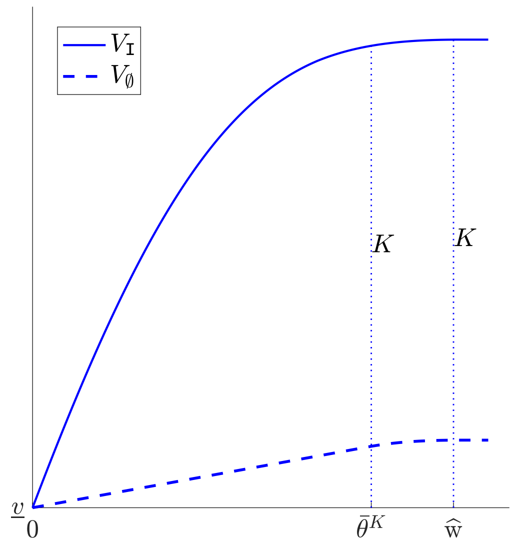



**Figure 3** **Value functions with** $r = 0.2$**,** $\rho = 0.5$**,** $c = b = 0.2$**,** $R = 2$**,** $\Delta\mu = 0.7$**,** $K = 4$**, and** $\mu = 2$**. In this case,** $\bar{w} = 1.14$**,** $\mathcal{V}_{\widehat{\mathrm{w}}}(\widehat{\mathrm{w}}) = 17.67$ **and** $\underline{v} = 13$**.**

utility function $V_{\emptyset}(w) - w$ is always negative. Therefore, the optimal contract for the initial state $\emptyset$ is to set the promised utility at 0, or, not to hire the agent.

For completeness, we present the following theorem, which states that the same contracts that are optimal for high switching cost are also optimal under medium switching cost.

THEOREM 3. *Under Conditions 1 and (M1), contracts* $\Gamma^*(w_0^*, \mathsf{I}; 0, \breve{w}(\widehat{\mathrm{w}}), \widehat{\mathrm{w}}, \widehat{\mathrm{w}})$ *and* $\underline{\Gamma}$ *are optimal if the initial state* $\mathcal{E}_{0-}$ *is* $\mathsf{I}$ *and* $\emptyset$*, respectively.*

REMARK 3. Here, it is interesting to compare with Remark 2 and consider the situation if the principal is forced to deliver a utility $w > 0$ to the agent starting at state $\emptyset$. If $w > \bar{\theta}^K$, then it is optimal to direct the agent to start working immediately, and then follow the contract $\Gamma^*(w, \mathsf{I}; 0, \breve{w}(\widehat{\mathrm{w}}), \widehat{\mathrm{w}}, \widehat{\mathrm{w}})$. If $w \in (0, \bar{\theta}^K)$, on the other hand, it is optimal to randomize the agent's continuation utility at time 0. Specifically, the agent is immediately terminated with probability $1 - w/\bar{\theta}^K$, and, with probability $w/\bar{\theta}^K$, the agent should start working following contract $\Gamma^*(\bar{\theta}^K, \mathsf{I}; 0, \breve{w}(\widehat{\mathrm{w}}), \widehat{\mathrm{w}}, \widehat{\mathrm{w}})$, from a starting promised utility $\bar{\theta}^K$. It is easy to verify that the corresponding principal's utility at time zero before the randomization is indeed $V_{\emptyset}(w) - w$, following (29). Therefore, although the optimal contracts are the same under high and medium switching cost cases when the principal can freely choose the initial promised utility, there are subtle differences between them. □

### 4.3. Low Switching Cost $K$

The previous two subsections show that when the switching cost $K \geqslant \underline{K}_1$, it is optimal not to switch the state from $\emptyset$ to $\mathsf{I}$. Equivalently, the lower threshold $\underline{\theta} = 0$. More interesting and richer structure occurs when the switching cost $K$ satisfies the following condition,

$$0 < K < \underline{K}_1. \tag{L1}$$

Small $K$ means that switching the agent from resting to working is not too costly. It is clear that condition (L1) holds only if $\underline{K}_1$ is strictly positive, or, equivalently, slope $m^K$ defined in Lemma 4 is larger than 1 for small enough $K$.

In this case, the value function for state $\mathsf{I}$ is no longer $\mathcal{V}_{\hat{\mathrm{w}}}(w)$. Recall the function $\mathcal{V}_{\widetilde{w}}$ defined in Lemma 2 for any $\widetilde{w} \in (0, \bar{w})$. When the switching cost $K$ is low, the additional boundary condition that allows us to identify a particular $\widetilde{w}$ to obtain a value function is no longer at $w = 0$. Instead, we need to identify the threshold $\underline{\theta}$, at which point the value functions for states $\emptyset$ and $\mathsf{I}$ are connected. In particular, when the promised utility is below $\underline{\theta}$, the principal should direct the agent to rest, which allows the promised utility to increase, according to (11). For the resting state $\emptyset$, we use function $\mathsf{V}_c(w)$ defined in (23) as the value function. The next result allows us to identify all the parameters, including the constant $c$ in (23).

Proposition 4. *Under Conditions 1 and (L1), there exists a set of parameters* $(\mathsf{c}, \hat{\mathbf{w}}, \bar{\vartheta}, \underline{\vartheta})$ *with* $\mathsf{c} > 0$ *and* $\underline{\vartheta} < \bar{\vartheta} < \hat{\mathbf{w}} < \widehat{\mathrm{w}}$*, in which* $\widehat{\mathrm{w}}$ *is defined in Lemma 3, such that*

$$\mathcal{V}_{\hat{\mathbf{w}}}(\underline{\vartheta}) = \mathsf{V}_{\mathsf{c}}(\underline{\vartheta}), \tag{30}$$

$$\mathcal{V}'_{\hat{\mathbf{w}}}(\underline{\vartheta}) = \mathsf{V}'_{\mathsf{c}}(\underline{\vartheta}), \tag{31}$$

$$\mathcal{V}_{\hat{\mathbf{w}}}(\bar{\vartheta}) = \mathsf{V}_{\mathsf{c}}(\bar{\vartheta}) + K, \text{ and} \tag{32}$$

$$\mathcal{V}'_{\hat{\mathbf{w}}}(\bar{\vartheta}) = \mathsf{V}'_{\mathsf{c}}(\bar{\vartheta}) > 1, \tag{33}$$

*in which* $\mathcal{V}_{\hat{\mathbf{w}}}$ *is defined in Lemma 2 with* $\hat{\mathbf{w}}$ *replacing* $\widetilde{w}$*, and* $\mathsf{V}_{\mathsf{c}}$ *is defined in (23) with* $\mathsf{c}$ *replacing* $c$*. Moreover, we have* $\bar{\vartheta} > \breve{w}(\hat{\mathbf{w}})$*, in which* $\breve{w}(\hat{\mathbf{w}})$ *is defined in Lemma 2(i), with* $\hat{\mathbf{w}}$ *replacing* $\widetilde{w}$*.*

Equations (30) and (31) are called *value-matching* and *smooth-pasting* conditions in the optimal control literature, respectively, which occur at the promised utility threshold $\underline{\vartheta}$, when the state is switched from $\mathsf{I}$ to $\emptyset$. Similarly, (32) and (33) specify the value-matching and smooth-pasting conditions when the state is switched from $\emptyset$ to $\mathsf{I}$ at promised utility $\bar{\vartheta}$, except that the values between the two states differ by $K$. The proof of Proposition 4 is rather intricate, and takes four key steps, as shown in Appendix B.3.3. Relying on these key steps, we illustrate how to identify compute $\bar{\vartheta}$ and $\underline{\vartheta}$ in Appendix A.2.

Equipped with Proposition 4, we define the following optimal value functions:

$$V_{\mathsf{I}}(w) := \begin{cases} \mathsf{V}_{\mathsf{c}}(w), & w \in [0, \underline{\vartheta}), \\ \mathcal{V}_{\hat{\mathbf{w}}}(w), & w \geqslant \underline{\vartheta}, \end{cases} \quad \text{and} \quad V_{\emptyset}(w) := \begin{cases} \mathsf{V}_{\mathsf{c}}(w), & w \in [0, \bar{\vartheta}), \\ \mathcal{V}_{\hat{\mathbf{w}}}(w) - K, & w \geqslant \bar{\vartheta}. \end{cases} \tag{34}$$

Figure 4 depicts functions $V_{\mathsf{I}}$ and $V_{\emptyset}$ defined in (34). In particular, functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ are identical for $w \leqslant \underline{\vartheta}$. Furthermore, function $V_{\mathsf{I}}(w)$ is linear in the interval $[\underline{\vartheta}, \breve{w}(\hat{\mathbf{w}})]$, while $V_{\emptyset}(w)$ remains to be $\mathsf{V}_{\mathsf{c}}(w)$ for $w \leqslant \bar{\vartheta}$. For higher $w$ such that $w \geqslant \bar{\vartheta}$, however, function $V_{\emptyset}(w)$ is a parallel shift of $V_{\mathsf{I}}(w)$, where the two functions differ by $K$.

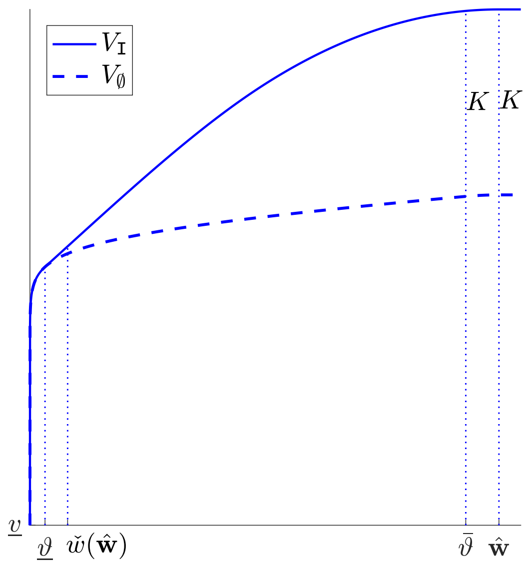



**Figure 4** **Value functions with** $r = 0.05$, $\rho = 1$, $c = b = 0.3$, $R = 112$, $\Delta\mu = 0.1$, $K = 40$, **and** $\mu = 1.95$. **In this case,** $\bar{w} = 5.85$, $\underline{\vartheta} = 0.18$, $\breve{w}(\hat{\mathbf{w}}) = 0.45$, $\bar{\vartheta} = 5.22$, **and** $\hat{\mathbf{w}} = 5.62$.

Now, we specify the optimal contract structure as follows. For any $w_0 \geqslant 0$, define contracts

$$\hat{\Gamma}_{\mathsf{I}}(w_0) := \begin{cases} \Gamma^* \left(w_0, \mathsf{I};\ \underline{\vartheta},\ (\underline{\vartheta} \vee \breve{w}(\hat{\mathbf{w}})),\ \bar{\vartheta},\ \hat{\mathbf{w}}\right), & w_0 > \underline{\vartheta}, \\ \Gamma^* \left(w_0, \emptyset;\ \underline{\vartheta},\ (\underline{\vartheta} \vee \breve{w}(\hat{\mathbf{w}})),\ \bar{\vartheta},\ \hat{\mathbf{w}}\right), & w_0 \leqslant \underline{\vartheta}, \end{cases} \quad \text{and} \tag{35}$$

$$\hat{\Gamma}_{\emptyset}(w_0) := \begin{cases} \Gamma^* \left(w_0, \mathsf{I};\ \underline{\vartheta},\ (\underline{\vartheta} \vee \breve{w}(\hat{\mathbf{w}})),\ \bar{\vartheta},\ \hat{\mathbf{w}}\right), & w_0 > \bar{\vartheta}, \\ \Gamma^* \left(w_0, \emptyset;\ \underline{\vartheta},\ (\underline{\vartheta} \vee \breve{w}(\hat{\mathbf{w}})),\ \bar{\vartheta},\ \hat{\mathbf{w}}\right), & w_0 \leqslant \bar{\vartheta}, \end{cases} \tag{36}$$

where we use notation $a \vee b$ to denote $\max\{a, b\}$ for any $a, b \in \mathbb{R}$. The term $\underline{\vartheta} \vee \breve{w}(\hat{\mathbf{w}})$ as the threshold $\breve{w}$ of Definition 1 implies that random switching from $\mathsf{I}$ to $\emptyset$ occurs under this contract if and only if $\underline{\vartheta} < \breve{w}(\hat{\mathbf{w}})$. If $\underline{\vartheta} \geqslant \breve{w}(\hat{\mathbf{w}})$, on the other hand, contracts $\hat{\Gamma}_{\mathsf{I}}(w_0)$ and $\hat{\Gamma}_{\emptyset}(w_0)$ demonstrate the control band structure. Note that the initial promised utility value $w_0$ affects which state to start the contracts with. For contract $\hat{\Gamma}_{\mathsf{I}}(w_0)$, the threshold is $\underline{\vartheta}$. In contrast, for contract $\hat{\Gamma}_{\emptyset}(w_0)$, the threshold is $\bar{\vartheta}$. The following result implies that these contracts are indeed related to the optimal ones.

PROPOSITION 5. *Under Conditions 1 and (L1), functions $V_{\mathsf{I}}$ and $V_{\emptyset}$ defined in (34) satisfy the optimality condition (20)–(22). Furthermore, for any $w \geqslant 0$, we have*

$$U\left(\hat{\Gamma}_{\mathsf{I}}(w), \mathsf{I}\right) = V_{\mathsf{I}}(w) - w, \text{ and} \tag{37}$$

$$U\left(\hat{\Gamma}_{\emptyset}(w), \emptyset\right) = V_{\emptyset}(w) - w. \tag{38}$$

Note that it is quite involved to verify that $V_{\mathsf{I}}(w)$ satisfies $\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}} \geqslant 0$ in condition (20) for $w \in [0, \underline{\vartheta}]$. As one can imagine, we need to show that the function $\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}}$ is always monotone in this interval. However, this function is not convex. In the proof presented in Appendix B.3.5, we have to establish that either $\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}}$'s first-order-derivative is negative, or its second-order-derivative is positive, throughout this interval. Together with the fact that the function $\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}}$ takes a non-negative value and negative derivative at $\underline{\vartheta}$, this guarantees $\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}} \geqslant 0$. Corresponding proofs in the existing literature, such as Duckworth and Zervos (2001), Vath and Pham (2007) are much simpler in comparison. In particular, Vath and Pham (2007) relies on showing convexity/concavity to verify variational inequalities.

Note that following expression (33) of Proposition 4, the slope of $V_{\emptyset}(w)$ at $w = \bar{\vartheta}$ is larger than 1. (Figure 4 does not appear this way, because of different scales of the $x$ and $y$-axes. —The value of $\hat{\mathbf{w}}$ is around 5, while the difference $V_{\mathsf{I}}(\widehat{w}) - V_{\mathsf{I}}(0)$ is around a few thousands.) This implies that the principal's utility function $V_{\emptyset}(w) - w$ for state $\emptyset$ is maximized at a point larger than $\bar{\vartheta}$. The same point maximizes function $V_{\mathsf{I}}(w) - w$ as well, because $V_{\emptyset}(w) - w$ and $V_{\mathsf{I}}(w) - w$ are "parallel" and differ by $K$ for $w \geqslant \bar{\vartheta}$ according to (34). Therefore, we have the following result.

THEOREM 4. *Under Conditions 1 and (L1), for any initial state $\varepsilon_0 \in \{\mathsf{I}, \emptyset\}$, contract $\hat{\Gamma}_{\varepsilon_0}(\mathsf{w}_0^*)$ is optimal, in which $\mathsf{w}_0^* \in [\bar{\vartheta}, \hat{\mathbf{w}}]$ is a maximizer of both functions $V_{\emptyset}(w) - w$ and $V_{\mathsf{I}}(w) - w$. In particular, if the initial state is $\emptyset$, the principal should ask the agent to start working right away.*

Theorem 4 demonstrates that when the promised utility is low due to underperformance, the principal should punish the agent with resting for a period of time, rather than terminating the contract forever. Following the optimal contract structure, the agent is never terminated. As discussed in the introduction, this is because contract termination, as a threaten to mitigate moral hazard, is in fact even more inefficient than letting the agent to rest and then paying the fixed switching cost to restart working once in a while.

Now let us summarize the optimal contract structures obtained in this section. Under Condition 1, the principal should hire the agent to start working immediately only if the switching cost $K$ is lower than $\underline{K}_1$. The corresponding optimal dynamic contract starts at the promised utility $\mathsf{w}_0^*$ and follows the general dynamic outlined in Definition 1 with parameters $\underline{\theta} = \underline{\vartheta}$, $\breve{w} = (\underline{\vartheta} \vee \breve{w}(\hat{\mathbf{w}}))$, $\bar{\theta} = \bar{\vartheta}$, and $\widehat{w} = \hat{\mathbf{w}}$. In Appendix A.2 we discuss how to compute contract parameters $\underline{\vartheta}$, $\breve{w}(\hat{\mathbf{w}})$, $\bar{\vartheta}$ and $\hat{\mathbf{w}}$.

## 5. Optimal Contract under Condition 2

Under Condition 2, the optimal value functions may not be differentiable on the entire $\mathbb{R}_+$ anymore. Similar to the previous section, we explore the optimal contract structures for different values of the switching cost.

### 5.1. High and Medium Switching Cost $K$

Under Condition 2, function $\mathcal{V}_{\hat{w}}$ from Lemma 3 no longer exists, because the boundary condition $\mathcal{V}_{\tilde{w}}(0) = \underline{v}$ does not hold for any $\tilde{w} \in [0, \bar{w}]$. In this case, the principal needs to either set the agent's promised utility at $\bar{w}$, or 0, but never in between. The corresponding value function is linear over the interval $[0, \bar{w}]$, connecting $\underline{v}$ at $w = 0$ and $\bar{V}$ at $w = \bar{w}$, where $\bar{V}$ is defined in (9). Therefore, define piece-wise linear function

$$V_{\mathsf{I}}(w) := \begin{cases} \underline{v} + \dfrac{\bar{V} - \underline{v}}{\bar{w}} \cdot w, & w \in [0, \bar{w}], \\ \bar{V}, & w \geqslant \bar{w}. \end{cases} \tag{39}$$

Under Condition 2, it is clear that the slope $\dfrac{\bar{V} - \underline{v}}{\bar{w}} > 1$. If

$$K \geqslant \bar{K}_2 := \bar{V} - \underline{v}, \tag{H2}$$

we claim, and will later show, that the principal's value function for state $\mathsf{I}$ is $V_{\mathsf{I}}(w) - w$, and for state $\emptyset$ is $\underline{v} - w$.

Now consider the case that $K$ is neither too high or too low. That is,

$$\underline{K}_2 \leqslant K < \bar{K}_2, \text{ in which } \underline{K}_2 := \bar{V} - \underline{v} - \bar{w}. \tag{M2}$$

In this case the value function for state $\emptyset$ changes from linear under (H2) to the following piece-wise linear function,

$$V_{\emptyset}(w) = \begin{cases} \underline{v} + \dfrac{\bar{V} - \underline{v} - K}{\bar{w}} \cdot w, & w \in [0, \bar{w}], \\ \bar{V} - K, & w \geqslant \bar{w}. \end{cases} \tag{40}$$

This is in contrast to the differentiable value function of Section 4.2. Note that under (M2), the slope $\dfrac{\bar{V} - \underline{v} - K}{\bar{w}} < 1$, which implies that the principal's utility function, $V_{\emptyset}(w) - w$ is monotonically decreasing.

THEOREM 5. *If Condition 2 holds, under (H2), functions $V_{\mathsf{I}}(w)$ as defined in (39) and $V_{\emptyset}(w) = \underline{v}$ satisfy the optimality condition (20)–(22); under (M2), on the other hand, functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ as defined in (39) and (40), respectively, satisfy (20)–(22).*

*Furthermore, if $K > \bar{V} - \underline{v} - \bar{w}$, for all $w \geqslant 0$, we have*

$$U\left(\bar{\Gamma}, \mathsf{I}\right) = \bar{V} - \bar{w} \geqslant V_{\mathsf{I}}(w) - w, \text{ and} \tag{41}$$

$$U\left(\underline{\Gamma}, \emptyset\right) = \underline{v} \geqslant \begin{cases} \underline{v} - w, & \text{under (H2)}, \\ V_{\emptyset}(w) - w, & \text{under (M2)}, \end{cases} \tag{42}$$

*in which* $\bar{\Gamma}$, $\underline{\Gamma}$, *and* $V_\emptyset(w)$ *are defined in* (15), (16), *and* (40), *respectively.*

*Therefore, under Condition* 2 *and* $K > \underline{K}_2$, *it is optimal to implement contract* $\bar{\Gamma}$ *for the initial state* $\mathcal{E}_{0-} = \mathsf{I}$, *and* $\underline{\Gamma}$ *for* $\mathcal{E}_{0-} = \emptyset$.

Remark 4. As a reminder, contract $\bar{\Gamma}$ directs the agent to work forever and pays $\beta$ for each arrival. Its optimality for state $\mathsf{I}$ is in sharp contrast with the correponding optimal contract in Section 4 when $K > \underline{K}_1$, where the agent is terminated within finite time with probability 1. This is intuitive, because under Condition 2, the revenue $R$ per arrival is higher, compared with Condition 1, and therefore the principal is willing to always keep the agent working. Next, we show that if $K$ is lower, then contract $\bar{\Gamma}$ may be optimal even for state $\emptyset$. □

### 5.2. Low Switching Cost $K$

We now consider the arguably more interesting case of $K$ being low, or,

$$0 < K < \underline{K}_2. \tag{L2}$$

The following result, which is similar to Lemma 4, uses the resting state $\emptyset$'s value function $\mathsf{V}_c$ to identify parameters for the value functions later.

Lemma 5. *Under Conditions* 2 *and* (L2), *there exists a set of* $K$*-dependent parameters* $(c_K, m_K, \underline{\theta}_K)$ *with* $c_K > 0$ *and* $m_K \in \big(0, (\bar{V} - \underline{v})/\bar{w}\big)$, *such that*

$$\mathsf{V}_{c_K}(\underline{\theta}_K) = \bar{V} + m_K(w - \bar{w}), \tag{43}$$

$$\mathsf{V}'_{c_K}(\underline{\theta}_K) = m_K, \text{ and} \tag{44}$$

$$\mathsf{V}_{c_K}(\bar{w}) = \bar{V} - K, \tag{45}$$

*in which* $\mathsf{V}_{c_K}$ *is defined in* (23) *with* $c_K$ *replacing* $c$. *Furthermore, we have* $c_K$ *is decreasing in* $K$, $m_K$ *is increasing in* $K$, *and* $\underline{\theta}_K$ *is decreasing in* $K$ *with* $\lim_{K\downarrow 0} \underline{\theta}_K = \bar{w}$.

It is easy to verify that Condition 2 is equivalent to $\dfrac{\bar{V} - \underline{v}}{\bar{w}} \geqslant \dfrac{\rho - r}{\rho - r - \mu} > 1$. Given $m_K \leqslant \dfrac{\bar{V} - \underline{v}}{\bar{w}}$ following Lemma 5, we first consider parameter settings that satisfy the following condition, which is more restrictive than Condition 2,

$$m_K \geqslant \frac{\rho - r}{\rho - r - \mu} > 1. \tag{mH}$$

Following the monotonicity of $m_K$ in $K$, (mH) implies that

$$\check{K}_2 \leqslant K < \underline{K}_2, \text{ in which } \check{K}_2 := \inf\left\{ K \in (0, \underline{K}_2] \;\middle|\; m_K \geqslant \frac{\rho - r}{\rho - r - \mu} \right\}. \tag{46}$$

Under Condition (mH), we define the following function for state I, which is smooth-pasting between the function $\mathsf{V}_{c_K}$ and a linear piece for the interval $[\underline{\theta}_K, \bar{w}]$,

$$V_{\mathsf{I}}(w) := \begin{cases} \mathsf{V}_{c_K}(w), & w \in [0, \underline{\theta}_K], \\ \bar{V} + m_K(w - \bar{w}), & w \in [\underline{\theta}_K, \bar{w}], \\ \bar{V}, & w > \bar{w}. \end{cases} \tag{47}$$

We further extend function $\mathsf{V}_{c_K}$ to include $w > \bar{w}$,

$$V_{\emptyset}(w) := \begin{cases} \mathsf{V}_{c_K}(w), & w \in [0, \bar{w}], \\ \bar{V} - K, & w > \bar{w}. \end{cases} \tag{48}$$

Figure 5 gives an example of the societal value functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$, as defined in (47) and (48), respectively. As we can see, the two functions are the same for $w < \underline{\theta}_K$, and have the same derivative at $w = \underline{\theta}_K$. The two functions then diverge for $w > \underline{\theta}_K$, with function $V_{\mathsf{I}}(w)$ being piece-wise linear in this interval. For $w \geqslant \bar{w}$, both functions become constant, and differ by exactly $K$.

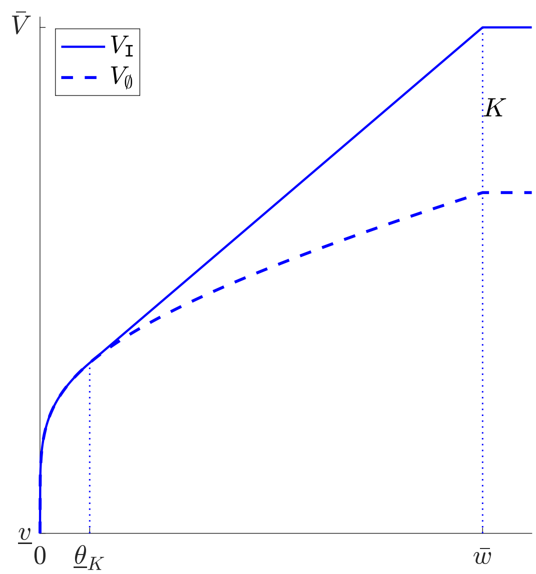



**Figure 5** **Value functions with** $r = 0.2$**,** $\rho = 0.5$**,** $c = b = 0.3$**,** $R = 10$**,** $\Delta\mu = 0.2$**,** $K = 1.6$**, and** $\mu = 0.6$**. In this case,** $\bar{w} = 0.9$**,** $\underline{\theta}_K = 0.1$**,** $\bar{V} = 24.9$ **and** $\underline{v} = 20$**.**

Following (mH), and the fact that $\mathsf{V}_c'(w) > 1$ for any $c$ and $w$, the derivatives of both functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ are higher than 1 for any $w \in [0, \bar{w})$. Therefore, both functions $V_{\mathsf{I}}(w) - w$ and $V_{\emptyset}(w) - w$ are maximized at $\bar{w}$. Hence, in the following, we show that contract $\bar{\Gamma}$ is optimal.

Theorem 6. *Under Conditions (2), (L2) and (mH), functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$, as defined in (47) and (48), respectively, satisfy the optimality condition (20)–(22).*

*Furthermore, for any $w \geqslant 0$, we have*

$$U\left(\bar{\Gamma}, \mathsf{I}\right) = \bar{V} - \bar{w} \geqslant V_{\mathsf{I}}(w) - w, \text{ and} \tag{49}$$

$$U\left(\bar{\Gamma}, \emptyset\right) = \bar{V} - \bar{w} - K \geqslant V_{\emptyset}(w) - w, \tag{50}$$

*in which $\bar{\Gamma}$ is defined in (15). Therefore, it is optimal to implement contract $\bar{\Gamma}$ regardless of whether the initial state is $\mathcal{E}_{0-} = \mathsf{I}$ or $\mathcal{E}_{0-} = \emptyset$.*

Comparing Theorem 6 with Theorem 5, we see that if $K \geqslant \underline{K}_2$, then it is optimal for the principal to not hire the agent to start working; if $K < \underline{K}_2$ and (mH) holds, on the other hand, it is optimal to start the work from the very beginning and keep the agent working forever.

If condition (mH) does not hold, or, equivalently,

$$0 < m_K < \frac{\rho - r}{\rho - r - \mu}, \tag{mL}$$

the function $\mathcal{V}_{\tilde{w}}$ defined in Lemma 2 is still the optimal value function for state $\mathsf{I}$, and the results are identical to those in Proposition of Section 4.3, as summarized in the next theorem.

THEOREM 7. *Under Conditions 2, (L2), and (mL), Propositions 4 and 5 and Theorem 4 still hold.*

The following result further implies that as if $R$ is large enough, we have the fixed switching cost $K \in (\breve{K}_2, \underline{K}_2)$, when contract $\bar{\Gamma}$ becomes optimal.

PROPOSITION 6. *Fixing model parameters $\rho$, $r$, $\mu$, $\Delta\mu$, $c$ and $b$, the threshold $\breve{K}_2$ is non-increasing in $R$ for $R \geqslant \hat{R}$, and reaches $0$ at a point $\bar{R} > \hat{R}$. Furthermore, $\underline{K}_2$ is increasing in $R$ and diverge to infinity with $R$.*

The proof of Proposition 6 in the appendix provides closed form expressions for $\breve{K}_2$ and $\bar{R}$.

REMARK 5. At this point, it is helpful to use Figure 6 to summarize the optimal contract structures under different parameter settings from Sections 4 and 5. As we can see, if the switching cost is above $\underline{K}_1$ or $\underline{K}_2$ (Region I), it is optimal for the principal not to hire the agent at state $\emptyset$. If $R$ is high enough such that $\breve{K}_2 \leqslant K \leqslant \underline{K}_2$ (Region II), then it is optimal for the principal to hire the agent and start paying $\beta$ for each arrival. If $K$ is lower than $\underline{K}_1$ or $\breve{K}_2$ (Region III), the optimal contract take the general form of $\Gamma^*\left(\mathrm{w}_0^*, \mathsf{I};\ \underline{\vartheta},\ (\underline{\vartheta} \wedge \breve{w}(\hat{\mathbf{w}})),\ \bar{\vartheta},\ \hat{\mathbf{w}}\right)$. This figure depicts model parameters such that $\rho - r > \mu$. If $\rho - r \leqslant \mu$, on the other hand, $\hat{R}$ is essentially infinity and Region II disappears from the graph. ◻

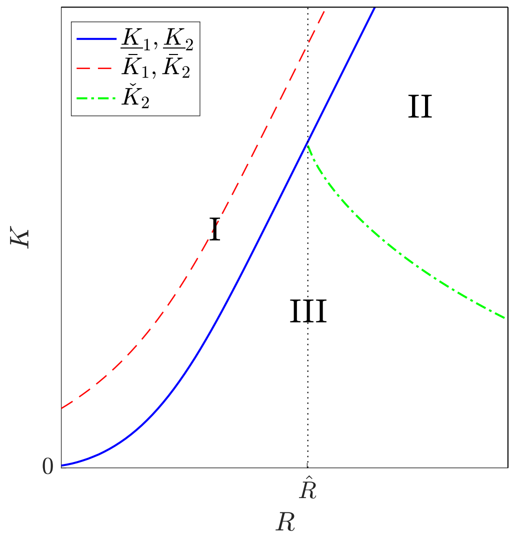



**Figure 6** $r = 0.2$, $\rho = 1.5$, $c = b = 0.2$, $R \in [0.6, 1.1]$, $\Delta\mu = 0.7$, **and** $\mu = 1$.

## 6. Switching Cost $K$ Approaching Zero

In this section, we discuss impacts of the switching cost $K$ on the optimal contract, especially when $K$ approaches zero. For this purpose, we focus on the following condition.

Condition 3. Model parameters (not including $K$) satisfy either Condition 1 and $\underline{K}_1 > 0$, or Condition 2 and $\breve{K}_2 > 0$.

Note that if Condition 3 does not hold, when $K = 0$, the principal should either not hire the agent, or always motivate the agent to work.

Proposition 7. *Under Condition 1 and* (L1)*, or under Condition 2,* (L2) *and* (mL)*, $\underline{\vartheta}$ and $\bar{\vartheta}$ defined in Proposition 4 are decreasing and increasing in $K$, respectively. Furthermore, under Condition 3, these two values converge to the same value as $K$ approaches* 0*, or,*

$$\theta_0 := \lim_{K \downarrow 0} \underline{\vartheta} = \lim_{K \downarrow 0} \bar{\vartheta}. \tag{51}$$

The monotonicity of $\underline{\vartheta}$ and $\bar{\vartheta}$ implies that the limit $\theta_0$ is an upper or lower bound for these thresholds. In Appendix A.2, we demonstrate an algorithm to compute the optimal contract for general $K$ values, in which computing $\theta_0$ is the first step.

Proposition 7 also implies that, as $K$ approaches 0, the control band between $\underline{\vartheta}$ and $\bar{\vartheta}$ diminishes. Consequently, switching occurs more and more frequently. In the limit as $K$ becomes zero, the number of switchings approaches infinity in a finite time period after the promised utility reaches the threshold $\theta_0$. A similar, although not identical, phenomenon in the optimal contract structures arises in the Brownian motion uncertainty case, as demonstrated in Zhu (2013), where the promised utility becomes "sticky" when the promised utility reaches a threshold.

Intuitively, a high switching frequency control policy appears impractical. Therefore, it is instructive to reflect on basic modeling choices. If the switching cost is fairly low, it is often a good practice to ignore it when building the first model. However, if the corresponding optimal switching frequency is extremely high, any cost associated with switching cannot be ignored any more.

Although the optimal control is not practical if $K=0$, we can still study the corresponding optimal value function, which sheds lights on how much the resting option helps, compared with always inducing the agent to work until potential termination. Following Proposition 7, we have the following result, which allows us to construct the optimal value function for $K=0$, regardless of whether other model parameters satisfy Conditions 1 or 2.

THEOREM 8. *Under Condition 3, the following quantities are well defined:*

$$\hat{\mathbf{w}}_0 := \lim_{K\downarrow 0} \hat{\mathbf{w}}, \quad \textit{and} \quad \mathsf{c}_0 := \lim_{K\downarrow 0} \mathsf{c}, \tag{52}$$

*in which* $\hat{\mathbf{w}}$ *and* $\mathsf{c}$ *are defined according to Proposition 4. Further define function*

$$\mathfrak{V}_{\theta_0}(w) := \begin{cases} \mathsf{V}_{\mathsf{c}_0}(w), & w \in [0, \theta_0], \\ \mathcal{V}_{\hat{\mathbf{w}}_0}(w), & w > \theta_0. \end{cases}$$

*Functions* $V_{\mathsf{l}} = V_{\emptyset} = \mathfrak{V}_{\theta_0}$ *satisfy* (20)–(22) *in which we set* $K=0$.

Theorem 8 implies that as $K$ approaches zero, the optimal value functions for positive $K$ values converge to a value function $\mathfrak{V}_{\theta_0}$, which is an upper bound of the optimal value function for $K=0$. Therefore, function $\mathfrak{V}_{\theta_0}$ serves as a benchmark for potential benefits of the switching option. Proposition 8 in Appendix A.2 describes how to compute the function $\mathfrak{V}_{\theta_0}$ directly, rather than to treat it as the limit of a sequence of functions.

Following Theorem 8, we define the optimal principal's utility under $K=0$ as

$$\bar{U} := \max_{w \geqslant 0} \left\{ \mathfrak{V}_{\theta_0}(w) - w \right\}. \tag{53}$$

It is worth comparing this value with the principal's utility without the resting option following Cao et al. (2021), defined as

$$\underline{U} := \begin{cases} \max_{w\geqslant 0} \left\{ \mathcal{V}_{\hat{\mathrm{w}}}(w) - w \right\}, & \text{under Condition 1}, \\ \bar{V} - \bar{w}, & \text{under Condition 2}. \end{cases} \tag{54}$$

Therefore, it is clear that if model parameters do not satisfy Condition 3, the switching option does not bring any value to the principal. Under Condition 3, we conduct a numerical test to compute the relative difference, $(\bar{U} - \underline{U})/\underline{U}$.

In particular, we consider the following model parameters. Fix $\rho = 1$, $R = 10$ and $c = b$. Take $r$ from the set $\{0.01, 0.1, 0.5, 0.9, 0.99\}$, $\mu$ from $\{0.1, 0.55, 1.1, 1.45, 1.9\}$, $\Delta\mu/\mu$ from $\{0.1, 0.5, 0.9\}$, and $c/(R\Delta\mu)$ from $\{0.1, 0.5, 0.9\}$, so that model parameters satisfy Assumption 1 and $r < \rho$. Among

**Table 1** **Parameters of the cases with relative difference that is greater than 10%**

| $r$ | $\mu$ | $\Delta\mu/\mu$ | $c/(R\Delta\mu)$ | Relative difference |
|---|---|---|---|---|
| 0.01 | 1.9 | 0.9 | 0.5 | 68.74% |
| 0.01 | 1.45 | 0.9 | 0.5 | 58.05% |
| 0.01 | 1 | 0.9 | 0.5 | 42.14% |
| 0.1 | 1.9 | 0.9 | 0.5 | 30.06% |
| 0.1 | 1.45 | 0.9 | 0.5 | 28.36% |
| 0.1 | 1 | 0.9 | 0.5 | 24.47% |
| 0.01 | 0.55 | 0.9 | 0.5 | 20.40% |
| 0.1 | 0.55 | 0.9 | 0.5 | 14.61% |

these 225 cases, 85 of them satisfy Condition 3. The mean of the relative differences among these 85 cases is 3.71%. However, in 8 cases, the relative difference exceeds 10%. We list the parameter of these 8 cases in Table 1. As we can see, these cases correspond to $r$ being very low (taking values 0.01 and 0.1), $\mu$ not too low (no lower than 0.55), $\Delta\mu$ is close to $\mu$ (ratio being 0.9), and $c/(R\Delta\mu)$ is neither close to 0 nor to 1. The maximum improvement of considering the switching option can be as high as 68.74%.

## 7. Concluding Remarks

We have fully solved the optimal contract design problem that dynamically schedules an agent to work and rest over time, depending on past arrival times. A natural modeling extension is to include a benefit payoff that the agent collects while resting. Although seemingly simple, such an extension appears highly non-trivial, given how intricate the proofs for Propositions 4 and 5 are. We suspect that the current policy structure remains optimal, but leave it as an open question. Furthermore, our model assumes that the principal undertakes the fixed switching cost. There could be settings where this cost is incurred to the agent and not observable to the principal. In such a setting, even if the principal reimburses this cost, the contract needs to mitigate the incentive for the agent to divert this fund for other purposes instead of switching on effort. Such a model poses additional challenges, and is left to future investigation.

## Endnotes

1. We use "work" and "exert effort" interchangeably in this paper.

2. Here we implicity assume that the continuous part of $L$, $L^c$ is absolutely continuous with repsect to the Lebesgue measure on $\mathbb{R}_+$.

3. Notation $W_t(\Gamma, \nu)$ represents the agent' continuation utility after observing either an arrival or a random switching that occurs at time $t$, which may trigger an instantaneous payment at time $t$. Hence, in its definition (4), we use the Lebesgue-Stieltjes integral $\int_{t+}^{\infty}$ to exclude the possible instantaneous payment at time $t$.

4. Technically speaking, all time indices in the d$t$ term in (PK) should be $t-$. However, it does not make any difference as there is no jump in the d$t$ term. This kind of confusion also appears in other places, making no harm to the results.

5. Condition 2 corresponds, but is not identical, to Equation (13) of Cao et al. (2021). The difference is due to model assumptions. The model in Cao et al. (2021) assumes that the effort cost $c$ is not immediately reimbursed, while it is in our model.

## References


Abreu, D., D. Pearce, and E. Stacchetti (1990). Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica 58*(5), 1041–1063.

Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica 78*(1), 73–118.

Brekke, K. A. and B. Oksendal (1994). Optimal switching in an economic activity under uncertainty. *SIAM Journal on Control and Optimization 32*(4), 1021–1036.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer Series in Statistics. Springer.

Cao, P., F. Tian, and P. Sun (2021). Comment on "optimal contract to induce continued effort". *Management Science*. Forthcoming.

Chen, M., P. Sun, and Y. Xiao (2020). Optimal monitoring schedule in dynamic contracts. *Operations Research 68*(5), 1285–1624.

Chen, W., S. Gupta, M. Dawande, and G. Janakiraman (2020). 3 years, 2 papers, 1 course off: Optimal non-monetary reward mechanisms. *Available at SSRN 3647569*.

Cvitanić, J., D. Possamaï, and N. Touzi (2016). Moral hazard in dynamic risk management. *Management Science 63*(10), 3328–3346.

Demarzo, P. M. and Y. Sannikov (2006). Optimal security design and dynamic capital structure in a continuous-time agency model. *Journal of Finance 61*(6), 2681–2724.

Duckworth, K. and M. Zervos (2001). A model for investment decisions with switching costs. *Annals of Applied Probability 11*(1), 239–260.

Georgiadis, G., Y. Kim, and H. D. Kwon (2020). The absence of attrition in the war of attrition under complete information. Working paper, Available at SSRN: https://ssrn.com/abstract=3353450 or http://dx.doi.org/10.2139/ssrn.3353450.

Grochulski, B. and Y. Zhang (2019). Termination as an incentive device. Working paper.

Hernandez Santibanez, N., D. Possamai, and C. Zhou (2020). Bank monitoring incentives under moral hazard and adverse selection. *Journal of Optimization Theory and Applications 184*(3), 988–1035.

Hu, Z. and W. Tang (2021). Size matters, so does duration: The interplay between offer size and offer deadline. *Management Science 67*(8), 4935–4960.

Kwon, H. D. (2019). Game of variable contributions to the common good under uncertainty. *Operations Research*. Forthcoming.

Kwon, H. D., W. Xu, A. Agrawal, and S. Muthulingam (2016). Impact of bayesian learning and externalities on strategic investment. *Management Science 62*(2), 550–570.

Myerson, R. B. (2015). Moral hazard in high office and the dynamics of aristocracy. *Econometrica 83*(6), 2083–2126.

Pages, H. (2013). Bank monitoring incentives and optimal abs. *Journal of Financial Intermediation 22*(1), 30–54.

Pages, H. and D. Possamai (2014). A mathematical treatment of bank monitoring incentives. *Finance and Stochastics 18*(1), 39–73.

Protter, P. E. (2003). *Stochastic Integration and Differential Equations* (second edition ed.), Volume 21 of *Stochastic Modelling and Applied Probability*. Springer.

Sannikov, Y. (2008). A continuous-time version of the principal-agent problem. *Review of Economic Studies 75*(3), 957–984.

Spear, S. and S. Srivastava (1987). On repeated moral hazard with discounting. *Review of Economic Studies 54*(4), 599–617.

Sun, P. and F. Tian (2018). Optimal contract to induce continued effort. *Management Science 64*(9), 4193–4217.

Tian, F., P. Sun, and I. Duenyas (2021). Optimal contract for machine repair and maintenance. *Operations Research 69*(3), 916–949.

Vath, V. L. and H. Pham (2007). Explicit solution to an optimal switching problem in the two-regime case. *SIAM Journal on Control and Optimization 46*(2), 395–426.

Vath, V. L., H. Pham, and S. Villeneuve (2008). A mixed singular/switching control problem for a dividend policy with reversible technology investment. *Annals of Applied Probability 18*(3), 1164–1200.

Zhu, J. Y. (2013). Optimal contracts with shirking. *Review of Economic Studies 80*(2), 812–839.

Zipkin, P. (2000). *Foundations of Inventory Management*. Springer Series in Statistics. McGraw-Hill.

Zorc, S. and I. Tsetlin (2020). Deadlines, offer timing, and the search for alternatives. *Operations Research 68*(3), 927–948.

Zorc, S., I. Tsetlin, and S. Hasija (2019). The when and how of delegated search. Working paper.

# Online Appendices for "Punish Underperformance with Resting – Optimal Dynamic Contracts in the Presence of Switching Cost"

In this document, we present some further discussions in Appendix A, and all proofs omitted in the paper in Appendix B.

## Appendix A: Further Discussions

Appendix A contains four parts. Appendix A.1 gives a heuristic derivation of the optimality condition (20)–(22) for the optimal value functions $V_{\mathsf{I}}$ and $V_{\emptyset}$, which appears in Section 3. Appendix A.2 demonstrates how to compute the optimal contract parameters. Besides, we discuss two extensions to the basic model studied in the paper. More precisely, in Appendix A.3, we consider the equal time discount case, and in Appendix A.4, we study the case in which there is also a fixed cost to switch from state $\mathsf{I}$ to $\emptyset$.

### A.1. A Heuristic Derivation of (20)–(22)

In this section we provide a heuristic derivation of the principal's utility function and of the main features of the optimal contract. Some arguments are borrowed from Section 4.1 in Biais et al. (2010). Let $F_{\mathsf{I}}(w)$ and $F_{\emptyset}(w)$ be the principal's optimal utility function that yields an agent's utility $w$ when the initial state is $\mathsf{I}$ and $\emptyset$, respectively.

First, we characterize the evolution of the principal's utility function $F_{\mathcal{E}_{t-}}(W_{t-})$. Since the principal discounts the future utility flow at rate $r$, his expected flow rate of utility at time $t$ is $rF_{\mathcal{E}_{t-}}(W_{t-})$. This must be equal to the sum of expected cash flow, the (possible) switching cost, and the expected rate of change in his continuation utility over $(t-\mathrm{d}t, t]$. Hence, we have

$$rF_{\mathcal{E}_{t-}}(W_{t-})\mathrm{d}t = [\bar{\nu}_t R - (c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}}]\mathrm{d}t - \mathrm{d}L_t + \mathbb{E}_{t-}[-\kappa(\mathcal{E}_{t-}, \mathcal{E}_t) + \mathrm{d}F_{\mathcal{E}_t}(W_t)], \tag{55}$$

where $\mathbb{E}_{t-}[\cdot] := \mathbb{E}[\cdot|\mathcal{F}_{t-}]$.

Following the discussions in Section 3, we assume that for any $\varepsilon \in \{\mathsf{I}, \emptyset\}$, $F_\varepsilon(\cdot)$ is concave and differentiable on $\mathbb{R}_+$. The actual value function might not be differentiable on the entire domain $\mathbb{R}_+$, which is an issue frequently arising in the optimal control literature, and often addressed by the viscosity solution approach. Since this section is devoted to a heuristic derivation of the optimality equation for the optimal utility function $F_\varepsilon$, we assume that $F_\varepsilon$ is smooth enough temporarily.

Recall that $\mathrm{d}L_t = \ell_t \mathrm{d}t + \Delta L_t$. Note that under any admissible contract, $\mathbb{1}_{\nu_t=\underline{\mu}} = \mathbb{1}_{\mathcal{E}_t=\emptyset}$ and $\mathbb{1}_{\nu_t=\mu} = \mathbb{1}_{\mathcal{E}_t=\mathsf{I}}$. Using (PK) and regarding $F_\varepsilon(w)$ as a function of $(w,\varepsilon)$, we are able to apply calculus of point process to the process $(W, \mathcal{E})$ to obtain

$$\begin{aligned}\mathrm{d}F_{\mathcal{E}_t}(W_t) = &\big(\rho W_{t-} + b\mathbb{1}_{\mathcal{E}_t=\mathsf{I}} - H_t\bar{\nu}_t + q_t H_t^q - \ell_t\big)F'_{\mathcal{E}_{t-}}(W_{t-})\mathrm{d}t \\ &+ \big[F_{\mathcal{E}_{t-}}(W_{t-} - \Delta L_t) - F_{\mathcal{E}_{t-}}(W_{t-})\big] + \big[F_{\mathcal{E}_{t-}}(W_{t-} + H_t) - F_{\mathcal{E}_{t-}}(W_{t-})\big]\mathrm{d}N_t \\ &+ \big[F_{\mathcal{E}_{t-}}(W_{t-} - H_t^q) - F_{\mathcal{E}_{t-}}(W_{t-})\big]\mathrm{d}Q_t + \big[F_{\mathcal{E}_t}(W_t) - F_{\mathcal{E}_{t-}}(W_t)\big].\end{aligned}$$

Plugging the above formula into (55) and using $\mathbb{E}_{t-}\mathrm{d}N_t = \bar{\nu}_t\mathrm{d}t$, and $\mathbb{E}_{t-}\mathrm{d}Q_t = q_t\mathrm{d}t$, we have

$$\begin{aligned} rF_{\mathcal{E}_{t-}}(W_{t-})\mathrm{d}t = \Big[ & R\bar{\nu}_t - (c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}} - \ell_t + \big(\rho W_{t-} + b\mathbb{1}_{\mathcal{E}_t=\mathsf{I}} - H_t\bar{\nu}_t + q_t H_t^q - \ell_t\big)F'_{\mathcal{E}_{t-}}(W_{t-}) \\ & + \big(F_{\mathcal{E}_{t-}}(W_{t-}+H_t) - F_{\mathcal{E}_{t-}}(W_{t-})\big)\bar{\nu}_t + \big(F_{\mathcal{E}_{t-}}(W_{t-}-H_t^q) - F_{\mathcal{E}_{t-}}(W_{t-})\big)q_t\Big]\mathrm{d}t \\ & - \Delta L_t + F_{\mathcal{E}_{t-}}(W_{t-}-\Delta L_t) - F_{\mathcal{E}_{t-}}(W_{t-}) + \mathbb{E}_{t-}\big[-\kappa(\mathcal{E}_{t-},\mathcal{E}_t) + F_{\mathcal{E}_t}(W_t) - F_{\mathcal{E}_{t-}}(W_t)\big]. \end{aligned} \tag{56}$$

Here, $\ell_t$, $\Delta L_t$, $H_t$, $H_t^q$, $q_t$ and $\mathcal{E}_t$ are all control variables. Besides, the contract might be terminated at time $t$ by paying off the promised utility to the agent instantaneously. Hence, we have $F_{\mathcal{E}_t}(W_t) \geqslant \underline{v} - W_t$. That is, $F_\varepsilon(w) \geqslant \underline{v} - w$ for any $w \in \mathbb{R}_+$ and $\varepsilon \in \{\mathsf{I}, \emptyset\}$.

We first optimize the constant-order terms on the right-hand side in (56). Considering that the optimized constant-order terms should be zero, we have

$$\max_{\Delta L_t \geqslant 0}\big\{-\Delta L_t + F_{\mathcal{E}_{t-}}(W_{t-}-\Delta L_t) - F_{\mathcal{E}_{t-}}(W_{t-})\big\} = 0, \text{ and} \tag{57}$$

$$\max_{\mathcal{E}_t \in \{\mathsf{I},\emptyset\}}\big\{-\kappa(\mathcal{E}_{t-},\mathcal{E}_t) + F_{\mathcal{E}_t}(W_t) - F_{\mathcal{E}_{t-}}(W_t)\big\} = 0. \tag{58}$$

Equation (57) yields that $F'_\varepsilon(w) \geqslant -1$ for any $w \in \mathbb{R}_+$ and $\varepsilon \in \{\mathsf{I}, \emptyset\}$. Let $\widehat{w}_\varepsilon = \inf\{w \geqslant 0 : F'_\varepsilon(w) = -1\}$. The concavity of $F_\varepsilon$ implies that it is optimal for the principal to pay $\Delta L_t = \max\{W_{t-} - \widehat{w}_{\mathcal{E}_{t-}}, 0\}$ instantaneously to the agent.

Equation (58) yields that $F_{\mathsf{I}}(w) \geqslant F_\emptyset(w)$ and $F_\emptyset(w) \geqslant F_{\mathsf{I}}(w) - K$ for any $w \in \mathbb{R}_+$. Besides, $\mathcal{E}_t \neq \mathcal{E}_{t-}$ only if $-\kappa(\mathcal{E}_{t-}, \mathcal{E}_{t-}^c) + F_{\mathcal{E}_{t-}^c}(W_t) - F_{\mathcal{E}_{t-}}(W_t) = 0$, where $\varepsilon^c$ is $\mathsf{I}$ if $\varepsilon = \emptyset$ and is $\emptyset$ if $\varepsilon = \mathsf{I}$.

Next, we consider the controls such that $\Delta L_t = 0$ and $\mathcal{E}_t = \mathcal{E}_{t-}$. If we plug these values into (56), the symbol "=" should be replaced by "⩽" due to the suboptimality of these controls. Comparing the d$t$-order terms on both sides of the resulting inequality yields

$$\begin{aligned} rF_{\mathcal{E}_{t-}}(W_{t-}) \geqslant \max\Big\{ & R\bar{\nu}_t - (c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}} - \ell_t + \big(\rho W_{t-} + b\mathbb{1}_{\mathcal{E}_t=\mathsf{I}} - H_t\bar{\nu}_t + H_t^q q_t - \ell_t\big)F'_{\mathcal{E}_{t-}}(W_{t-}) \\ & + \big(F_{\mathcal{E}_{t-}}(W_{t-}+H_t) - F_{\mathcal{E}_{t-}}(W_{t-})\big)\bar{\nu}_t + \big(F_{\mathcal{E}_{t-}}(W_{t-}-H_t^q) - F_{\mathcal{E}_{t-}}(W_{t-})\big)q_t\Big\}, \end{aligned} \tag{59}$$

where the maximization is taken over the set of controls $(\ell_t, H_t, H_t^q, q_t)$ that satisfies $\ell_t \geqslant b\mathbb{1}_{\mathcal{E}_t=\mathsf{I}}$, the IR constraint (5), and the IC constraint (IC).

Inequality (59) can be written as two inequalities, for working and resting respectively. If $\mathcal{E}_{t-} = \mathsf{I}$, by omitting the time index, (59) becomes

$$\begin{aligned} rF_{\mathsf{I}}(w) \geqslant\ & R\mu - (c-b) + (\rho w + b)F'_{\mathsf{I}}(w) \\ & + \max\Big\{-\ell - (\ell + \mu h - qh^q)F'_{\mathsf{I}}(w) + \mu(F_{\mathsf{I}}(w+h) - F_{\mathsf{I}}(w)) + (F_{\mathsf{I}}(w-h^q) - F_{\mathsf{I}}(w))q\Big\}, \end{aligned} \tag{60}$$

where the maximization is taken over the set of $(\ell, h, h^q, q)$ that satisfies

$$\ell \geqslant b,\ \ h \geqslant \beta,\ \ h^q \leqslant w,\ \ q \geqslant 0. \tag{61}$$

If $\mathcal{E}_{t-} = \emptyset$, then (59) becomes

$$\begin{aligned} rF_\emptyset(w) \geqslant R\underline{\mu} + \rho wF'_\emptyset(w) + \max\Big\{ & -\ell - (\ell + \underline{\mu}h - qh^q)F'_\emptyset(w) + \underline{\mu}\big(F_\emptyset(w+h) - F_\emptyset(w)\big) \\ & + \big(F_\emptyset(w-h^q) - F_\emptyset(w)\big)q\Big\}, \end{aligned} \tag{62}$$

where the maximization is taken over the set of $(\ell, h, h^q, q)$ that satisfies

$$\ell \geqslant 0,\ \ h \geqslant -w,\ \ h^q \leqslant w,\ \ q \geqslant 0. \tag{63}$$

Recall that $V_{\mathsf{I}}(w) = F_{\mathsf{I}}(w) + w$ and $V_{\emptyset}(w) = F_{\emptyset}(w) + w$. Then, we have the following basic properties of $V_{\mathsf{I}}$ and $V_{\emptyset}$:

1. $V_{\mathsf{I}}(w) \geqslant \underline{v}$ and $V_{\emptyset}(w) \geqslant \underline{v}$ for any $w \in \mathbb{R}_+$.
2. $V_{\mathsf{I}}'(w) \geqslant 0$ and $V_{\emptyset}'(w) \geqslant 0$ for any $w \in \mathbb{R}_+$ (this follows from the fact that $F_{\mathsf{I}}'(w) \geqslant -1$ and $F_{\emptyset}'(w) \geqslant -1$).
3. Both $V_{\mathsf{I}}$ and $V_{\emptyset}$ are concave on $\mathbb{R}_+$.
4. $V_{\mathsf{I}}$ (resp. $V_{\emptyset}$) will take constant value on $[\widehat{w}_{\mathsf{I}}, \infty)$ (resp. $[\widehat{w}_{\emptyset}, \infty)$).
5. $V_{\mathsf{I}}(w) \geqslant V_{\emptyset}(w)$ and $V_{\emptyset}(w) \geqslant V_{\mathsf{I}}(w) - K$ for any $w \in \mathbb{R}_+$.

We first analyze (60), which can be rewritten as follows in terms of $V_{\mathsf{I}}$:

$$\begin{aligned} rV_{\mathsf{I}}(w) \geqslant R\mu - c - (\rho - r)w + (\rho w + b)V_{\mathsf{I}}'(w) + \max\Big\{ -\ell V_{\mathsf{I}}'(w) + \big(V_{\mathsf{I}}(w+h) - V_{\mathsf{I}}(w) - hV_{\mathsf{I}}'(w)\big)\mu \\ + \big(V_{\mathsf{I}}(w - h^q) - V_{\mathsf{I}}(w) + h^q V_{\mathsf{I}}'(w)\big)q\Big\}, \end{aligned} \tag{64}$$

where the maximization is taken over the constraints (61).

Optimizing the right-hand side of (64) with respect to $\ell$, we have $\ell^* = \arg\max_{\ell \geqslant b}\{-\ell V_{\mathsf{I}}'(w)\} = b$ if $w \in [0, \widehat{w}_{\mathsf{I}})$, where we use the fact that $V_{\mathsf{I}}'(w) > 0$ for $w \in [0, \widehat{w}_{\mathsf{I}})$.

Optimizing the right-hand side of (64) with respect to $h$, we have $h^* = \arg\max_{h \geqslant \beta}\{V_{\mathsf{I}}(w+h) - V_{\mathsf{I}}'(w)h\} = \beta$, by noting that $V_{\mathsf{I}}(w+h) - V_{\mathsf{I}}'(w)h$ is decreasing in $h$ on $[0, \infty)$, since $V_{\mathsf{I}}'(w+h) - V_{\mathsf{I}}'(w) \leqslant 0$ for any $h \geqslant 0$ due to the concavity of $V_{\mathsf{I}}$.

Note that $\max_{h^q \leqslant w}\{V_{\mathsf{I}}(w - h^q) - V_{\mathsf{I}}(w) + h^q V_{\mathsf{I}}'(w)\} = 0$. Hence, (64) reduces to

$$rV_{\mathsf{I}}(w) \geqslant R\mu - c - (\rho - r)w - \rho(\bar{w} - w)V_{\mathsf{I}}'(w) + \mu(V_{\mathsf{I}}(w+\beta) - V_{\mathsf{I}}(w)), \tag{65}$$

for $w \in \mathbb{R}_+$, which can be rewritten as $(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) \geqslant 0$ by using the operator $\mathcal{A}_{\mathsf{I}}$ defined in (18).

We next analyze (62), which can be rewritten as follows in terms of $V_{\emptyset}$:

$$\begin{aligned} rV_{\emptyset}(w) \geqslant R\underline{\mu} - (\rho - r)w + \rho w V_{\emptyset}'(w) + \max\Big\{ -\ell V_{\emptyset}'(w) + \underline{\mu}\big(V_{\emptyset}(w+h) - V_{\emptyset}(w) - hV_{\emptyset}'(w)\big) \\ + \big(V_{\emptyset}(w - h^q) - V_{\emptyset}(w) + h^q V_{\emptyset}'(w)\big)q\Big\}, \end{aligned} \tag{66}$$

where the maximization is taken over the constraints (63).

Optimizing the right-hand side of (66) with respect to $\ell$, we have $\ell^* = \arg\max_{\ell \geqslant 0}\{-\ell V_{\emptyset}'(w)\} = 0$ if $w \in [0, \widehat{w}_{\emptyset})$. Optimizing the right-hand side of (66) with respect to $h$, we have $h^* = \arg\max_{h \geqslant -w}\{-V_{\emptyset}'(w)h + V_{\emptyset}(w+h)\} = 0$, by noting that $-V_{\emptyset}'(w)h + V_{\emptyset}(w+h)$ is increasing in $h$ for $h < 0$ and decreasing in $h$ for $h > 0$ due to the concavity of $V_{\emptyset}$. Also, we have $\max_{h^q \leqslant w}\{V_{\emptyset}(w - h^q) - V_{\emptyset}(w) + h^q V_{\emptyset}'(w)\} = 0$. Consequently, (66) can further reduce to

$$rV_{\emptyset}(w) \geqslant R\underline{\mu} - (\rho - r)w + \rho w V_{\emptyset}'(w), \tag{67}$$

where can be rewritten as $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) \geqslant 0$.

Summarizing the above discussions leads us to consider (20)–(22).

### A.2. Computing Contract Parameters

For $K = 0$, we have the following result.

PROPOSITION 8. *(i) Under Condition 1 and $\underline{K}_1 > 0$, we have $\theta_0 = \underline{\theta}^0$, where $\theta_0$ and $\underline{\theta}^0$ are defined in Proposition 7 and Lemma 8, respectively. Correspondingly, we have $\hat{\mathbf{w}}_0 = \widetilde{w}(\theta_0)$ and $\mathsf{c}_0 = C(\theta_0)$, in which functions $\widetilde{w}(\cdot)$ and $C(\cdot)$ are defined in Lemma 7.*

*(ii) Under Condition 2 and $\breve{K}_2 > 0$, define a lower bound*

$$\breve{\underline{\theta}} := \frac{(\bar{V} - \underline{v})(\rho - r - \mu) - (\rho - r)\bar{w}}{\mu(\rho/r - 1)}.$$

*Similar to Lemmas 7 and 8, for any $\underline{\theta} \in (\breve{\underline{\theta}}, \bar{w})$, there exist unique values $\widetilde{w}(\underline{\theta}) \in (\underline{\theta}, \bar{w})$ and $C(\underline{\theta})$, such that if we set $\hat{\mathbf{w}} = \widetilde{w}(\underline{\theta})$, $\mathsf{c} = C(\underline{\theta})$, and $\underline{\vartheta} = \underline{\theta}$, the value-matching and smooth-pasting conditions (30) and (31) are satisfied. Furthermore, value $\underline{\theta}^0 := \inf\{\underline{\theta} \in (\breve{\underline{\theta}}, \bar{w}) : \widetilde{w}'(\underline{\theta}) \geqslant 0\}$ is well-defined, and we have $\theta_0 = \underline{\theta}^0$, $\hat{\mathbf{w}}_0 = \widetilde{w}(\theta_0)$, and $\mathsf{c}_0 = C(\theta_0)$.*

For any $\underline{\theta} \in (0, \bar{w})$, function $h(\widetilde{w}, \underline{\theta})$, as defined in (93), is decreasing in $\widetilde{w}$ with $h(\widetilde{w}(\underline{\theta}), \underline{\theta}) = 0$. Hence, $\widetilde{w}(\underline{\theta})$ can be efficiently found by a binary search procedure, starting from lower bound $\underline{\theta}$ and upper bound $\bar{w}$. Consequently, $C(\underline{\theta})$ can also be immediately computed as $C_1(\widetilde{w}(\underline{\theta}), \underline{\theta})$, with $C_1(\widetilde{w}, \underline{\theta})$ defined in (92). Therefore, following Proposition 8, in order to determine the optimal contract parameters for $K = 0$ under Condition 3, we only need to find $\underline{\theta}^0$. Based on the definition of $\underline{\theta}^0$ (see part (ii) of Proposition 8), this value can be determined by a line search to check at which point $\widetilde{w}(\underline{\theta})$ is no longer increasing, starting from 0 under Condition 1 and $\underline{K}_1 > 0$, or from $\breve{\underline{\theta}}$ under Condition 2 and $\breve{K}_2 > 0$.

Computation of the optimal contract parameters for $K > 0$ is more complex. We only demonstrate how to compute the control band parameters $(\mathsf{c}, \hat{\mathbf{w}}, \bar{\vartheta}, \underline{\vartheta})$ under Conditions 1 and (L1) or under Conditions 2, (L2), and (mL), as the optimal contract in other cases takes a simpler form. Take the case under Conditions 1 and (L1) for illustration. Note that for any $\underline{\theta} \in (0, \underline{\theta}^0)$, the value $\bar{\theta}(\underline{\theta})$ can be determined by (89), using a line search procedure. Hence, function $\psi(\underline{\theta})$, as defined in (90), can be readily computed for each $\underline{\theta} \in (0, \underline{\theta}^0)$. Since by Lemma 10 function $\psi(\underline{\theta})$ is decreasing in $\underline{\theta}$ with $\psi(\underline{\vartheta}) = K$, the quantity $\underline{\vartheta}$ can be efficiently found by a binary search procedure, starting from lower bound 0 and upper bound $\underline{\theta}^0$. The three other parameters, $\mathsf{c}$, $\hat{\mathbf{w}}$, $\bar{\vartheta}$, are thus immediately computed as $C(\underline{\vartheta})$, $\widetilde{w}(\underline{\vartheta})$ and $\bar{\theta}(\underline{\vartheta})$. For the case under Conditions 2, (L2), and (mL), the only difference is that initial lower bound for the binary search is $\breve{\underline{\theta}}$.

The above procedure can be summarized by the following four subroutines.

**Subroutine 1.** Given $\underline{\theta} \in (0, \bar{w})$, compute $\tilde{w}(\underline{\theta})$: Binary search on $[\underline{\theta}, \tilde{w}]$, to determine $\tilde{w}(\underline{\theta})$ according to $h(\tilde{w}(\underline{\theta}), \theta) = 0$ where function $h(\tilde{w}, \underline{\theta})$ is defined in (93).

**Subroutine 2.** Given $\underline{\theta} \in (0, \bar{w})$, compute $C(\underline{\theta})$: Following **Subroutine 1**, we obtain $\tilde{w}(\underline{\theta})$. Then, $C(\underline{\theta}) = C_1(\tilde{w}(\underline{\theta}), \underline{\theta})$ with $C_1(\tilde{w}, \underline{\theta})$ defined in (92).

**Subroutine 3.** Given $\underline{\theta} \in (0, \underline{\theta}^0)$, compute $\bar{\theta}(\underline{\theta})$: Following **Subroutines 1** and **2**, we obtain $\tilde{w}(\underline{\theta})$ and $C(\underline{\theta})$. Then, we calculate $\bar{\theta}(\underline{\theta})$ by (89), using a line search procedure.

**Subroutine 4.** Given $\underline{\theta} \in (0, \underline{\theta}^0)$, compute $\psi(\underline{\theta})$: Following **Subroutines 1–3**, we obtain $\tilde{w}(\underline{\theta})$, $C(\underline{\theta})$ and $\bar{\theta}(\underline{\theta})$. Then, we compute $\psi(\underline{\theta})$, as defined in (90).

With the above four steps, the optimal control band parameters can be computed by Algorithm 1 below.

**Algorithm 1** Compute $(\mathsf{c}, \hat{\mathbf{w}}, \bar{\vartheta}, \underline{\vartheta})$.

1: Line search to determine $\underline{\theta}^0$ according to $\tilde{w}'(\theta) = 0$, in which function $\tilde{w}(\theta)$ is computed according to **Subroutine 1**.

2: Binary search to determine $\underline{\vartheta}$ according to $\psi(\underline{\vartheta}) = K$ where $\psi(\underline{\vartheta})$ can be computed following **Subroutine 4**.

3: Following **Subroutines 1–3**, we obtain $\hat{\mathbf{w}} = \tilde{w}(\underline{\vartheta})$, $\mathsf{c} = C(\underline{\vartheta})$ and $\underline{\vartheta} = \bar{\theta}(\underline{\vartheta})$, respectively.

### A.3. Equal Discount Rate

In the study of dynamic contracts without the switching options, Sun and Tian (2018) claimed, without a formal proof, that under equal discount rates, it is optimal for the principal to always induce the agent to work before contract termination. In our context with switching, this claim corresponds to never switching the agent to resting and then working again. Here, we provide a formal proof that validates this claim, for any $K \geqslant 0$.

When the two player's discount rates are the same, that is, $r = \rho$, various expressions in the main part of the paper become simpler. For example, the value $\bar{V}$ defined in (53) becomes

$$\bar{V}_e := \frac{\mu R - c}{r}, \tag{68}$$

and the differential equation (24), which plays an essential role in deciding the optimal value functions, becomes

$$0 = (\mu + r)V_e(w) - \mu V_e((w + \beta) \wedge \bar{w}) + r(\bar{w} - w)V_e'(w) - (\mu R - c). \tag{69}$$

According to Lemma 3 of Sun and Tian (2018), differential equation (69) with boundary condition $V_e(0) = \underline{v}$ has a unique solution $V_e$ on $[0, \bar{w}]$, which is increasing and strictly concave, with $V_e(w) = \bar{V}_e$ for all $w \geqslant \bar{w}$. Theorem 1 still holds, in which the operators $\mathcal{A}_{\mathsf{l}}$ and $\mathcal{A}_{\emptyset}$ are simplified to

$$(\mathcal{A}_{\mathsf{l}} f)(w) = (\mu + r)f(w) - \mu f(w + \beta) + r(\bar{w} - w)f'(w) - (\mu R - c), \text{ and}$$
$$(\mathcal{A}_{\emptyset} f)(w) = rf(w) - rwf'(w) - R\underline{\mu},$$

respectively, for differentiable function $f$.

Furthermore, when $r = \rho$, effectively Condition 1 holds. Consequently, analysis for the equal discount case resembles that in Section 4. In particular, we will show that the value function for state $\mathsf{l}$ is $V_e$ defined above. Furthermore, the upper thresholds $\bar{K}$ of (H1) becomes

$$\bar{K}_e := \bar{V}_e - \underline{v}. \tag{70}$$

In order to define the lower threshold for the switching cost, we need to define the value function for state $\emptyset$. Note that when $r = \rho$, function $\mathcal{V}_{\hat{\mathrm{w}}}$ becomes $V_e$, with $\hat{\mathrm{w}}$ being $\bar{w}$ and $\breve{w}(\hat{\mathrm{w}})$ being 0. Hence, following Lemma 4, if $K < \bar{K}_e$, there exists $K$-dependent values $\bar{\theta}^K \in [0, \bar{w}]$ and $m^K \in [0, V_e'(0)]$ such that

$$V_e(\bar{\theta}^K) = m^K \bar{\theta}^K + K + \underline{v}, \text{ and } V_e'(\bar{\theta}^K) = m^K.$$

Then, similar to (29), we define the following societal value function for the resting state,

$$V_\emptyset(w) = \begin{cases} m^K w + \underline{v}, & w \in \left[0, \bar{\theta}^K\right], \\ V_e(w) - K, & w \in \left[\bar{\theta}^K, \bar{w}\right]. \end{cases} \tag{71}$$

Figure 7 depicts the value functions. It is clear that $V_\emptyset$ is linear over the interval $[0, \bar{\theta}^K]$. Furthermore, $V_\mathsf{I}(w)$ and $V_\emptyset(w)$ are "parallel" with a difference of $K$ for $w \geqslant \bar{\theta}^K$.

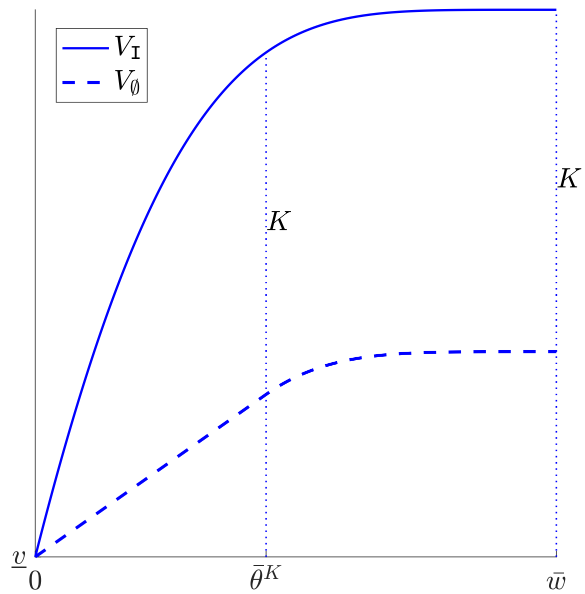



**Figure 7 Value functions with** $r = 0.5$**,** $\rho = 0.5$**,** $c = b = 0.2$**,** $R = 2$**,** $\Delta\mu = 0.7$**,** $K = 1.5$**, and** $\mu = 2$**. In this case,** $\bar{\theta}^K = 0.51$**,** $\bar{w} = 1.14$**,** $\bar{V}_e = 7.6$ **and** $\underline{v} = 5.2$**.**

The following theorem summarizes the optimality results.

Theorem 9. *Consider* $r = \rho$*. For any* $w \geqslant 0$*, we have*

$$U(\Gamma^*(w, \mathsf{I}; 0, 0, \bar{w}, \bar{w}), \mathsf{I}) = V_e(w), \text{ and } U(\underline{\Gamma}, \emptyset) = \underline{v}.$$

*If* $K \geqslant \bar{K}_e$*, functions* $V_\mathsf{I} = V_e$ *and* $V_\emptyset = \underline{v}$ *satisfy* (20)–(22).

*If* $K < \bar{K}_e$*, on the other hand, functions* $V_\mathsf{I} = V_e$ *and* $V_\emptyset$ *as defined in* (71) *satisfy* (20)–(22)*. Furthermore, if* $V_e'(\bar{\theta}^K) > 1$*, for any* $w \geqslant \bar{\theta}^K$ *we have*

$$U(\Gamma^*(w, \mathsf{I}; 0, 0, \bar{w}, \bar{w}), \emptyset) = V_\emptyset(w) - w.$$

Therefore, in the equal discount case, contract $\Gamma^*(w_e^*, \mathsf{I}; 0, 0, \bar{w}, \bar{w})$ is optimal for the initial state $\mathsf{I}$, as well as for the initial state $\emptyset$, if $K < \bar{K}_e$ and $m^K > 1$, in which $w_e^* \in [0, \bar{w}]$ is the unique maximizer of function $V_e$ such that $w_e^* > \bar{\theta}^K$. If $K \geqslant \bar{K}_e$, or $K < \bar{K}_e$ and $m^K \leqslant 1$, on the other hand, it is optimal for the principal not to hire the agent for the initial state $\emptyset$. Note that because the threshold $\underline{\theta}$ in contract $\Gamma^*(w, \mathsf{I}; 0, 0, \bar{w}, \bar{w})$ is zero, the principal does not direct the agent to stop working until the promised utility has reached 0. At this point the promised utility cannot become positive again and the contract is terminated. Therefore, in all these cases, it is never optimal for the principal to direct the agent to stop working and restart later.

### A.4. Positive Switching Cost From On to Off

Now we briefly discuss a generalization of our basic model, which involves a fixed cost, call it $\mathcal{K}$, for the principal to direct the agent to stop working, including terminating the contract. Instead of providing a comprehensive summary of all results, we provide the key ideas and leave some details for the reader to fill in.

The general contract structure, $\Gamma^*$ of Definition 1, remains optimal. In order to identify the specific parameters of the policy structure, we describe the optimal value functions.

First of all, in the verification theorem, condition (21) is revised to $-\mathcal{K} \leqslant V_{\mathsf{I}} - V_{\emptyset} \leqslant K$, and the second inequality in (22) changes to $V_{\emptyset}(0) \geqslant \underline{v} - \mathcal{K}$. The key idea for constructing the value functions is that when $w < \underline{\theta}$, function $V_{\mathsf{I}}$ is a downward parallel shift of $V_{\emptyset}$ by $\mathcal{K}$. Accordingly, the value-matching and smooth-pasting conditions of Proposition 4 become

$$V_{\mathsf{I}}(\underline{\theta}) = V_{\emptyset}(\underline{\theta}) - \mathcal{K}, \quad V_{\mathsf{I}}'(\underline{\theta}) = V_{\emptyset}'(\underline{\theta}),$$
$$V_{\mathsf{I}}(\bar{\theta}) = V_{\emptyset}(\bar{\theta}) + K, \quad \text{and} \quad V_{\mathsf{I}}'(\bar{\theta}) = V_{\emptyset}'(\bar{\theta}).$$

Figure 8 depicts the value functions. Similar to Figure 4, function $V_{\mathsf{I}}(w)$ is linear in the interval $w \in [\underline{\theta}, \breve{w}]$. Furthermore, for $w \leqslant \underline{\theta}$, function $V_{\mathsf{I}}(w)$ is a downward parallel shift from $V_{\emptyset}(w)$ by $\mathcal{K}$, while for $w \geqslant \bar{\theta}$, function $V_{\emptyset}(w)$ is a downward parallel shift from $V_{\mathsf{I}}(w)$ by $K$.

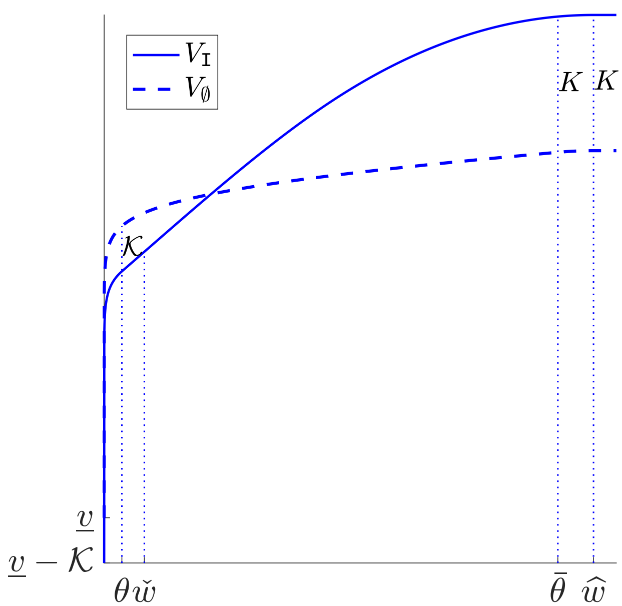



**Figure 8** **Value functions with** $r = 0.05$, $\rho = 1$, $c = b = 0.3$, $R = 112$, $\Delta\mu = 0.1$, $\mathcal{K} = 10$, $K = 30$, **and** $\mu = 1.95$. **In this case,** $\bar{w} = 5.85$, $\underline{\theta} = 0.2$, $\breve{w} = 0.46$, $\bar{\theta} = 5.21$, **and** $\widehat{w} = 5.62$.

## Appendix B: Proofs of all the Results

This section collects all proofs for the results in the paper and in Appendix A.

### B.1. Proofs of the Results in Section 2

*Proof of Proposition 1.* The proof of part (i) is exactly the same as that of Proposition 1 in Cao et al. (2021), in which random termination instead of random switching may take place. The proof of part (ii) is similar to that of Lemma 6 in Sun and Tian (2018). Hence, we omit both of them for brevity. □

*Proof of Lemma 1.* If we can show that (PK) holds under contract $\Gamma^*(w_0, \varepsilon_0; \underline{\theta}, \breve{w}, \bar{\theta}, \widehat{w})$, then (17) follows immediately from (2) and (4) with $t = 0$. In fact, (PK) holds by setting $H_t = \beta \mathbb{1}_{\mathcal{E}_{t-}=\mathsf{I}}$ and $H_t^q = (\breve{w} - \underline{\theta})\, \mathbb{1}_{W_{t-}=\breve{w}, \mathcal{E}_{t-}=\mathsf{I}}$. □

### B.2. Proofs of the Results in Section 3

*Proof of Theorem 1.* Fix any contract $\Gamma \in \mathfrak{C}$. For simplicity, we omit $\Gamma$ and $\nu$ from quantities of interest. The agent's promised utility follows a process $W$ whose dynamics is described by (PK) with $\nu_t = \mu$ for $\mathcal{E}_t = \mathsf{I}$ and $\nu_t = \underline{\mu}$ for $\mathcal{E}_t = \emptyset$.

Recall that $\mathrm{d}L_t = \ell_t \mathrm{d}t + \Delta L_t$. Write $\phi(w, \varepsilon) = V_\varepsilon(w) - w$ for any $w \in \mathbb{R}_+$ and $\varepsilon \in \{\mathsf{I}, \emptyset\}$. Applying the change of variable formula (see, for example, Theorem 70 of Chapter IV in Protter 2003, pp. 214) for processes of locally bounded variation to the process $(W, \mathcal{E})$ and using (PK), we have

$$e^{-rT}\phi(W_T, \mathcal{E}_T) = \phi(W_{0-}, \mathcal{E}_{0-}) + \int_{0+}^{T} e^{-rt}\Big[(\rho W_{t-} + b\mathbb{1}_{\nu_t=\mu} - H_t\nu_t + q_t H_t^q - \ell_t)\cdot D_{t-} \\ - rV_{\mathcal{E}_{t-}}(W_{t-})\Big]\mathrm{d}t + \sum_{0\leqslant t\leqslant T} e^{-rt}\Delta\phi(W_t, \mathcal{E}_t),$$

for any $T \geqslant 0$, where $D_{t-}$ is the left-derivative of $\phi(w, \mathcal{E}_{t-})$ with respect to $w$ at $W_{t-}$, that is, $D_{t-} = V'_{\mathcal{E}_{t-}}(W_{t-}) - 1$, by recalling that we use $f'(w)$ to represent the left-derivative of $f$ at $w$ for any absolutely continuous function defined on $\mathbb{R}_+$. Besides, we have

$$\begin{aligned}\Delta\phi(W_t, \mathcal{E}_t) =& \phi(W_t, \mathcal{E}_t) - \phi(W_t, \mathcal{E}_{t-}) \\ &+ \phi(W_{t-} + H_t\mathrm{d}N_t - H_t^q\mathrm{d}Q_t - \Delta L_t, \mathcal{E}_{t-}) - \phi(W_{t-} + H_t\mathrm{d}N_t - H_t^q\mathrm{d}Q_t, \mathcal{E}_{t-}) \\ &+ \phi(W_{t-} + H_t\mathrm{d}N_t - H_t^q\mathrm{d}Q_t, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-}) \text{ for } t > 0,\end{aligned}$$

and

$$\Delta\phi(W_0, \mathcal{E}_0) = \phi(W_0, \mathcal{E}_0) - \phi(W_0, \mathcal{E}_{0-}) + \phi(W_0, \mathcal{E}_{0-}) - \phi(W_{0-}, \mathcal{E}_{0-})$$

by noting that $\mathrm{d}N_0 = \mathrm{d}Q_0 = 0$ with probability 1.

Define $M^N = \{M_t^N\}_{t\geqslant 0}$ and $M^Q = \{M_t^Q\}_{t\geqslant 0}$ by

$$M_t^N = N_t - \int_0^t \nu_s \mathrm{d}s,\ \ M_t^Q = Q_t - \int_0^t q_s \mathrm{d}s.$$

Note that

$$
\begin{aligned}
&\sum_{0<t\leqslant T}\big[\phi(W_{t-}+H_t\mathrm{d}N_t-H_t^q\mathrm{d}Q_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})\big]\\
&=\int_{0+}^T e^{-rt}\Big\{[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}N_t+[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}Q_t\Big\}\\
&=\int_{0+}^T e^{-rt}\big[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})\big]\mathrm{d}M_t^N+\int_{0+}^T e^{-rt}\big[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})\big]\nu_t\mathrm{d}t\\
&\quad+\int_{0+}^T e^{-rt}\big[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})\big]\mathrm{d}M_t^Q+\int_{0+}^T e^{-rt}\big[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})\big]q_t\mathrm{d}t.
\end{aligned}
$$

where the first equality uses the fact that $\{t\in[0,T]:\mathrm{d}N_t=\mathrm{d}Q_t=1\}$ has a Lebesgue measure 0 with probability 1. This suggests that

$$
\begin{aligned}
e^{-rT}\phi(W_T,\mathcal{E}_T)=&\phi(W_{0-},\mathcal{E}_{0-})+\int_{0+}^T e^{-rt}[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^N+\\
&+\int_{0+}^T e^{-rt}[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^Q+A_1+A_2+A_3+A_4+A_5,
\end{aligned}
\tag{72}
$$

where

$$
\begin{aligned}
A_1:=&\int_{0+}^T e^{-rt}\Big\{(\rho W_{t-}+b\mathbb{1}_{\nu_t=\mu}-H_t\nu_t-\ell_{t-})\cdot\big(V'_{\mathcal{E}_{t-}}(W_{t-})-1\big)-r\phi(W_{t-},\mathcal{E}_{t-})\\
&+[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\nu_t\Big\}\mathrm{d}t,\\
A_2:=&\sum_{0<t\leqslant T}e^{-rt}\Big[\phi\big(W_{t-}+H_t\mathrm{d}N_t-H_t^q\mathrm{d}Q_t-\Delta L_t,\mathcal{E}_{t-}\big)-\phi\big(W_{t-}+H_t\mathrm{d}N_t-H_t^q\mathrm{d}Q_t,\mathcal{E}_{t-}\big)\Big],\\
A_3:=&\sum_{0\leqslant t\leqslant T}e^{-rt}[\phi(W_t,\mathcal{E}_t)-\phi(W_t,\mathcal{E}_{t-})],\\
A_4:=&\int_{0+}^T e^{-rt}q_t\Big\{H_t^q\big(V'_{\mathcal{E}_{t-}}(W_{t-})-1\big)+\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})\Big\}\mathrm{d}t,\\
A_5:=&\phi(W_0,\mathcal{E}_{0-})-\phi(W_{0-},\mathcal{E}_{0-}).
\end{aligned}
$$

Below we treat each term separately.

Consider first $A_1$. If $\mathcal{E}_{t-}=\mathsf{l}$, then $\nu_{t-}=\mu$ and $\phi(W_{t-},\mathcal{E}_{t-})=V_{\mathsf{l}}(W_{t-})-W_{t-}$. Since the contract $\Gamma$ is incentive compatible, we have $H_t\geqslant\beta$ from Proposition 1 (ii). Consequently, we have

$$
\begin{aligned}
&(\rho W_{t-}+b\mathbb{1}_{\nu_t=\mu}-H_t\nu_t-\ell_t)\cdot\big(V'_{\mathcal{E}_{t-}}(W_{t-})-1\big)-r\phi(W_{t-},\mathcal{E}_{t-})+[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\nu_t\\
=&(\rho W_{t-}+b-H_t\mu-\ell_t)\cdot\big(V_{\mathsf{l}}'(W_{t-})-1\big)-r\cdot(V_{\mathsf{l}}(W_{t-})-W_{t-})+[V_{\mathsf{l}}(W_{t-}+H_t)-V_{\mathsf{l}}(W_{t-})-H_t]\cdot\mu\\
=&\rho W_{t-}\cdot\big(V_{\mathsf{l}}'(W_{t-})-1\big)-r\cdot(V_{\mathsf{l}}(W_{t-})-W_t)-(\ell_t-b)\cdot\big(V_{\mathsf{l}}'(W_{t-})-1\big)\\
&+\big[V_{\mathsf{l}}(W_{t-}+H_t)-V_{\mathsf{l}}(W_{t-})-V_{\mathsf{l}}'(W_{t-})H_t\big]\cdot\mu\\
\leqslant&\rho W_{t-}\cdot(V_{\mathsf{l}}'(W_{t-})-1)-r\cdot(V_{\mathsf{l}}(W_{t-})-W_{t-})+\ell_t-b+[V_{\mathsf{l}}(W_{t-}+\beta)-V_{\mathsf{l}}(W_{t-})-V_{\mathsf{l}}'(W_{t-})\beta]\cdot\mu\\
=&-\Big[(\mu+r)V_{\mathsf{l}}(W_{t-})-\mu V_{\mathsf{l}}(W_{t-}+\beta)+\rho(\bar{w}-W_{t-})V_{\mathsf{l}}'(W_{t-})-(\mu R-c)+(\rho-r)W_{t-}\Big]+\ell_t-[R\mu-(c-b)]\\
=&-(\mathcal{A}_{\mathsf{l}}V_{\mathsf{l}})(W_{t-})+\ell_t-[R\mu-(c-b)]\\
\leqslant&\ell_t-[R\mu-(c-b)].
\end{aligned}
$$

In the above, the first inequality follows from (i) $V_{\mathsf{I}}(W_{t-}) \geqslant 0$ (this follows from the fact that $V_{\mathsf{I}}$ is nondecreasing) and (ii) $H_t \geqslant \beta$, and $\beta = \arg\max_{h\geqslant\beta}\{V_{\mathsf{I}}(w+h) - V_{\mathsf{I}}(w) - V_{\mathsf{I}}'(w)\cdot h\}$ due to the concavity of $V_{\mathsf{I}}$; the last inequality follows from (20).

If $\mathcal{E}_{t-} = \emptyset$, then $\nu_{t-} = \underline{\mu}$. It follows from (5) that $H_t \geqslant -W_{t-}$. Therefore, we have

$$
\begin{aligned}
&(\rho W_{t-} + b\mathbb{1}_{\nu_t=\mu} - H_t\nu_t - \ell_t)\cdot\big(V'_{\mathcal{E}_{t-}}(W_{t-}) - 1\big) - r\phi(W_{t-},\mathcal{E}_{t-}) + [\phi(W_{t-}+H_t,\mathcal{E}_{t-}) - \phi(W_{t-},\mathcal{E}_{t-})]\nu_t\\
=&(\rho W_{t-} - H_t\underline{\mu} - \ell_t)\cdot\big(V'_{\emptyset}(W_{t-}) - 1\big) - r\cdot(V_{\emptyset}(W_{t-}) - W_{t-}) + \big[V_{\emptyset}(W_{t-}+H_t) - V_{\emptyset}(W_{t-}) - H_t\big]\cdot\underline{\mu}\\
=&\rho W_{t-}\cdot\big(V'_{\emptyset}(W_{t-}) - 1\big) - r\cdot(V_{\emptyset}(W_{t-}) - W_{t-}) - \ell_t\cdot\big(V'_{\emptyset}(W_{t-}) - 1\big)\\
&+\big[V_{\emptyset}(W_{t-}+H_t) - V_{\emptyset}(W_{t-}) - V'_{\emptyset}(W_{t-})H_t\big]\cdot\underline{\mu}\\
\leqslant&\rho W_{t-}\cdot\big(V'_{\emptyset}(W_{t-}) - 1\big) - r\cdot(V_{\emptyset}(W_{t-}) - W_{t-}) + \ell_t\\
=&-\Big[rV_{\emptyset}(W_{t-}) - \rho W_{t-}\cdot V'_{\emptyset}(W_{t-}) + (\rho - r)W_{t-} - R\underline{\mu}\Big] + \ell_t - R\underline{\mu}\\
=&-(\mathcal{A}_{\emptyset}V_{\emptyset})(W_{t-}) + \ell_t - R\underline{\mu}\\
\leqslant&\ell_t - R\underline{\mu},
\end{aligned}
$$

where the first inequality follows from (i) $V'_{\emptyset}(W_{t-}) \geqslant 0$ (this follows from the fact that $V_{\emptyset}$ is nondecreasing) and (ii) $H_t \geqslant -W_{t-}$, and $0 = \arg\max_{h\geqslant -w}\{V_{\emptyset}(w+h) - V_{\emptyset}(w) - V'_{\emptyset}(w)\cdot h\}$ due to the concavity of $V_{\emptyset}$, and the last inequality follows from (20).

Combining the above two cases yields

$$
\begin{aligned}
&(\rho W_{t-} + b\mathbb{1}_{\nu_t=\mu} - H_t\nu_t - \ell_t)\cdot\big(V'_{\mathcal{E}_{t-}}(W_{t-}) - 1\big) - r\phi(W_{t-},\mathcal{E}_{t-}) + [\phi(W_{t-}+H_t,\mathcal{E}_{t-}) - \phi(W_{t-},\mathcal{E}_{t-})]\nu_t\\
\leqslant&\ell_t - [R\nu_t - (c-b)\mathbb{1}_{\nu_t=\mu}]
\end{aligned}
\tag{73}
$$

for any $t > 0$.

Consider next $A_2$. We have

$$
\begin{aligned}
&\phi(W_{t-} + H_t\mathrm{d}N_t - H^q_t\mathrm{d}Q_t - \Delta L_t,\mathcal{E}_{t-}) - \phi(W_{t-} + H_t\mathrm{d}N_t - H^q_t\mathrm{d}Q_t,\mathcal{E}_{t-})\\
=&V_{\mathcal{E}_{t-}}(W_{t-} + H_t\mathrm{d}N_t - H^q_t\mathrm{d}Q_t - \Delta L_t) - V_{\mathcal{E}_{t-}}(W_{t-} - H^q_t\mathrm{d}Q_t + H_t\mathrm{d}N_t) + \Delta L_t\\
\leqslant&\Delta L_t,\ \forall t > 0,
\end{aligned}
\tag{74}
$$

where the inequality follows from the facts that $\Delta L_t \geqslant 0$ and that $V_{\varepsilon}$ is nondecreasing for any $\varepsilon \in \{\mathsf{I},\emptyset\}$.

Consider now $A_3$. By considering four possible value combinations of $(\mathcal{E}_{t-},\mathcal{E}_t)$ and using (21), we have

$$
\phi(W_t,\mathcal{E}_t) - \phi(W_t,\mathcal{E}_{t-}) = V_{\mathcal{E}_t}(W_t) - V_{\mathcal{E}_{t-}}(W_t) \leqslant \kappa(\mathcal{E}_{t-},\mathcal{E}_t). \tag{75}
$$

Consider next $A_4$. We have

$$
\begin{aligned}
&H^q_t\big(V'_{\mathcal{E}_{t-}}(W_{t-}) - 1\big) + \phi(W_{t-} - H^q_t,\mathcal{E}_{t-}) - \phi(W_{t-},\mathcal{E}_{t-})\\
=&H^q_tV'_{\mathcal{E}_{t-}}(W_{t-}) + V_{\mathcal{E}_{t-}}(W_{t-} - H^q_t) - V_{\mathcal{E}_{t-}}(W_{t-}) \leqslant 0,
\end{aligned}
$$

where the inequality follows from the concavity of $V_{\varepsilon}$ for any $\varepsilon \in \{\mathsf{I},\emptyset\}$. This, combining with $q_t \geqslant 0$, yields

$$
A_4 = \int_{0+}^{T} e^{-rt}q_t\Big\{H^q_t\big(V'_{\mathcal{E}_{t-}}(W_{t-}) - 1\big) + \phi(W_{t-} - H^q_t,\mathcal{E}_{t-}) - \phi(W_{t-},\mathcal{E}_{t-})\Big\}\mathrm{d}t \leqslant 0. \tag{76}
$$

Consider finally $A_5$. It follows from (2) and (4) with $t=0$ that $\mathbb{E}[W_0+\Delta L_0]=W_{0-}$. Therefore, we have

$$\begin{aligned}\mathbb{E}[\phi(W_0,\mathcal{E}_{0-})]-\phi(W_{0-},\mathcal{E}_{0-})=&\mathbb{E}[V_{\mathcal{E}_{0-}}(W_0)]-V_{\mathcal{E}_{0-}}(W_{0-})-\big(\mathbb{E}[W_{0-}]-W_{0-}\big)\\ \leqslant&V_{\mathcal{E}_{0-}}\big(\mathbb{E}[W_0]\big)-V_{\mathcal{E}_{0-}}(W_{0-})+\mathbb{E}[\Delta L_0]\leqslant\mathbb{E}[\Delta L_0],\end{aligned}\tag{77}$$

where the first inequality follows from the concavity of $V_\varepsilon$ for any $\varepsilon\in\{\mathsf{I},\emptyset\}$ and Jensen's inequality, the second inequality follows from the facts that $V_\varepsilon$ is nondecreasing and $W_{0-}=\mathbb{E}[W_0+L_0]\geqslant\mathbb{E}[W_0]$.

Combining (72)–(76), we have

$$\begin{aligned}e^{-rT}\phi(W_T,\mathcal{E}_T)\leqslant&\phi(W_{0-},\mathcal{E}_{0-})+\int_{0+}^T e^{-rt}[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^N\\ &+\int_{0+}^T e^{-rt}[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^Q\\ &+\int_{0+}^T e^{-rt}[\ell_t-(R\nu_t-(c-b)\mathbb{1}_{\nu_t=\mu})]\mathrm{d}t+\sum_{0<t\leqslant T}e^{-rt}\Delta L_t\\ &+\sum_{0\leqslant t\leqslant T}e^{-rt}\kappa(\mathcal{E}_{t-},\mathcal{E}_t)+\phi(W_0,\mathcal{E}_{0-})-\phi(W_{0-},\mathcal{E}_{0-})\end{aligned}$$

for any $T>0$, which can be rewritten as

$$\begin{aligned}\phi(W_{0-},\mathcal{E}_{0-})\geqslant&e^{-rT}\phi(W_T,\mathcal{E}_t)-\int_0^T e^{-rt}[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^N\\ &-\int_{0+}^T e^{-rt}[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^Q\\ &+\int_{0+}^T e^{-rt}(R\mathrm{d}N_t-\mathrm{d}L_t-(c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}}\mathrm{d}t)-\sum_{0\leqslant t\leqslant T}e^{-rt}\kappa(\mathcal{E}_{t-},\mathcal{E}_t)\\ &+\phi(W_{0-},\mathcal{E}_{0-})-\phi(W_0,\mathcal{E}_{0-}).\end{aligned}$$

Taking expectation in the above inequality yields

$$\begin{aligned}\phi(W_{0-},\mathcal{E}_{0-})\geqslant&\mathbb{E}[e^{-rT}\phi(W_T,\mathcal{E}_T)]-\mathbb{E}\left[\int_{0+}^T e^{-rt}[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^N\right]\\ &-\mathbb{E}\left[\int_{0+}^T e^{-rt}[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^Q\right]\\ &+\mathbb{E}\left[\int_{0+}^T e^{-rt}\big(R\mathrm{d}N_t-\mathrm{d}L_t-(c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}}\mathrm{d}t\big)-\sum_{0\leqslant t\leqslant T}e^{-rt}\kappa(\mathcal{E}_{t-},\mathcal{E}_t)\right]\\ &+\phi(W_{0-},\mathcal{E}_{0-})-\mathbb{E}\phi(W_0,\mathcal{E}_{0-})\\ \geqslant&\mathbb{E}[e^{-rT}\phi(W_T,\mathcal{E}_T)]-\mathbb{E}\left[\int_{0+}^T e^{-rt}[\phi(W_{t-}+H_t,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^N\right]\\ &-\mathbb{E}\left[\int_{0+}^T e^{-rt}[\phi(W_{t-}-H_t^q,\mathcal{E}_{t-})-\phi(W_{t-},\mathcal{E}_{t-})]\mathrm{d}M_t^Q\right]\\ &+\mathbb{E}\left[\int_0^T e^{-rt}(R\mathrm{d}N_t-\mathrm{d}L_t-(c-b)\mathbb{1}_{\mathcal{E}_t=\mathsf{I}}\mathrm{d}t)-\sum_{0\leqslant t\leqslant T}e^{-rt}\kappa(\mathcal{E}_{t-},\mathcal{E}_t)\right]\end{aligned}\tag{78}$$

for any $T>0$, where the last inequality follows from (77).

We claim that it suffices to consider the case that

$$\mathbb{E}\left[\int_{0+}^\infty e^{-rt}|H_t|\nu_t\mathrm{d}t\right]<\infty.\tag{79}$$

Otherwise, we have $\mathbb{E}\left[\int_{0+}^{\infty} e^{-rt}|H_t|\nu_t \mathrm{d}t\right] = \infty$. It follows from (PK) and (WU) that $\mathrm{d}L_t \geqslant (H_t - \bar{W})^+ \mathrm{d}N_t$ for $t > 0$. Hence, we have

$$\mathbb{E}\left[\int_0^{\infty} e^{-rt}\mathrm{d}L_t\right] \geqslant \mathbb{E}\left[\int_{0+}^{\infty} e^{-rt}(H_t - \bar{W})^+\mathrm{d}N_t\right] = \mathbb{E}\left[\int_{0+}^{\tau} e^{-rt}(H_t - \bar{W})^+\nu_t\mathrm{d}t\right]$$
$$\geqslant \mathbb{E}\left[\int_{0+}^{\infty} e^{-rt}(|H_t| - \bar{W})\nu_t\mathrm{d}t\right] \geqslant \mathbb{E}\left[\int_{0+}^{\infty} e^{-rt}|H_t|\nu_t\mathrm{d}t\right] - \frac{\bar{W}\mu}{r} = \infty,$$

where the equality follows from (2.3), Chapter II in Brémaud (1981), the second inequality follows from $H_t \geqslant -W_{t-} \geqslant -R\mu/r$ in view of (5) and (WU), and the third inequality follows from $\nu_t \leqslant \mu$. Then, we have

$$U(\Gamma, \mathcal{E}_{0-}) \leqslant \mathbb{E}^{\bar{\nu}(\Gamma)}\left[\int_0^{\infty} e^{-rt}\left[R\mathrm{d}N_t - \mathrm{d}L_t\right]\middle|\mathcal{E}_{0-}\right] \leqslant \frac{R\mu}{r} - \mathbb{E}^{\bar{\nu}(\Gamma)}\left[\int_0^{\infty} e^{-rt}\mathrm{d}L_t\middle|\mathcal{E}_{0-}\right] = -\infty,$$

and thus the desired result follows immediately.

Given (79), we have

$$\mathbb{E}\left[\int_{0+}^{\infty} e^{-rt}|\phi(W_{t-} + H_t, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-})|\nu_t\right]\mathrm{d}t$$
$$\leqslant \max_{w>0, \varepsilon\in\{\mathsf{I},\emptyset\}}\{|V_\varepsilon'(w) - 1|\} \cdot \mathbb{E}\left[\int_{0+}^{\infty} e^{-rt}|H_t|\nu_t\mathrm{d}t\right] < \infty,$$

where $\max_{w>0,\varepsilon\in\{\mathsf{I},\emptyset\}}\{|V_\varepsilon'(w) - 1|\} < \infty$ follows from the concavity of $V_\varepsilon$ and the fact that $V_\varepsilon' \geqslant 0$. It follows from Lemma L3, Chapter II in Brémaud (1981) that $\widetilde{M} = \{\widetilde{M}_t\}_{t\geqslant 0}$, defined by

$$\widetilde{M}_t = \int_{0+}^{t} e^{-rs}[\phi(W_{s-} + H_s, \mathcal{E}_{s-}) - \phi(W_{s-}, \mathcal{E}_{s-})]\mathrm{d}M_s^N,$$

is an $\mathcal{F}$-martingale. Hence, $\mathbb{E}\widetilde{M}_T = \mathbb{E}\widetilde{M}_0 = 0$, that is,

$$\mathbb{E}\left[\int_{0+}^{T} e^{-rt}[\phi(W_{t-} + H_t, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-})]\mathrm{d}M_t^N\right] = 0.$$

Similarly, using (1), we can show that

$$\mathbb{E}\left[\int_{0+}^{T} e^{-rt}[\phi(W_{t-} - H_t^q, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-})]\mathrm{d}M_t^Q\right] = 0.$$

It follows from (22) and the fact that both $V_\mathsf{I}$ and $V_\emptyset$ are nondecreasing that $\phi(w,\varepsilon) \geqslant \underline{v} - w$ for any $\varepsilon \in \{\mathsf{I}, \emptyset\}$. Letting $T \to \infty$ in (78) and using (WU), we have $\phi(W_{0-}, \mathcal{E}_{0-}) \geqslant U(\Gamma, \mathcal{E}_{0-})$ with $W_{0-} = u(\Gamma, \nu, \mathcal{E}_{0-})$. Hence, the desired result is obtained. ▫

A byproduct of the proof of Theorem 1 is the following result. In the remaining of this appendix, whenever we need to prove that certain contract achieves the upper bound, we use this result together with Lemma 1.

Proposition 9. *Suppose that the conditions stated in Theorem 1 hold. Furthermore, suppose that there exists a contract* $\Gamma^\diamond \in \mathfrak{C}$ *such that the corresponding agent's promised utility* $W_t$ *satisfies*

$$(\rho W_{t-} + b\mathbb{1}_{\nu_t=\mu} - H_t\nu_t - \ell_t)\big(V_{\mathcal{E}_{t-}}'(W_{t-}) - 1\big) - r\phi(W_{t-}, \mathcal{E}_{t-}) + [\phi(W_{t-} + H_t, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-})]\nu_t$$
$$= \ell_t - [R\nu_t - (c-b)\mathbb{1}_{\nu_t=\mu}], \tag{80}$$

$$\phi(W_{t-} + H_t\mathrm{d}N_t - H_t^q\mathrm{d}Q_t - \Delta L_t, \mathcal{E}_{t-}) - \phi(W_{t-} + H_t\mathrm{d}N_t - H_t^q\mathrm{d}Q_t, \mathcal{E}_{t-}) = \Delta L_t, \tag{81}$$

$$\phi(W_t, \mathcal{E}_t) - \phi(W_t, \mathcal{E}_{t-}) = \kappa(\mathcal{E}_{t-}, \mathcal{E}_t), \tag{82}$$

$$q_t\Big\{H_t^q\big(V_{\mathcal{E}_{t-}}'(W_{t-}) - 1\big) + \phi(W_{t-} - H_t^q, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-})\Big\} = 0, \tag{83}$$

*for any* $t > 0$ *and*

$$\mathbb{E}[\phi(W_0, \mathcal{E}_{0-})] - \phi(W_{0-}, \mathcal{E}_{0-}) = \mathbb{E}[\Delta L_0]. \tag{84}$$

*Then, for any value* $w \in [0, \infty)$ *and initial state* $\mathcal{E}_{0-} \in \{\mathsf{I}, \emptyset\}$*, such that* $u(\Gamma^{\diamond}, \bar{\nu}, \mathcal{E}_{0-}) = w$*, we have*

$$U(\Gamma^{\diamond}, \mathcal{E}_{0-}) = V_{\mathcal{E}_{0-}}(w) - w.$$

*Proof.* Equalities (80)–(84) demonstrate that all the inequalities in the proof of Theorem 1, (73)–(77), hold as equalities under contract $\Gamma^{\diamond}$. The desired result can be shown by going through the proof of Theorem 1, with all inequalities replaced by equalities. ▫

*Proof of Lemma 2.* This result follows almost the same logic as that for the proof of Lemmas 2 and 3 in Cao et al. (2021), which uses Lemma 6 below. However, there are minor differences as both $\beta$ and $\bar{w}$ in Cao et al. (2021) take different values from ours. We first present Lemma 6 here because it is frequently used in the subsequent analysis. Its proof can be found at the end of this section.

Lemma 6. *For any* $\widetilde{w} \in [0, \bar{w})$*, there exists a unique function* $V_{\widetilde{w}}$ *in* $C^1([0, \widetilde{w}])$ *that solves the differential equation* (24) *on* $[0, \widetilde{w}]$ *with boundary condition* (25)*. Further extend the domain of* $V_{\widetilde{w}}$ *to* $\mathbb{R}_+$ *by letting* $V_{\widetilde{w}}(w) = V_{\widetilde{w}}(\widetilde{w})$ *for all* $w > \widetilde{w}$*. Function* $V_{\widetilde{w}}(w)$ *is non-decreasing in* $w$*, and has the following properties.*

- (i) *For any* $\widetilde{w}_1$ *and* $\widetilde{w}_2$ *such that* $0 < \widetilde{w}_1 < \widetilde{w}_2 < \bar{w}$*, we have* $V_{\widetilde{w}_1}(w) > V_{\widetilde{w}_2}(w)$ *and* $V'_{\widetilde{w}_1}(w) < V'_{\widetilde{w}_2}(w)$ *for* $w \in [0, \widetilde{w}_1)$.
- (ii) $V_{\widetilde{w}}(\cdot) \in C^1(\mathbb{R}_+) \cap C^2(\mathbb{R}_+ \backslash \{\widetilde{w}\})$.
- (iii) *For any given* $w \geqslant 0$*, define function* $v(\widetilde{w}) := V_{\widetilde{w}}(w)$*. We have* $v(\cdot) \in C^1([0, \bar{w}))$.
- (iv) *If* $\rho \leqslant r + \mu$*, then for any* $w \geqslant 0$*,* $V_{\widetilde{w}}(w)$ *approaches negative infinity, and* $V'_{\widetilde{w}}(w)$ *approaches positive infinity, as* $\widetilde{w}$ *approaches* $\bar{w}$ *from below.*
- (v) *If* $\rho > r + \mu$*, then for any* $w \in [0, \bar{w}]$*,*

$$\lim_{\widetilde{w} \uparrow \bar{w}} V_{\widetilde{w}}(w) = \bar{V} - \frac{\rho - r}{\rho - r - \mu}(\bar{w} - w),$$

*where* $\bar{V}$ *is defined in* (9)*. Furthermore,* $\bar{V} - \dfrac{\rho - r}{\rho - r - \mu}\bar{w} > \underline{v}$ *is equivalent to*

$$R > \left( \frac{c}{b} + \frac{(\rho - r)(\rho - \mu)\mu}{(\mu - \underline{\mu})(\rho - r - \mu)\rho} \right) \beta.$$

Below, for the sake of brevity, we only highlight the differences from Cao et al. (2021). Similar to Appendix B.4 in Cao et al. (2021), we start by considering the case that $\mu + r \leqslant \rho$. Here, $\bar{w} < \beta$ is implied by $\rho > \mu$. The following argument in Appendix B.4 in Cao et al. (2021) remains intact, until that their inequality (B.22), which should be modified to

$$rV_{\widetilde{w}}(w^c) > \mu R - c - (\rho - r)(w^c - \beta) + [\rho(w^c - \beta) + r\beta] V'_{\widetilde{w}}(w^c).$$

As a result, the two cases to be considered will be $\rho(w^c - \beta) + r\beta \geqslant 0$ versus $\rho(w^c - \beta) + r\beta < 0$. The former case reaches a contradiction, following the same logic as in Cao et al. (2021), and the latter case implies

$w^c < (1-r/\rho)\beta$, demonstrating a possible value $\breve{w}(\widetilde{w}) \in [0, \widetilde{w})$ (which is $w^c$) such that $V''_{\widetilde{w}} < 0$ over $(\breve{w}(\widetilde{w}), \widetilde{w})$ and $V''_{\widetilde{w}} > 0$ over $[0, \breve{w}(\widetilde{w}))$. Hence, part (i) follows immediately. Besides, it is evident that $\mathcal{V}''_{\widetilde{w}}(\breve{w}(\widetilde{w})) = 0$ if $\breve{w}(\widetilde{w}) > 0$.

Both parts (ii) and (iii) follow immediately from Lemma 3 in Cao et al. (2021), by noting that all remaining arguments in Appendix B.4 of Cao et al. (2021) are valid. We mention that although both Lemmas 2 and 3 in Cao et al. (2021) are stated under Condition 2, this condition is not needed in showing properties in our Lemma 2. In fact, this condition is only required for guaranteeing that $\mathcal{V}_{\widehat{w}}(0) = \underline{v}$ is satisfied for some $\widehat{w}$ in $[0, \bar{w})$; see our Lemma 3. □

*Proof of Lemma 6.* The existence and uniqueness of a function satisfying the claimed property has been shown as the first step in the proof of Proposition 4 in Sun and Tian (2018), which is omitted for brevity. Next, we show that such a function $V_{\widetilde{w}}$ has properties (i)–(v).

(i) This property has been proved in Step 2 from the proof of Proposition 4 in Sun and Tian (2018).

(ii) It follows from (24) and the boundary condition at $\widetilde{w}$ that $V_{\widetilde{w}}(\cdot) \in C^1(\mathbb{R}_+)$. Taking derivative in (24) with respect to $w$ yields

$$(\mu + r)V'_{\widetilde{w}}(w) - \mu V'_{\widetilde{w}}\big((w+\beta) \wedge \widetilde{w}\big) + \rho(\bar{w} - w)V''_{\widetilde{w}}(w) - \rho V'_{\widetilde{w}}(w) + \rho - r = 0 \tag{85}$$

for $w \in [0, \widetilde{w})$, which implies that $V_{\widetilde{w}}(\cdot) \in C^2([0, \widetilde{w}))$. Moreover, $V''_{\widetilde{w}}(\widetilde{w}-) = -(\rho - r)/[\rho(\bar{w} - \widetilde{w})] < 0$. Besides, by the definition of $V_{\widetilde{w}}$ on $(\widetilde{w}, \infty)$, we have $V''_{\widetilde{w}}(w) = 0$ for $w > \widetilde{w}$ and $V''_{\widetilde{w}}(\widetilde{w}+) = 0$. Hence, $V_{\widetilde{w}}(\cdot) \in C^2(\mathbb{R}_+ \backslash \{\widetilde{w}\})$.

(iii) Fix any $w \geqslant 0$. If $\widetilde{w} \leqslant w$, then $v(\widetilde{w}) = V_{\widetilde{w}}(w) = \bar{V}(\widetilde{w})$, which implies that $v(\cdot) \in C^1([0, w \wedge \bar{w}))$. Hence, the desired property is obtained if $w \geqslant \bar{w}$.

Now suppose that $w < \bar{w}$ and $\widetilde{w} \in (w, \bar{w})$. By the above discussion, we have $v(\cdot) \in C^1([0, w])$. For any $w' \in [w, \widetilde{w}]$, it follows from (24) that

$$\rho V'_{\widetilde{w}}(w') = -\frac{(\mu + r)V_{\widetilde{w}}(w')}{\bar{w} - w'} + \frac{\mu V_{\widetilde{w}}\big((w'+\beta) \wedge \widetilde{w}\big)}{\bar{w} - w'} + \frac{(\mu R - c) - (\rho - r)w'}{\bar{w} - w'}.$$

Integrating the above equation with respect to $w'$ from $w$ to $\widetilde{w}$ yields

$$\begin{aligned}\rho\big(V_{\widetilde{w}}(\widetilde{w}) - V_{\widetilde{w}}(w)\big) = -(\mu + r)\int_w^{\widetilde{w}} \frac{V_{\widetilde{w}}(w')}{\bar{w} - w'}\mathrm{d}w' + \mu \int_w^{\widetilde{w}} \frac{V_{\widetilde{w}}\big((w'+\beta) \wedge \widetilde{w}\big)}{\bar{w} - w'}\mathrm{d}w' \\ + \int_w^{\widetilde{w}} \frac{(\mu R - c) - (\rho - r)w'}{\bar{w} - w'}\mathrm{d}w'.\end{aligned}$$

First, using the above equality, we can obtain that $v(\widetilde{w}) = V_{\widetilde{w}}(w)$ is continuous in $\widetilde{w}$ on $[w, \bar{w})$. Then, again using this equality, we conclude that $V_{\widetilde{w}}(w)$ is continuously differentiable in $\widetilde{w}$ on $[w, \bar{w})$, which, combining with $v(\cdot) \in C^1([0, w])$, yields that $v(\cdot) \in C^1([0, \bar{w})$.

(iv) This also has been shown in Step 2 from the proof of Proposition 4 in Sun and Tian (2018).

(v) According to the proof of Lemmas 2 and 3 in Cao et al. (2021), as $\widetilde{w} \uparrow \bar{w}$, we have $b_{\widetilde{w}} \downarrow 0$ and thus

$$V_{\widetilde{w}}(w) \to \bar{U} - \frac{\rho - r}{\rho - r - \mu}(\bar{w} - w)$$

for any $w \in [0, \bar{w})$.

At the end, we show that $V_{\widetilde{w}}(w)$ is non-decreasing in $w$ on $\mathbb{R}_+$ by a contradictory argument. Suppose that $w^p = \sup\{w \in \mathbb{R}_+ : V'_{\widetilde{w}}(w) < 0\}$ exists. Recall from (ii) that $V''_{\widetilde{w}}(\widetilde{w}-) < 0$. Hence, $w^p \in [0, \widetilde{w})$, $V'_{\widetilde{w}}(w^p) = 0$ and $V'_{\widetilde{w}} > 0$ on $(w^p, \widetilde{w})$. It follows from (24) at $w^p$ that

$$
\begin{aligned}
rV_{\widetilde{w}}(w^p) &= \mu R - c - (\rho - r)w^p + \mu\Big[V_{\widetilde{w}}\big((w^p + \beta) \wedge \widetilde{w}\big) - V_{\widetilde{w}}(w^p)\Big] \\
&> \mu R - c - (\rho - r)w^p = rV_{\widetilde{w}}(\widetilde{w}),
\end{aligned}
$$

where the inequality follows from that $V'_{\widetilde{w}} > 0$ on $(w^p, \widetilde{w})$. This reaches a contradiction with $V_{\widetilde{w}}(w^p) < V_{\widetilde{w}}(\widetilde{w})$. □

## B.3. Proofs of the Results in Section 4

*Proof of Lemma 3.* This follows from Lemma 3 in Cao et al. (2021). □

### B.3.1. Proofs of the Results in Section 4.1

*Proof of Proposition 2.* The proof consists of two parts.

*Part 1.* First, we verify (26) and (27). Following the definition of $\underline{\Gamma}$ as in (16), (27) trivially holds. Following Lemma 1, in order to show (26), we apply Proposition 9 by verifying that (80)–(84) all hold.

Fix any $t \geqslant 0$. By comparing (PK) with the dynamics of $W$ and $L$ in Definition 1 (i) and (ii), we have $\nu_t = \mu$, $\ell_t = b$ and $H_t = \beta$ if $\mathcal{E}_{t-} = \mathsf{I}$, and $\nu_t = \underline{\mu}$, $\ell_t = 0$ and $H_t = 0$ if $\mathcal{E}_{t-} = \emptyset$. This implies that (80) holds.

For (81), we consider the following three cases: (i) if $\mathcal{E}_{t-} = \emptyset$, then $I_t = 0$ and (81) trivially holds; (ii) if $\mathcal{E}_{t-} = \mathsf{I}$ and $I_t = 0$, then (81) trivially holds; (iii) if $\mathcal{E}_{t-} = \mathsf{I}$ and $I_t > 0$, then by Definition 1 (ii), $\mathrm{d}N_t = 1$ and $I_t = W_{t-} + \beta - \widehat{\mathrm{w}}$, which implies $W_{t-} + \beta > \widehat{\mathrm{w}}$. Hence, we have

$$
\begin{aligned}
&\phi(W_{t-} + H_t \mathrm{d}N_t - I_t, \mathsf{I}) - \phi(W_{t-} + H_t \mathrm{d}N_t, \mathsf{I}) \\
=&\phi(\widehat{\mathrm{w}}, \mathsf{I}) - \phi(W_{t-} + \beta, \mathsf{I}) \\
=&\mathcal{V}_{\widehat{\mathrm{w}}}(\widehat{\mathrm{w}}) - \widehat{\mathrm{w}} - [\mathcal{V}_{\widehat{\mathrm{w}}}((W_{t-} + \beta) \wedge \widehat{\mathrm{w}}) - (W_{t-} + \beta)] \\
=&W_{t-} + \beta - \widehat{\mathrm{w}} = I_t,
\end{aligned}
$$

and thus (81) also holds.

For (82), we also consider three cases: (i) if $W_t > 0$, then (82) holds since $\mathcal{E}_t = \mathcal{E}_{t-} = \mathsf{I}$; (ii) if $W_t = 0$ and $\mathcal{E}_{t-} = \mathsf{I}$, then (82) holds since $\mathcal{E}_t = \emptyset$ and $\kappa(\mathsf{I}, \emptyset) = 0$; (iii) if $W_t = 0$ and $\mathcal{E}_{t-} = \emptyset$, then (82) holds since $\mathcal{E}_t = \emptyset$.

For (83), note that $q_t > 0$ only if $\breve{w}(\widehat{\mathrm{w}}) > 0$, $W_{t-} = \breve{w}(\widehat{\mathrm{w}})$ and $\mathcal{E}_{t-} = \mathsf{I}$. Hence, we have

$$
\begin{aligned}
&H_t^q\big(V'_{\mathcal{E}_{t-}}(W_{t-}) - 1\big) + \phi(W_{t-} - H_t^q, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-}) \\
=&\breve{w}(\widehat{\mathrm{w}})\mathcal{V}'_{\widehat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) + \mathcal{V}_{\widehat{\mathrm{w}}}(0) - \mathcal{V}_{\widehat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) = 0,
\end{aligned}
$$

where the last equality follows from property (ii) in Lemma 2. Hence, (83) holds. Finally, (84) holds since $W_0 = \widehat{\mathrm{w}} \wedge W_{0-}$.

*Part 2.* Next, we show that functions $V_{\mathsf{I}}(w) = \mathcal{V}_{\widehat{\mathrm{w}}}(w)$ and $V_{\emptyset}(w) = \underline{v}$ satisfy the optimality condition (20)–(22).

First, we show that $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) \geqslant 0$ for any $w \in \mathbb{R}_+$. If $w \in [\breve{w}(\widehat{\mathrm{w}}), \widehat{\mathrm{w}})$, by the definition of $\mathcal{V}_{\hat{\mathrm{w}}}$, we have $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) = 0$. If $w \in [\widehat{\mathrm{w}}, \infty)$, then we have

$$\begin{aligned}(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) &= (\mu + r)\mathcal{V}_{\hat{\mathrm{w}}}(\widehat{\mathrm{w}}) - \mu\mathcal{V}_{\hat{\mathrm{w}}}(\widehat{\mathrm{w}}) - (\mu R - c) + (\rho - r)w \\ &= (\rho - r)(w - \widehat{\mathrm{w}}) \geqslant 0.\end{aligned}$$

If $w \in [0, \breve{w}(\widehat{\mathrm{w}}))$ (if we discuss this case, it is implicitly assumed that $\breve{w}(\widehat{\mathrm{w}}) > 0$), then we have $\mathcal{V}_{\hat{\mathrm{w}}}(w) = \underline{v} + \mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}}))w$. Consequently,

$$\begin{aligned}(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) &= (\mu + r)(\underline{v} + \mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}}))w) - \mu\mathcal{V}_{\hat{\mathrm{w}}}(w + \beta) + \rho(\bar{w} - w)\mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) - (\mu R - c) + (\rho - r)w \\ &=: g_{\mathsf{I}}(w).\end{aligned}$$

Obviously, $g_{\mathsf{I}}(\breve{w}(\widehat{\mathrm{w}})) = 0$. Moreover, for $w \in [0, \breve{w}(\widehat{\mathrm{w}}))$, we have

$$\begin{aligned}g'_{\mathsf{I}}(w) &= (\mu + r)\mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) - \mu\mathcal{V}'_{\hat{\mathrm{w}}}(w + \beta) - \rho\mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) + \rho - r \\ &= (\rho - r)(1 - \mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}}))) + \mu(\mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) - \mathcal{V}'_{\hat{\mathrm{w}}}(w + \beta)) \\ &\leqslant (\rho - r)\big(1 - \mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}}))\big) + \mu\big(\mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) - \mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}}) + \beta)\big) = 0,\end{aligned}$$

where the inequality follows from the concavity of $\mathcal{V}_{\hat{\mathrm{w}}}$, and the last equality follows from that $\rho(\bar{w} - \breve{w}(\widehat{\mathrm{w}}))\mathcal{V}''_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) = (\rho - r)(\mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) - 1) + \mu[\mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}}) + \beta) - \mathcal{V}'_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}}))]$ (due to (85)) and $\mathcal{V}''_{\hat{\mathrm{w}}}(\breve{w}(\widehat{\mathrm{w}})) = 0$. Consequently, $g_{\mathsf{I}}(w) \geqslant 0$ for all $w \in [0, \breve{w}(\widehat{\mathrm{w}}))$.

Therefore, we have $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) \geqslant 0$ for any $w \in \mathbb{R}_+$. Obviously, we have $(\mathcal{A}_{\emptyset}V_{\emptyset})(w) = (\rho - r)w \geqslant 0$. Hence, (20) holds.

It follows from the facts that $V_{\mathsf{I}}(w) \geqslant V_{\mathsf{I}}(0) = \underline{v} = V_{\emptyset}(w)$ and $V_{\emptyset}(w) = \underline{v} \geqslant \bar{V}(\widehat{\mathrm{w}}) - K = \mathcal{V}_{\hat{\mathrm{w}}}(\widehat{\mathrm{w}}) - K \geqslant V_{\mathsf{I}}(w) - K$ (due to (H1)) that both (21) and (22) hold. □

### B.3.2. Proof of the Results in Section 4.2

*Proofs of Lemma 4.* Define

$$g(w, K) := \mathcal{V}_{\hat{\mathrm{w}}}(w) - \mathcal{V}'_{\hat{\mathrm{w}}}(w)w - \underline{v} - K. \tag{86}$$

Then, we have

$$g(\breve{w}(\widehat{\mathrm{w}}), K) = -K < 0, \tag{87}$$

where the equality follows from that $\mathcal{V}_{\hat{\mathrm{w}}}(w)$ is linear in $w$ on $[0, \breve{w}(\widehat{\mathrm{w}})]$, and

$$g(\widehat{\mathrm{w}}, K) = \mathcal{V}_{\hat{\mathrm{w}}}(\widehat{\mathrm{w}}) - \underline{v} - K > 0,$$

where the equality follows from $\mathcal{V}'_{\hat{\mathrm{w}}}(\widehat{\mathrm{w}}) = 0$ and the inequality follows from the assumption that (H1) does not hold. Furthermore, we have

$$\frac{\partial g(w, K)}{\partial w} = -\mathcal{V}''_{\hat{\mathrm{w}}}(w)w > 0 \ \text{ for } \ w \in (\breve{w}(\widehat{\mathrm{w}}), \widehat{\mathrm{w}}),$$

where the inequality follows from that $\mathcal{V}_{\hat{\mathrm{w}}}$ is strictly concave on $(\breve{w}(\hat{\mathrm{w}}), \hat{\mathrm{w}})$. Since $g(w,K)$ is continuous in $w$ (recall that $\mathcal{V}_{\hat{\mathrm{w}}}(\cdot)$ is continuously differentiable), for any $K>0$, there exists a unique $\bar{\theta}^K \in (\breve{w}(\hat{\mathrm{w}}), \hat{\mathrm{w}})$ such that $g(\bar{\theta}^K, K) = 0$. Hence, (28) holds if we define $m^K := \mathcal{V}'_{\hat{\mathrm{w}}}(\bar{\theta}^K)$. Furthermore, by the Implicit Function Theorem, we have

$$\frac{\mathrm{d}\bar{\theta}^K}{\mathrm{d}K} = -\frac{\frac{\partial g(w,K)}{\partial K}}{\frac{\partial g(w,K)}{\partial w}} = \frac{1}{\frac{\partial g(w,K)}{\partial w}} > 0,$$

which implies that $\bar{\theta}^K$ is increasing in $K$. Since $\mathcal{V}'_{\hat{\mathrm{w}}}(w)$ is decreasing in $w$, we have $m^K = \mathcal{V}'_{\hat{\mathrm{w}}}(\bar{\theta}^K)$ is decreasing in $K$. Finally, $\lim_{K\downarrow 0} \bar{\theta}^K = \breve{w}(\hat{\mathrm{w}})$ is implied by (87). □

*Proof of Proposition 3.* First, (26) and (27) hold as the corresponding argument in the proof of Proposition 2 is still valid. Next, we verify that functions $V_{\mathsf{I}}(w) = \mathcal{V}_{\hat{\mathrm{w}}}(w)$ and $V_{\emptyset}(w)$ as defined in (29) satisfy the optimality condition (20)–(22). Note that Lemma 4 implies that $m^K \leqslant 1$ under condition (M1).

Obviously, (22) holds since $V_{\mathsf{I}}(0) = V_{\emptyset}(0) = \underline{v}$. Note that it has been shown in the proof of Proposition 2 that $(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) \geqslant 0$ for any $w \in \mathbb{R}_+$. Below, we will establish the second part of (20), as well as (21), by considering the following three cases.

*Case 1:* $w \in [0, \bar{\theta}^K)$. In this case, we have $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) = (\rho - r)(1-m^K)w \geqslant 0$. Besides, we have

$$\begin{aligned} V_{\emptyset}(w) &= \left(1 - \frac{w}{\bar{\theta}^K}\right)\underline{v} + \frac{w}{\bar{\theta}^K}\cdot(\mathcal{V}_{\hat{\mathrm{w}}}(\bar{\theta}^K) - K) \\ &\leqslant \left(1 - \frac{w}{\bar{\theta}^K}\right)\mathcal{V}_{\hat{\mathrm{w}}}(0) + \frac{w}{\bar{\theta}^K}\cdot\mathcal{V}_{\hat{\mathrm{w}}}(\bar{\theta}^K) \leqslant \mathcal{V}_{\hat{\mathrm{w}}}(w) = V_{\mathsf{I}}(w), \end{aligned}$$

where the second inequality follows from the concavity of $\mathcal{V}_{\hat{\mathrm{w}}}$. Moreover, we have

$$\begin{aligned} V_{\emptyset}(w) - V_{\mathsf{I}}(w) + K &= m^K w + \underline{v} + K - \mathcal{V}_{\hat{\mathrm{w}}}(w) \\ &= \mathcal{V}_{\hat{\mathrm{w}}}(\bar{\theta}^K) - m^K\cdot(\bar{\theta}^K - w) - \mathcal{V}_{\hat{\mathrm{w}}}(w) \\ &= \int_w^{\bar{\theta}^K} (\mathcal{V}'_{\hat{\mathrm{w}}}(y) - m^K)\mathrm{d}y > 0, \end{aligned}$$

where the second equality follows from the first equality in (28), and the inequality follows from the concavity of $\mathcal{V}_{\hat{\mathrm{w}}}$ and $m^K = \mathcal{V}'_{\hat{\mathrm{w}}}(\bar{\theta}^K)$.

*Case 2:* $w \in [\bar{\theta}^K, \hat{\mathrm{w}})$. First, we show that $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) \geqslant 0$ in this case. Note that $V_{\emptyset}(w) = \mathcal{V}_{\hat{\mathrm{w}}}(w) - K$ on $[\bar{\theta}^K, \hat{\mathrm{w}})$. Hence, $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) \geqslant 0$ is equivalent to

$$f_1(w) := r(\mathcal{V}_{\hat{\mathrm{w}}}(w) - K) - \rho w \mathcal{V}'_{\hat{\mathrm{w}}}(w) + (\rho - r)w - R\underline{\mu} \geqslant 0.$$

For $w \in [\breve{w}(\hat{\mathrm{w}}), \hat{\mathrm{w}}]$, it holds that $(\mathcal{A}_{\mathsf{I}}\mathcal{V}_{\hat{\mathrm{w}}})(w) = 0$. That is,

$$f_2(w) := (\mu + r)\mathcal{V}_{\hat{\mathrm{w}}}(w) - \mu\mathcal{V}_{\hat{\mathrm{w}}}(w+\beta) + \rho(\bar{w} - w)\mathcal{V}'_{\hat{\mathrm{w}}}(w) - (\mu R - c) + (\rho - r)w = 0. \tag{88}$$

Recall that $\bar{\theta}^K \in (\breve{w}(\hat{\mathrm{w}}), \hat{\mathrm{w}})$. Hence, it suffices to show that

$$\begin{aligned} f_3(w) &:= f_2(w) - f_1(w) \\ &= \mu\big(\mathcal{V}_{\hat{\mathrm{w}}}(w) - \mathcal{V}_{\hat{\mathrm{w}}}(w+\beta)\big) + \rho\bar{w}\mathcal{V}'_{\hat{\mathrm{w}}}(w) + rK - (R\Delta\mu - c) < 0 \end{aligned}$$

for $w \in [\bar{\theta}^K, \hat{\mathrm{w}})$.

It follows from (28) that $f_1(\bar{\theta}^K) = (\rho - r)\bar{\theta}^K(1 - m^K) > 0$. Hence, $f_3(\bar{\theta}^K) < 0$. Hence, it is enough to show that $f_3'(w) \leqslant 0$ for $w \in [\bar{\theta}^K, \widehat{\mathrm{w}})$, i.e.,

$$\mu(\mathcal{V}_{\widehat{\mathrm{w}}}'(w) - \mathcal{V}_{\widehat{\mathrm{w}}}'(w+\beta)) + \rho\bar{w}\mathcal{V}_{\widehat{\mathrm{w}}}''(w) \leqslant 0.$$

Taking derivative with respect to $w$ in (88) yields

$$(\mu + r)\mathcal{V}_{\widehat{\mathrm{w}}}'(w) - \mu\mathcal{V}_{\widehat{\mathrm{w}}}'(w+\beta) + \rho(\bar{w} - w)\mathcal{V}_{\widehat{\mathrm{w}}}''(w) - \rho\mathcal{V}_{\widehat{\mathrm{w}}}'(w) + \rho - r = 0$$

for $w \in [\breve{w}(\widehat{\mathrm{w}}), \widehat{\mathrm{w}}]$. Hence, for $w \in [\bar{\theta}^K, \widehat{\mathrm{w}})$, we have

$$\mu(\mathcal{V}_{\widehat{\mathrm{w}}}'(w) - \mathcal{V}_{\widehat{\mathrm{w}}}'(w+\beta)) + \rho\bar{w}\mathcal{V}_{\widehat{\mathrm{w}}}''(w) = (\rho - r)(\mathcal{V}_{\widehat{\mathrm{w}}}'(w) - 1) + \rho w\mathcal{V}_{\widehat{\mathrm{w}}}''(w) \leqslant 0,$$

where the inequality follows from the fact that $\mathcal{V}_{\widehat{\mathrm{w}}}'(\bar{\theta}^K) = m^K \leqslant 1$ and the concavity of $\mathcal{V}_{\widehat{\mathrm{w}}}$. Hence, we have $(\mathcal{A}_{\emptyset}V_{\emptyset})(w) \geqslant 0$ for $w \in [\bar{\theta}^K, \widehat{\mathrm{w}})$. Note that $V_{\emptyset}(w) - V_{\mathsf{I}}(w) + K = 0$. Hence, (21) trivially holds.

*Case 3:* $w \in [\widehat{\mathrm{w}}, \infty)$. It is straightforward to see that $(\mathcal{A}_{\emptyset}V_{\emptyset})(w) = r(\mathcal{V}_{\widehat{\mathrm{w}}}(\widehat{\mathrm{w}}) - K) + (\rho - r)w - R\underline{\mu} > r\underline{v} - R\underline{\mu} = 0$ and $V_{\emptyset}(w) - V_{\mathsf{I}}(w) + K = 0$. □

### B.3.3. Proofs of the Results in Section 4.3

*Proof of Proposition 4.* The proof of Proposition 4 is rather intricate, which has a total of four key steps. These steps illustrate how to identify thresholds $\bar{\vartheta}$ and $\underline{\vartheta}$ in computation. Furthermore, these steps help us establish $\underline{\theta}^0$ in Proposition 8.

In Step 1, fixing any $\underline{\theta}$, we identify bound $\hat{\mathbf{w}}$ and slope $\mathsf{c}$ as functions of $\underline{\theta}$ to satisfy (30) and (31).

Lemma 7. *For any $\underline{\theta} \in (0, \bar{w})$, there exist unique values $\widetilde{w}(\underline{\theta}) \in (\underline{\theta}, \bar{w})$ and $C(\underline{\theta})$, in place of $\hat{\mathbf{w}}$ and $\mathsf{c}$, respectively, such that value-matching and smooth-pasting conditions (30) and (31) are satisfied at $\underline{\vartheta} = \underline{\theta}$.*

Step 2 determines an interval to further identify $\underline{\theta}$.

Lemma 8. *Value $\underline{\theta}^0 := \inf\{\underline{\theta} \in (0, \bar{w}) : \widetilde{w}'(\underline{\theta}) \geqslant 0\}$ is well-defined. Furthermore, we have $\widetilde{w}(\underline{\theta})$ is strictly decreasing, and $C(\underline{\theta})$ strictly increasing, for $\underline{\theta} \in (0, \underline{\theta}^0)$, with $\widetilde{w}'(\underline{\theta}^0) = 0$. Moreover, $C(\underline{\theta}) > 0$ for any $\underline{\theta} \in (0, \underline{\theta}^0)$.*

Next, in Step 3, we define the upper threshold $\bar{\theta}$ as a function of $\underline{\theta}$, such that smooth pasting condition (33) is satisfied.

Lemma 9. *We have*

(i) *for any $\underline{\theta} \in (0, \underline{\theta}^0)$, the threshold*

$$\bar{\theta}(\underline{\theta}) := \inf\left\{w > \underline{\theta} : \mathcal{V}_{\widetilde{w}(\underline{\theta})}'(w) \leqslant 1 + C(\underline{\theta})r/\rho \cdot w^{r/\rho - 1}\right\} \tag{89}$$

*is well defined;*

(ii) *as a function of $\underline{\theta}$, threshold $\bar{\theta}(\underline{\theta})$ is decreasing and continuous in $\underline{\theta}$ on $[0, \underline{\theta}^0)$;*

(iii) $\lim_{\underline{\theta} \uparrow \underline{\theta}^0} \bar{\theta}(\underline{\theta}) = \underline{\theta}^0$.

Finally, in Step 4, we find the appropriate $\underline{\vartheta}$ to satisfy (32), and define $(\mathsf{c}, \hat{\mathbf{w}}, \bar{\vartheta})$ as $(C(\underline{\vartheta}), \widetilde{w}(\underline{\vartheta}), \bar{\theta}(\underline{\vartheta}))$. To this end, we define function

$$\psi(\underline{\theta}) := \mathcal{V}_{\widetilde{w}(\underline{\theta})}(\bar{\theta}(\underline{\theta})) - \left[\underline{v} + \bar{\theta}(\underline{\theta}) + C(\underline{\theta})\left(\bar{\theta}(\underline{\theta})\right)^{r/\rho}\right]. \tag{90}$$

In order to satisfy (32), we hope to identify the value $\underline{\vartheta}$ such that $\psi(\underline{\vartheta}) = K$, which is guaranteed in the following result.

LEMMA 10. *Function $\psi(\underline{\theta})$ is continuous and decreasing in $\underline{\theta}$ on $(0, \underline{\theta}^0)$, and satisfies*

$$\lim_{\underline{\theta}\uparrow\underline{\theta}^0} \psi(\underline{\theta}) = 0, \text{ and } \lim_{\underline{\theta}\downarrow 0} \psi(\underline{\theta}) > K.$$

*Consequently, there exists a value $\underline{\vartheta} \in (0, \underline{\theta}^0)$ such that $\psi(\underline{\vartheta}) = K$.*

The proofs of Lemmas 7–10 appear at the end of this subsection. According to these results, $(\hat{\mathbf{w}}, \mathsf{c}, \underline{\vartheta}, \bar{\vartheta})$ defined by $\hat{\mathbf{w}} = \widetilde{w}(\underline{\vartheta})$, $\mathsf{c} = C(\underline{\vartheta})$ and $\bar{\vartheta} = \bar{\theta}(\underline{\vartheta})$ satisfies (30)–(33). Besides, it follows from $\bar{\theta}(\underline{\theta}) < \widetilde{w}(\underline{\theta})$ for $\underline{\theta} \in [0, \underline{\theta}^0)$ that $\bar{\vartheta} < \hat{\mathbf{w}}$, which implies $\hat{\mathbf{w}} = \widetilde{w}(\underline{\vartheta}) < \widetilde{w}(0) = \widehat{\mathrm{w}}$ by noting that $\widetilde{w}(\underline{\theta})$ is decreasing in $\underline{\theta}$ on $[0, \underline{\theta}^0)$. To end the proof, we need to show that $\bar{\vartheta} > \breve{w}(\hat{\mathbf{w}})$. If it fails to hold, then we have $\mathcal{V}'_{\hat{\mathbf{w}}}(\underline{\vartheta}) = \mathcal{V}'_{\hat{\mathbf{w}}}(\bar{\vartheta}) = \mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}}))$. On the other side, it follows from $\mathsf{c} > 0$ and $\underline{\vartheta} < \bar{\vartheta}$ that $\mathsf{V}'_{\mathsf{c}}(\underline{\vartheta}) > \mathsf{V}'_{\mathsf{c}}(\bar{\vartheta})$. This contradicts with (31) and (33). □

*Proof of Proposition 5.* The proof consists of two parts.

*Part 1.* We verify (37) and (38). Similar to the proof of Proposition 2, we apply Lemma 1 and Proposition 9 by verifying that (80)–(84) all hold.

First, we show (37) by considering the case that $\mathcal{E}_{0-} = \mathsf{I}$. Equality (80) holds by noting that (i) $\ell_t = b\mathbb{1}_{\nu_t=\mu}$; (ii) for any $t > 0$, $W_{t-} \in [\underline{\vartheta}, \hat{\mathbf{w}}]$ if $\mathcal{E}_{t-} = 1$ and $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) = 0$ if $w \in [\underline{\vartheta}, \hat{\mathbf{w}}]$; (iii) for any $t > 0$, $W_{t-} \in (0, \bar{\vartheta}]$ if $\mathcal{E}_{t-} = 0$ and $(\mathcal{A}_{\emptyset}V_{\emptyset})(w) = 0$ if $w \in (0, \bar{\vartheta}]$.

Equality (81) holds by noting that $\Delta L_t > 0$ only if $\mathcal{E}_{t-} = 1$ and $W_t + H_t \mathrm{d}N_t - H_t^q \mathrm{d}Q_t > \hat{\mathbf{w}}$, as well as that $V_{\mathsf{I}}(w) = V_{\mathsf{I}}(\hat{\mathbf{w}})$ for any $w \geqslant \hat{\mathbf{w}}$.

Equality (82) holds by noting that for any $t \geqslant 0$, (i) $\mathcal{E}_t = 1 - \mathcal{E}_{t-} = 1$ only if $W_{t-} \in [\bar{\vartheta}, \hat{\mathbf{w}}]$ and $V_{\mathsf{I}}(w) - V_{\emptyset}(w) = K$ if $w \in [\bar{\vartheta}, \hat{\mathbf{w}}]$; and (ii) $\mathcal{E}_t = 1 - \mathcal{E}_{t-} = 0$ only if $W_{t-} \in (0, \underline{\vartheta}]$ and $V_{\emptyset}(w) - V_{\mathsf{I}}(w) = 0$ if $w \in (0, \underline{\vartheta}]$.

Note that $q_t > 0$ only if $\breve{w}(\hat{\mathbf{w}}) > 0$, $W_{t-} = \breve{w}(\hat{\mathbf{w}})$ and $\mathcal{E}_{t-} = \mathsf{I}$. Hence, if $q_t > 0$, then we have

$$\begin{aligned} &H_t^q \cdot \left(V'_{\mathcal{E}_{t-}}(W_{t-}) - 1\right) + \phi(W_{t-} - H_t^q, \mathcal{E}_{t-}) - \phi(W_{t-}, \mathcal{E}_{t-}) \\ =&(\breve{w}(\hat{\mathbf{w}}) - \underline{\vartheta})\mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) + \mathcal{V}_{\hat{\mathbf{w}}}(\underline{\vartheta}) - \mathcal{V}_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) = 0, \end{aligned}$$

where the last equality follows from property (ii) in Lemma 2. Hence, (83) holds.

Finally, (84) holds by noting that (i) if $W_{0-} \leqslant \hat{\mathbf{w}}$, then $\mathbb{E}[\phi(W_0, \mathcal{E}_{0-})] - \phi(W_{0-}, \mathcal{E}_{0-}) - \mathbb{E}\Delta L_0 = \phi(W_{0-}, \mathcal{E}_{0-}) - \phi(W_{0-}, \mathcal{E}_{0-}) = 0$; and (ii) if $W_{0-} > \hat{\mathbf{w}}$, then $\mathbb{E}[\phi(W_0, \mathcal{E}_{0-})] - \phi(W_{0-}, \mathcal{E}_{0-}) - \mathbb{E}\Delta L_0 = \phi(\hat{\mathbf{w}}, \mathcal{E}_{0-}) - \phi(W_{0-}, \mathcal{E}_{0-}) - (W_{0-} - \hat{\mathbf{w}}) = 0$.

For the case that $\mathcal{E}_{0-} = \emptyset$, by using the same argument as that for the case of $\mathcal{E}_{0-} = \mathsf{I}$, we find that (80)–(84) all hold, which establishes (38).

*Part 2.* Next, we show that functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ as defined in (34) satisfy the optimality condition (20)–(22). First, (22) holds by noting that $V_{\mathsf{I}}(0) = V_{\emptyset}(0) = \underline{v}$. To verify (20) and (21), we consider the following three cases separately: $w \in [0, \underline{\vartheta})$, $w \in [\underline{\vartheta}, \hat{\mathbf{w}})$, and $w \in [\hat{\mathbf{w}}, \infty)$. We study the case of $w \in [\underline{\vartheta}, \hat{\mathbf{w}})$ before $w \in [0, \underline{\vartheta})$.

*Case 1:* $w \in [\underline{\vartheta}, \hat{\mathbf{w}})$. First, we prove that

$$(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) \geqslant 0 \text{ on } [\underline{\vartheta}, \hat{\mathbf{w}}). \tag{91}$$

Obviously, we have $(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) = 0$ for $w \in [\breve{w}(\hat{\mathbf{w}}), \hat{\mathbf{w}})$. It remains to show that (91) holds for $w \in [\underline{\vartheta}, \breve{w}(\hat{\mathbf{w}}))$ if $\underline{\vartheta} < \breve{w}(\hat{\mathbf{w}}) < \tilde{w}$. For $w \in [\underline{\vartheta}, \breve{w}(\hat{\mathbf{w}}))$, $\mathcal{V}_{\hat{\mathbf{w}}}$ is linear and thus we have

$$\begin{aligned}(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) &= (\mu + r)\mathcal{V}_{\hat{\mathbf{w}}}(w) - \mu \mathcal{V}_{\hat{\mathbf{w}}}(w + \beta) + \rho(\bar{w} - w)\mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) - (\mu R - c) + (\rho - r)w \\ &=: g_{\mathsf{I}}(w).\end{aligned}$$

Note that

$$\begin{aligned}g'_{\mathsf{I}}(w) &= (\mu + r)\mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) - \mu \mathcal{V}'_{\hat{\mathbf{w}}}(w + \beta) - \rho \mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) + \rho - r \\ &= (\rho - r)\big(1 - \mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}}))\big) + \mu\big(\mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) - \mathcal{V}'_{\hat{\mathbf{w}}}(w + \beta)\big) \\ &\leqslant (\rho - r)\big(1 - \mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}}))\big) + \mu\big(\mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) - \mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w} + \beta)\big) = 0,\end{aligned}$$

where the inequality follows from the concavity of $\mathcal{V}_{\hat{\mathbf{w}}}$ and the last equality follows from

$$0 = \rho(\bar{w} - \breve{w}(\hat{\mathbf{w}}))\mathcal{V}''_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) = (\rho - r)\big(\mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}})) - 1\big) + \mu\big[\mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}}) + \beta) - \mathcal{V}'_{\hat{\mathbf{w}}}(\breve{w}(\hat{\mathbf{w}}))\big].$$

Consequently, $g_{\mathsf{I}}(w) \geqslant 0$ for all $w \in [\underline{\vartheta}, \breve{w}(\hat{\mathbf{w}}))$, which yields (91). Note that $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) = 0$ if $w \in [\underline{\vartheta}, \bar{\vartheta}]$. Hence, (20) holds by the following result, whose proof can be found in Appendix B.3.4.

Lemma 11. *Under the conditions stated in Proposition 5, we have* $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) \geqslant 0$ *for* $w \in [\bar{\vartheta}, \hat{\mathbf{w}})$.

If $w \in [\bar{\vartheta}, \hat{\mathbf{w}})$, then $V_{\mathsf{I}}(w) - V_{\emptyset}(w) = K > 0$. To establish (21), we need to show that $0 \leqslant V_{\mathsf{I}}(w) - V_{\emptyset}(w) \leqslant K$ if $w \in [\underline{\vartheta}, \bar{\vartheta}]$.

Let $\Phi(w) := V_{\mathsf{I}}(w) - V_{\emptyset}(w)$ and $\chi(w) := \mathcal{V}'_{\hat{\mathbf{w}}}(w) - 1 - \mathsf{c} \cdot r/\rho \cdot w^{r/\rho - 1}$. Obviously, we have $\Phi(\underline{\vartheta}) = 0$ and $\chi(\underline{\vartheta}) = \chi(\bar{\vartheta}) = 0$. It follows from the proof of property (i) in Lemma 9 that $\Phi'(w) = \chi(w) > 0$ for any $w \in (\underline{\vartheta}, \bar{\vartheta})$. Hence, for any $w \in [\underline{\vartheta}, \bar{\vartheta})$, we have $\Phi(w) \geqslant \Phi(\underline{\vartheta}) = 0$ and $\Phi(w) \leqslant \Phi(\bar{\vartheta}) = K$.

*Case 2:* $w \in [0, \underline{\vartheta})$. We claim that

Lemma 12. *Under the conditions stated in Proposition 5, we have* $(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) \geqslant 0$ *for* $w \in (0, \underline{\vartheta})$.

Its proof is rather involved, which is relegated to Appendix B.3.5. Obviously, we have $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) = 0$ on $[0, \underline{\vartheta})$. Hence, (20) holds. Inequality (21) also holds by noting that $V_{\mathsf{I}}(w) = V_{\emptyset}(w)$ in this case.

*Case 3:* $w \in [\hat{\mathbf{w}}, \infty)$. Using the boundary condition $\mathcal{V}_{\hat{\mathbf{w}}}(\hat{\mathbf{w}}) = (\mu R - c - (\rho - r)\hat{\mathbf{w}})/r$, we have

$$(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) = r\mathcal{V}_{\hat{\mathbf{w}}}(\hat{\mathbf{w}}) - (\mu R - c) + (\rho - r)w = (\rho - r)(w - \hat{\mathbf{w}}) \geqslant 0,$$

and

$$\begin{aligned}(\mathcal{A}_{\emptyset} V_{\emptyset})(w) &= r(\mathcal{V}_{\hat{\mathbf{w}}}(\hat{\mathbf{w}}) - K) + (\rho - r)w - R\underline{\mu} \\ &= \mu R - c - (\rho - r)\hat{\mathbf{w}} + (\rho - r)w - R\underline{\mu} - rK \\ &= R\Delta\mu - c + (\rho - r)(w - \hat{\mathbf{w}}) - rK \geqslant 0,\end{aligned}$$

where the last inequality follows as $K < \bar{K}_1 = \bar{V}(\hat{\mathrm{w}}) - \underline{v} = (\mu R - c - (\rho - r)\hat{\mathrm{w}})/r - R\underline{\mu}/r < (R\Delta\mu - c)/r$. Hence, (20) holds. Inequality (21) holds since $V_{\mathrm{I}}(w) - V_{\emptyset}(w) = K$. ▫

*Proof of Lemma 7.* For any $\widetilde{w} \in [\underline{\theta}, \bar{w})$, define

$$C_1(\widetilde{w}, \underline{\theta}) = (\mathcal{V}_{\widetilde{w}}(\underline{\theta}) - \underline{v} - \underline{\theta})\underline{\theta}^{-r/\rho}, \text{ and} \tag{92}$$
$$C_2(\widetilde{w}, \underline{\theta}) = \rho/r \cdot (\mathcal{V}'_{\widetilde{w}}(\underline{\theta}) - 1)\underline{\theta}^{1-r/\rho}.$$

It follows from property (iii) in Lemma 2 that $C_1(\widetilde{w}, \underline{\theta})$ is decreasing in $\widetilde{w}$ and $C_2(\widetilde{w}, \underline{\theta})$ is increasing in $\widetilde{w}$ on $[\underline{\theta}, \bar{w})$.

Note that

$$\begin{aligned} C_1(\underline{\theta}, \underline{\theta}) &= (\mathcal{V}_{\underline{\theta}}(\underline{\theta}) - \underline{v} - \underline{\theta})\underline{\theta}^{-r/\rho} \\ &= \left(\frac{\mu R - c - (\rho - r)\underline{\theta}}{r} - \underline{v} - \underline{\theta}\right) \cdot \underline{\theta}^{-r/\rho} > -\frac{\rho}{r}\underline{\theta}^{1-r/\rho} = C_2(\underline{\theta}, \underline{\theta}), \end{aligned}$$

where the second and third equalities follow from the boundary conditions at $\underline{\theta}$ (see Lemma 6), and the inequality follows from Assumption 1. Besides, it follows from property (iv) in Lemma 6 that $C_1(\widetilde{w}, \underline{\theta}) \to -\infty$ and $C_2(\widetilde{w}, \underline{\theta}) \to \infty$ as $\widetilde{w} \uparrow \bar{w}$. In view of property (ii) of Lemma 6, both $C_1(\widetilde{w}, \underline{\theta})$ and $C_2(\widetilde{w}, \underline{\theta})$ are continuous in $\widetilde{w}$. Therefore, there exists a unique $\widetilde{w}(\underline{\theta}) \in (\underline{\theta}, \bar{w})$ such that $C_1(\widetilde{w}(\underline{\theta}), \underline{\theta}) = C_2(\widetilde{w}(\underline{\theta}), \underline{\theta})$. Let $C(\underline{\theta}) := C_1(\widetilde{w}(\underline{\theta}), \underline{\theta})$. Then, $\widetilde{w}(\underline{\theta})$ and $C(\underline{\theta})$ satisfy (30)–(31) as desired. ▫

*Proof of Lemma 8.* Define

$$h(\widetilde{w}, \underline{\theta}) = \mathcal{V}_{\widetilde{w}}(\underline{\theta}) - \underline{v} - \underline{\theta} - \rho/r \cdot (\mathcal{V}'_{\widetilde{w}}(\underline{\theta}) - 1)\underline{\theta}. \tag{93}$$

It follows from property (iii) in Lemma 2 that $h(\widetilde{w}, \underline{\theta})$ is decreasing in $\widetilde{w}$. Besides, $h(\widetilde{w}(\underline{\theta}), \underline{\theta}) = 0$. Note that $h(\widetilde{w}, \underline{\theta})$ is continuously differentiable in $\widetilde{w}$ and $\underline{\theta}$ (see properties (ii) and (iii) in Lemma 6). Hence, $\widetilde{w}(\underline{\theta})$ is continuously differentiable in $\underline{\theta}$.

Since $h(\widetilde{w}, 0) = \mathcal{V}_{\widetilde{w}}(0) - \underline{v}$, we have $\widetilde{w}(0) = \hat{\mathrm{w}} > 0$. Besides, it follows from $\widetilde{w}(\underline{\theta}) \in (\underline{\theta}, \bar{w})$ that $\lim_{\underline{\theta} \uparrow \bar{w}} \widetilde{w}(\underline{\theta}) = \bar{w} > \hat{\mathrm{w}}$.

Write $h_1(\widetilde{w}, \underline{\theta}) = \partial h(\widetilde{w}, \underline{\theta})/\partial \widetilde{w}$ and $h_2(\widetilde{w}, \underline{\theta}) = \partial h(\widetilde{w}, \underline{\theta})/\partial \underline{\theta}$. Then, we have $h_1(\widetilde{w}, \underline{\theta}) < 0$,

$$\begin{aligned} h_2(\widetilde{w}, \underline{\theta}) &= \frac{\rho - r - (\rho - r)\mathcal{V}'_{\widetilde{w}}(\underline{\theta}) - \rho\underline{\theta}\mathcal{V}''_{\widetilde{w}}(\underline{\theta})}{r}, \text{ and} \\ \widetilde{w}'(\underline{\theta}) &= -\frac{h_2(\widetilde{w}(\underline{\theta}), \underline{\theta})}{h_1(\widetilde{w}(\underline{\theta}), \underline{\theta})}. \end{aligned}$$

It follows from Condition (L1) and Lemma 4 that $m^K = \mathcal{V}'_{\hat{\mathrm{w}}}(\bar{\theta}^K) > 1$, which implies that $\mathcal{V}'_{\hat{\mathrm{w}}}(0) > 1$ by the concavity of $\mathcal{V}_{\hat{\mathrm{w}}}$. Therefore, we have $h_2(\widetilde{w}(0), 0) = (\rho - r - (\rho - r)\mathcal{V}'_{\hat{\mathrm{w}}}(0))/r < 0$ and thus $\widetilde{w}'(0) < 0$. This implies that $\widetilde{w}(\underline{\theta})$ is strictly decreasing in $\underline{\theta}$ when $\underline{\theta}$ is near 0.

It follows from $\lim_{\underline{\theta} \uparrow \bar{w}} \widetilde{w}(\underline{\theta}) = \bar{w} > \hat{\mathrm{w}} = \widetilde{w}(0)$ and the continuity of $\widetilde{w}'(\underline{\theta})$ in $\underline{\theta}$ that value $\underline{\theta}^0 := \inf\{\underline{\theta} \in (0, \bar{w}) : \widetilde{w}'(\underline{\theta}) \geqslant 0\}$ is well defined, and $\widetilde{w}'(\underline{\theta}) < 0$ for any $\underline{\theta} \in [0, \underline{\theta}^0)$ and $\widetilde{w}'(\underline{\theta}^0) = 0$. Consequently, we have

$$\rho - r - (\rho - r)\mathcal{V}'_{\widetilde{w}(w)}(w) - \rho w \mathcal{V}''_{\widetilde{w}(w)}(w) < 0, \text{ for any } w \in [0, \underline{\theta}^0); \text{ and} \tag{94}$$

$$\rho - r - (\rho - r)\mathcal{V}'_{\widetilde{w}(\underline{\theta}^0)}(\underline{\theta}^0) - \rho\underline{\theta}^0\mathcal{V}''_{\widetilde{w}(\underline{\theta}^0)}(\underline{\theta}^0) = 0. \tag{95}$$

Note that $\widetilde{w}(\underline{\theta})$ is strictly decreasing in $\underline{\theta}$ on $[0,\underline{\theta}^0)$. We claim that $C(\underline{\theta})$ is strictly increasing in $\underline{\theta}$ on $[0,\underline{\theta}^0)$. In fact, for any $\underline{\theta}\in(0,\underline{\theta}^0)$, we have

$$\begin{aligned}
C'(\underline{\theta}) =& C_1'(\widetilde{w}(\underline{\theta}),\underline{\theta}) = \frac{\mathrm{d}}{\mathrm{d}\underline{\theta}}\left[\left(\mathcal{V}_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - \underline{v} - \underline{\theta}\right)\underline{\theta}^{-r/\rho}\right] \\
=& \left(\mathcal{V}'_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) + \widetilde{w}'(\underline{\theta})\cdot\frac{\partial \mathcal{V}_{\widetilde{w}(\underline{\theta})}(\underline{\theta})}{\partial \widetilde{w}(\underline{\theta})} - 1\right)\underline{\theta}^{-r/\rho} - \frac{r}{\rho}\left(\mathcal{V}_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - \underline{v} - \underline{\theta}\right)\underline{\theta}^{-r/\rho-1} \\
=& \widetilde{w}'(\underline{\theta})\cdot\frac{\partial \mathcal{V}_{\widetilde{w}(\underline{\theta})}(\underline{\theta})}{\partial \widetilde{w}(\underline{\theta})} > 0,
\end{aligned}$$

where the last equality follows from $C_1(\widetilde{w}(\underline{\theta}),\underline{\theta}) = C_2(\widetilde{w}(\underline{\theta}),\underline{\theta})$, and the inequality follows from $\widetilde{w}'(\underline{\theta}) < 0$ and $\partial\mathcal{V}_{\widetilde{w}}(w)/\partial\widetilde{w} < 0$ in view of property (iii) in Lemma 2.

Note that $\mathcal{V}'_{\hat{w}}(0) > 1$. Hence, it follows from the continuity of $\widetilde{w}(\underline{\theta})$ in $\underline{\theta}$ and properties (ii) and (iii) of Lemma 6 that there exists $\epsilon > 0$ such that $\mathcal{V}'_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) > 1$ for any $\underline{\theta}\in(0,\epsilon)$. Consequently, $C(\underline{\theta}) = C_2(\widetilde{w}(\underline{\theta}),\underline{\theta}) > 0$ for any $\underline{\theta}\in(0,\epsilon)$. As a result, we have $C(\underline{\theta}) > 0$ for any $\underline{\theta}\in(0,\underline{\theta}^0)$ by noting that $C(\underline{\theta})$ is strictly increasing in $\underline{\theta}$ on $[0,\underline{\theta}^0)$. ▫

*Proof of Lemma 9.* (i) Define

$$\Psi(w,\underline{\theta}) = \mathcal{V}'_{\widetilde{w}(\underline{\theta})}(w) - 1 - C(\underline{\theta})r/\rho\cdot w^{r/\rho-1}.$$

It follows from Lemma 8 and property (iii) in Lemma 2 that $\Psi(w,\underline{\theta})$ is decreasing in $\underline{\theta}$ on $(0,\underline{\theta}^0)$. Therefore, for any $w\in(0,\underline{\theta})$, we have $\Psi(w,\underline{\theta}) < \Psi(w,w) = 0$, which implies that $\underline{\theta} = \inf\{w\geqslant 0 : \Psi(w,\underline{\theta}) = 0\}$ as $\Psi(\underline{\theta},\underline{\theta}) = 0$. Moreover, we have

$$\begin{aligned}
\frac{\partial\Psi}{\partial w}(\underline{\theta},\underline{\theta}) &= \mathcal{V}''_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - C(\underline{\theta})r/\rho\cdot(r/\rho-1)\underline{\theta}^{r/\rho-2} \\
&= \mathcal{V}''_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - C_2(\widetilde{w}(\underline{\theta}),\underline{\theta})r/\rho\cdot(r/\rho-1)\underline{\theta}^{r/\rho-2} \\
&= \mathcal{V}''_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - \rho/r\cdot(\mathcal{V}'_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - 1)\underline{\theta}^{1-r/\rho}r/\rho\cdot(r/\rho-1)\underline{\theta}^{r/\rho-2} \\
&= \mathcal{V}''_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - (\mathcal{V}'_{\widetilde{w}(\underline{\theta})}(\underline{\theta}) - 1)\cdot(r/\rho-1)/\underline{\theta} \\
&> 0,
\end{aligned}$$

where the last inequality follows from (94). This implies that $\Psi(w,\underline{\theta}) > 0$ for $w\in(\underline{\theta},\underline{\theta}+\epsilon)$ with some $\epsilon > 0$. It follows from property (ii) in Lemma 6 that $\Psi(w,\underline{\theta})$ is continuous in $w$. Besides, we have

$$\begin{aligned}
\Psi(\widetilde{w}(\underline{\theta}),\underline{\theta}) &= \mathcal{V}'_{\widetilde{w}(\underline{\theta})}(\widetilde{w}(\underline{\theta})) - (1 + C(\underline{\theta})r/\rho\cdot\widetilde{w}(\underline{\theta})^{r/\rho-1}) \\
&= -\left(1 + C(\underline{\theta})r/\rho\cdot\widetilde{w}(\underline{\theta})^{r/\rho-1}\right) < 0
\end{aligned}$$

and $\Psi(\underline{\theta},\underline{\theta}) = 0$. Hence,

$$\begin{aligned}
\bar{\theta}(\underline{\theta}) &:= \inf\{w > \underline{\theta} : \Psi(w,\underline{\theta}) \leqslant 0\} \\
&= \inf\left\{w > \underline{\theta} : \mathcal{V}'_{\widetilde{w}(\underline{\theta})}(w) \leqslant 1 + C(\underline{\theta})r/\rho\cdot w^{r/\rho-1}\right\}
\end{aligned}$$

is well defined and $\bar{\theta}(\underline{\theta}) < \widetilde{w}(\underline{\theta})$.

(ii) This follows immediately by noting that $\Psi(w,\underline{\theta})$ is decreasing in $\underline{\theta}$ on $[0,\underline{\theta}^0)$ and is continuous in $(w,\underline{\theta})$.

(iii) According to (95), we have $\frac{\partial\Psi}{\partial w}(\underline{\theta}^0,\underline{\theta}^0) = 0$, which implies that $\Psi(\cdot,\underline{\theta}^0)$ attains its local maximum at $\underline{\theta}^0$. Hence, we have $\lim_{\underline{\theta}\uparrow\underline{\theta}^0}\bar{\theta}(\underline{\theta}) = \underline{\theta}^0$ by noting that $\bar{\theta}(\underline{\theta}) \geqslant \underline{\theta}$. ▫

*Proof of Lemma 10.* Note that

$$\psi(\underline{\theta})=\int_{\underline{\theta}}^{\bar{\theta}(\underline{\theta})}[\mathcal{V}'_{\widetilde{w}(\underline{\theta})}(y)-(1+C(\underline{\theta})r/\rho\cdot y^{r/\rho-1})]\mathrm{d}y=\int_{\underline{\theta}}^{\bar{\theta}(\underline{\theta})}\Psi(y,\underline{\theta})\mathrm{d}y.$$

Fix any $\underline{\theta}^1<\underline{\theta}^2$ in $(0,\underline{\theta}^0)$. We have

$$\psi(\underline{\theta}^1)=\int_{\underline{\theta}^1}^{\bar{\theta}(\underline{\theta}^1)}\Psi(y,\underline{\theta}^1)\mathrm{d}y>\int_{\underline{\theta}^2}^{\bar{\theta}(\underline{\theta}^2)}\Psi(y,\underline{\theta}^1)\mathrm{d}y>\int_{\underline{\theta}^2}^{\bar{\theta}(\underline{\theta}^2)}\Psi(y,\underline{\theta}^2)\mathrm{d}y=\psi(\underline{\theta}^2),$$

where the first inequality follows from $\Psi(y,\underline{\theta}^1)>0$ for $y\in(\underline{\theta}^1,\underline{\theta}^2)\cup(\bar{\theta}(\underline{\theta}^2),\bar{\theta}(\underline{\theta}^1))$, and the second inequality holds because $\Psi(y,\underline{\theta})$ is decreasing in $\underline{\theta}$ on $(0,\underline{\theta}^0)$. Hence, $\psi(\underline{\theta})$ is decreasing in $\underline{\theta}$ on $(0,\underline{\theta}^0)$. The continuity of $\psi(\underline{\theta})$ follows from properties (ii) and (iii) in Lemma 6.

Since $\lim_{\underline{\theta}\uparrow\underline{\theta}^0}\bar{\theta}(\underline{\theta})=\underline{\theta}^0$, we have $\lim_{\underline{\theta}\uparrow\underline{\theta}^0}\psi(\underline{\theta})=0$. Note that $\widetilde{w}(0)=\hat{\mathrm{w}}$ and $C_2(\widetilde{w},0)=0$ for any $\widetilde{w}\in[0,\bar{w})$. Hence, we have $\lim_{\underline{\theta}\downarrow 0}C(\underline{\theta})=0$ and thus $\lim_{\underline{\theta}\downarrow 0}\bar{\theta}(\underline{\theta})=\inf\{w>0:\mathcal{V}'_{\hat{\mathrm{w}}}(w)=1\}$. This yields

$$\begin{aligned}\lim_{\underline{\theta}\downarrow 0}\psi(\underline{\theta})&=\int_0^\infty(\mathcal{V}'_{\hat{\mathrm{w}}}(y)-1)^+dy>\int_0^{\bar{\boldsymbol{\theta}}}(\mathcal{V}'_{\hat{\mathrm{w}}}(y)-1)dy\\&=\mathcal{V}_{\hat{\mathrm{w}}}(\bar{\theta}^K)-\underline{v}-\bar{\theta}^K=K+(m^K-1)\bar{\theta}^K>K,\end{aligned}$$

where the first inequality follows from the fact that $\mathcal{V}'_{\hat{\mathrm{w}}}(\bar{\theta}^K)=m^K>1$ and that $\mathcal{V}'_{\hat{\mathrm{w}}}$ is nonincreasing, and the last inequality holds since $m^K>1$. Consequently, it follows from the continuity of $\psi(\cdot)$ that there exists a $\underline{\vartheta}\in(0,\underline{\theta}^0)$ such that $\psi(\underline{\vartheta})=K$. ▫

**B.3.4. Proof of Lemma 11** For $w\in[\bar{\vartheta},\hat{\mathbf{w}})$, it holds that $V_\emptyset(w)=\mathcal{V}_{\hat{\mathbf{w}}}(w)-K$. Define $\varpi_0=\inf\{w>0:\mathcal{V}'_{\hat{\mathrm{w}}}(w)=1\}$, which is well-defined by noting that $\mathcal{V}'_{\hat{\mathrm{w}}}(0)>1$ and $\mathcal{V}'_{\hat{\mathrm{w}}}(\hat{\mathrm{w}})=0$. For any $w\in[\varpi_0,\hat{\mathbf{w}}]$, let

$$(\mathcal{A}_\emptyset V_\emptyset)(w)=r(\mathcal{V}_{\hat{\mathbf{w}}}(w)-K)-\rho w\mathcal{V}'_{\hat{\mathbf{w}}}(w)+(\rho-r)w-\underline{\mu}R=:g_\emptyset(w).$$

It follows from property (iii) in Lemma 2 and $\hat{\mathbf{w}}<\hat{\mathrm{w}}$ that $\mathcal{V}'_{\hat{\mathbf{w}}}(w)<\mathcal{V}'_{\hat{\mathrm{w}}}(w)$ and $\mathcal{V}_{\hat{\mathbf{w}}}(w)>\mathcal{V}_{\hat{\mathrm{w}}}(w)$ for $w\in[0,\hat{\mathbf{w}}]$. Hence, for $w\in[\varpi_0,\hat{\mathbf{w}}]$, we have

$$\begin{aligned}g'_\emptyset(w)&=r\mathcal{V}'_{\hat{\mathbf{w}}}(w)-\rho\mathcal{V}'_{\hat{\mathbf{w}}}(w)-\rho w\mathcal{V}''_{\hat{\mathbf{w}}}(w)+\rho-r\\&=(\rho-r)(1-\mathcal{V}'_{\hat{\mathbf{w}}}(w))-\rho w\mathcal{V}''_{\hat{\mathbf{w}}}(w)\geqslant 0,\end{aligned}$$

where the last inequality follows from that $\mathcal{V}'_{\hat{\mathbf{w}}}(w)\leqslant\mathcal{V}'_{\hat{\mathrm{w}}}(w)\leqslant 1$ for $w\geqslant\varpi_0$ and the concavity of $\mathcal{V}_{\hat{\mathbf{w}}}$. Besides, we have

$$\begin{aligned}g_\emptyset(\varpi_0)&>r(\mathcal{V}_{\hat{\mathbf{w}}}(\varpi_0)-K)-\rho\varpi_0+(\rho-r)\varpi_0-\underline{\mu}R\\&>r(\underline{v}+\varpi_0+K-K)-\rho\varpi_0+(\rho-r)\varpi_0-\underline{\mu}R=0,\end{aligned}$$

where the first inequality follows from $\mathcal{V}'_{\hat{\mathbf{w}}}(\varpi_0)<\mathcal{V}'_{\hat{\mathrm{w}}}(\varpi_0)=1$, and the second inequality follows from the fact that $\mathcal{V}_{\hat{\mathbf{w}}}(\varpi_0)>\mathcal{V}_{\hat{\mathrm{w}}}(\varpi_0)>\underline{v}+\varpi_0+K$ (the last inequality follows from $m^K>1$). As a result, we have $(\mathcal{A}_\emptyset V_\emptyset)(w)\geqslant 0$ for all $w\in[\varpi_0,\hat{\mathbf{w}}]$.

Next, we will prove that $(\mathcal{A}_\emptyset V_\emptyset)(w)\geqslant 0$ for $w\in[\bar{\vartheta},\varpi_0)$ by a contradictory argument.

Suppose, on the contrary, that there exists a number $\varpi\in(\bar{\vartheta},\varpi_0)$ such that $(\mathcal{A}_\emptyset V_\emptyset)(\varpi)<0$. Then, we have

$$(\mathcal{A}_\emptyset V_\emptyset)(\varpi)=r(\mathcal{V}_{\hat{\mathbf{w}}}(\varpi)-K)-\rho\varpi\cdot\mathcal{V}'_{\hat{\mathbf{w}}}(\varpi)+(\rho-r)\varpi-\underline{\mu}R<0,$$

and thus

$$\mathcal{V}'_{\hat{\mathbf{w}}}(\varpi) > \frac{(\rho - r)\varpi + r(\mathcal{V}_{\hat{\mathbf{w}}}(\varpi) - K - \underline{v})}{\rho\varpi}. \tag{96}$$

It follows from (89) that $\lim_{\underline{\theta}\downarrow 0}\bar{\theta}(\underline{\theta}) = \inf\{w > 0 : \mathcal{V}'_{\hat{\mathbf{w}}}(w) = 1\} = \varpi_0$. Note that $\varpi > \bar{\vartheta} = \bar{\theta}(\underline{\vartheta})$ and $\varpi < \varpi_0$. Hence, it follows from Lemma 9 that there exists a number $\underline{\theta}' \in (0, \underline{\vartheta})$ such that $\bar{\theta}(\underline{\theta}') = \varpi$. Using Lemmas 8 and 10, we have $\widetilde{w}(\underline{\theta}') > \widetilde{w}(\underline{\vartheta}) = \hat{\mathbf{w}}$, $C(\underline{\theta}') < C(\underline{\vartheta}) = \mathsf{c}$ and

$$\psi(\underline{\theta}') = \mathcal{V}_{\widetilde{w}(\underline{\theta}')}(\varpi) - \left[\underline{v} + \varpi + C(\underline{\theta}')\varpi^{r/\rho}\right] > \psi(\underline{\vartheta}) = K. \tag{97}$$

Moreover, we have

$$\mathcal{V}_{\hat{\mathbf{w}}}(\varpi) > \mathcal{V}_{\widetilde{w}(\underline{\theta}')}(\varpi) \text{ and } \mathcal{V}'_{\hat{\mathbf{w}}}(\varpi) < \mathcal{V}'_{\widetilde{w}(\underline{\theta}')}(\varpi), \tag{98}$$

in view of property (iii) in Lemma 2.

Consequently,

$$\begin{aligned}\mathcal{V}'_{\hat{\mathbf{w}}}(\varpi) &> \frac{(\rho - r)\varpi + r(\mathcal{V}_{\hat{w}(\underline{\theta}')}(\varpi) - K - \underline{v})}{\rho\varpi} \\ &> \frac{(\rho - r)\varpi + r\cdot[(\underline{v} + \varpi + C(\underline{\theta}')\varpi^{r/\rho}) - \underline{v}]}{\rho\varpi} \\ &= 1 + r/\rho\cdot C(\underline{\theta}')\varpi^{r/\rho - 1} = \mathcal{V}'_{\widetilde{w}(\underline{\theta}')}(\varpi),\end{aligned}$$

where the first inequality follows from (96) and (98), the second inequality follows from (97), and the last equality follows from $\varpi = \bar{\theta}(\underline{\theta}')$ and the definition of $\bar{\theta}(w)$. This reaches a contradiction with (98).

**B.3.5. Proof of Lemma 12** The proof of Lemma 12 is probably the most complex proof in the paper. As mentioned in the paragraph below Proposition 5, the key step is to establish Lemma 14, which states that either $\mathcal{A}_{\mathsf{l}}V_{\mathsf{l}}$'s first-order-derivative is negative, or its second-order-derivative is positive on $(0, \underline{\vartheta})$. This crucial result is obtained by studying a total of four cases, which are summarized as Lemmas 15–18.

Following from $V_{\mathsf{l}}(w) = \underline{v} + w + \mathsf{c}w^{r/\rho}$ for $w \in [0, \underline{\vartheta}]$ and

$$(\mathcal{A}_{\mathsf{l}}f - \mathcal{A}_{\emptyset}f)(w) = \mu(f(w) - f(w+\beta)) + \rho\bar{w}f'(w) - (R\Delta\mu - c),$$

we define

$$\begin{aligned}g_{\mathsf{l}}(w) := (\mathcal{A}_{\mathsf{l}}V_{\mathsf{l}})(w) &= \mu(V_{\mathsf{l}}(w) - V_{\mathsf{l}}(w+\beta)) + \rho\bar{w}V_{\mathsf{l}}'(w) - (R\Delta\mu - c) \\ &= \mu(\underline{v} + w + \mathsf{c}w^{r/\rho} - V_{\mathsf{l}}(w+\beta)) + \rho\bar{w}(1 + \mathsf{c}r/\rho\cdot w^{r/\rho-1}) - (R\Delta\mu - c)\end{aligned}$$

for $w \in (0, \underline{\vartheta}]$. From property (ii) in Lemma 6, we have $g_{\mathsf{l}} \in C^1((0, \underline{\vartheta})) \cap C^2((0, \underline{\vartheta})\backslash\{\hat{\mathbf{w}} - \beta, \underline{\vartheta} - \beta\})$. (Note that $V_{\mathsf{l}}$ may not be twice-differentiable at $\underline{\vartheta}$.) Besides, we have

$$g_{\mathsf{l}}'(w) = \mu\big(1 + \mathsf{c}r/\rho\cdot w^{r/\rho-1} - V_{\mathsf{l}}'(w+\beta)\big) + \rho\bar{w}\cdot\mathsf{c}r/\rho\cdot(r/\rho - 1)w^{r/\rho-2}, \tag{99}$$

$$g_{\mathsf{l}}''(w) = \mathsf{c}r/\rho\cdot(r/\rho - 1)w^{r/\rho-3}[\mu w + \rho\bar{w}(r/\rho - 2)] - \mu V_{\mathsf{l}}''(w+\beta). \tag{100}$$

Lemma 12 is equivalent to $g_{\mathsf{l}}(w) \geqslant 0$ for $w \in (0, \underline{\vartheta}]$. From (91) at $\underline{\vartheta}$, we have $g_{\mathsf{l}}(\underline{\vartheta}) \geqslant 0$. Moreover, the following holds.

Lemma 13. *We have*

$$g_{\mathsf{l}}'(\underline{\vartheta}) < 0. \tag{101}$$

*Proof.* Note that $\breve{w}(\hat{\mathbf{w}}) < \beta$ (see property (i) in Lemma 2), which implies $\breve{w}(\hat{\mathbf{w}}) < \underline{\vartheta} + \beta$.

If $\underline{\vartheta} \leqslant \breve{w}(\hat{\mathbf{w}})$, then

$$\begin{aligned} g_{\mathsf{l}}'(\underline{\vartheta}) &= \mu(V_{\mathsf{l}}'(\underline{\vartheta}) - V_{\mathsf{l}}'(\underline{\vartheta} + \beta)) + \rho\bar{w}V_{\mathsf{l}}''(\underline{\vartheta}) \\ &< \mu\big(V_{\mathsf{l}}'(\breve{w}(\hat{\mathbf{w}})) - V_{\mathsf{l}}'(\breve{w}(\hat{\mathbf{w}}) + \beta)\big) + \rho\bar{w}V_{\mathsf{l}}''(\underline{\vartheta}) \\ &= (\rho - r)(V_{\mathsf{l}}'(\breve{w}(\hat{\mathbf{w}})) - 1) - \rho(\bar{w} - \breve{w}(\hat{\mathbf{w}}))V_{\mathsf{l}}''(\breve{w}(\hat{\mathbf{w}})) + \rho\bar{w}V_{\mathsf{l}}''(\underline{\vartheta}) \\ &= (\rho - r)\mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1} + \rho\bar{w}\cdot \mathsf{c}r/\rho \cdot (r/\rho - 1)\underline{\vartheta}^{r/\rho - 2} \\ &= (\rho - r)\mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 2}(\underline{\vartheta} - \bar{w}) < 0, \end{aligned}$$

where the first inequality follows from $V_{\mathsf{l}}'(\underline{\vartheta}) = \mathcal{V}_{\hat{\mathbf{w}}}'(\underline{\vartheta}) = \mathcal{V}_{\hat{\mathbf{w}}}'(\breve{w}(\hat{\mathbf{w}})) = V_{\mathsf{l}}'(\breve{w}(\hat{\mathbf{w}}))$ and $V_{\mathsf{l}}'(\underline{\vartheta} + \beta) = \mathcal{V}_{\hat{\mathbf{w}}}'(\underline{\vartheta} + \beta) \geqslant \mathcal{V}_{\hat{\mathbf{w}}}'(\breve{w}(\hat{\mathbf{w}}) + \beta) = V_{\mathsf{l}}'(\breve{w}(\hat{\mathbf{w}}) + \beta)$, the second equality follows from the fact that $V_{\mathsf{l}}(\breve{w}(\hat{\mathbf{w}}))$ satisfies (24), as well as (85) at $\breve{w}(\hat{\mathbf{w}})$, and the third equality follows from $V_{\mathsf{l}}''(\breve{w}(\hat{\mathbf{w}})) = \mathcal{V}_{\hat{\mathbf{w}}}''(\breve{w}(\hat{\mathbf{w}})) = 0$, $V_{\mathsf{l}}'(\breve{w}(\hat{\mathbf{w}})) = V_{\mathsf{l}}'(\underline{\vartheta}) = 1 + \mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1}$ and $V_{\mathsf{l}}''(\underline{\vartheta}) = \mathsf{c}r/\rho \cdot (r/\rho - 1)\underline{\vartheta}^{r/\rho - 2}$.

If $\underline{\vartheta} > \breve{w}(\hat{\mathbf{w}})$, then it follows from (85) at $\underline{\vartheta}$ that

$$(\mu + r)V_{\hat{\mathbf{w}}}'(\underline{\vartheta}) - \mu V_{\hat{\mathbf{w}}}'\big((\underline{\vartheta} + \beta) \wedge \hat{\mathbf{w}}\big) + \rho(\bar{w} - \underline{\vartheta})V_{\hat{\mathbf{w}}}''(\underline{\vartheta}) - \rho V_{\hat{\mathbf{w}}}'(\underline{\vartheta}) + \rho - r = 0.$$

Note that $V_{\hat{\mathbf{w}}}'(\underline{\vartheta}) = 1 + \mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1}$ and $V_{\mathsf{l}}'(w + \beta) = V_{\hat{\mathbf{w}}}'\big((\underline{\vartheta} + \beta) \wedge \hat{\mathbf{w}}\big)$. Hence, we have

$$\begin{aligned} g_{\mathsf{l}}'(\underline{\vartheta}) &= \mu\big(1 + \mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1} - V_{\mathsf{l}}'(\underline{\vartheta} + \beta)\big) + \rho\bar{w}\cdot \mathsf{c}r/\rho \cdot (r/\rho - 1)\underline{\vartheta}^{r/\rho - 2} \\ &= (\rho - r)\mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1} - \rho(\bar{w} - \underline{\vartheta})V_{\hat{\mathbf{w}}}''(\underline{\vartheta}) + \rho\bar{w}\cdot \mathsf{c}r/\rho \cdot (r/\rho - 1)\underline{\vartheta}^{r/\rho - 2} \\ &< (\rho - r)\mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1} - \frac{(\bar{w} - \underline{\vartheta})(\rho - r)(1 - V_{\hat{\mathbf{w}}}'(\underline{\vartheta}))}{\underline{\vartheta}} + \rho\bar{w}\cdot \mathsf{c}r/\rho \cdot (r/\rho - 1)\underline{\vartheta}^{r/\rho - 2} \\ &= (\rho - r)\mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1} + \frac{(\bar{w} - \underline{\vartheta})(\rho - r)\mathsf{c}r/\rho \cdot \underline{\vartheta}^{r/\rho - 1}}{\underline{\vartheta}} + \rho\bar{w}\cdot \mathsf{c}r/\rho \cdot (r/\rho - 1)\underline{\vartheta}^{r/\rho - 2} \\ &= 0, \end{aligned}$$

where the inequality follows from (94) at $\underline{\vartheta}$ and $\hat{\mathbf{w}} = \widetilde{w}(\underline{\vartheta})$.

The proof is complete by combining the above two cases. ◻

Next, we show the following crucial result.

Lemma 14. *For any $w \in (0, \underline{\vartheta})$, either $g_{\mathsf{l}}'(w) \leqslant 0$ or $g_{\mathsf{l}}''(w) \geqslant 0$.*

The above result, combining with (101), yields that $g_{\mathsf{l}}'(w) \leqslant 0$ for any $w \in (0, \underline{\vartheta}]$, which immediately concludes the desired result. In fact, if it fails to hold, $w^{\dagger} := \sup\{w \in (0, \underline{\vartheta}) : g_{\mathsf{l}}'(w) > 0\}$ is well-defined, which further implies that $g_{\mathsf{l}}''(w^{\dagger}) < 0$. This contradicts with Lemma 14.

Lemma 14 follows immediately from Lemmas 15–18 below.

Lemma 15. *For any $w \in [0, \bar{\vartheta} - \beta]$, we have $g_{\mathsf{l}}'(w) < 0$.*

LEMMA 16. *For any* $w \in (\bar{\vartheta} - \beta, \underline{\vartheta})$ *such that* $\mathcal{V}'''_{\hat{\mathbf{w}}}(w+\beta) \leqslant 0$*, we have* $g''_{\mathsf{l}}(w) > 0$.

LEMMA 17. *For any* $w \in \big(0, (2 - r/\rho)\beta \wedge \underline{\vartheta}\big)$*, we have* $g''_{\mathsf{l}}(w) \geqslant 0$.

LEMMA 18. *For any* $w \in [(1 - r/\rho)\beta, \underline{\vartheta})$ *such that* $\mathcal{V}'''_{\hat{\mathbf{w}}}(w+\beta) > 0$*, we have* $g'_{\mathsf{l}}(w) \leqslant 0$.

It is worth mentioning that $\mathcal{V}_{\hat{\mathbf{w}}} \in C^1(\mathbb{R}_+) \cap C^2(\mathbb{R}_+ \backslash \{\hat{\mathbf{w}}\}) \cap C^3(\mathbb{R}_+ \backslash \{\hat{\mathbf{w}}, \hat{\mathbf{w}} - \beta, \breve{w}(\hat{\mathbf{w}})\}) \cap C^4(\mathbb{R}_+ \backslash \{\hat{\mathbf{w}}, \hat{\mathbf{w}} - \beta, \hat{\mathbf{w}} - 2\beta, \breve{w}(\hat{\mathbf{w}})\})$ and thus the 3rd-order-derivative of the function $\mathcal{V}_{\hat{\mathbf{w}}}$ might not exist at some points. Hence, in Lemmas 16 and 18, we follow the convention to use $\mathcal{V}'''_{\hat{\mathbf{w}}}(w)$ to represent the left-3rd-order-derivative of the function $\mathcal{V}_{\hat{\mathbf{w}}}$ at $w$. In their proofs, we also use the following technical lemma.

LEMMA 19. *If* $2\rho < r + \mu$ *or* $\rho > r + \mu$*, there exists a value* $\varsigma \in [\breve{w}(\hat{\mathbf{w}}), \hat{\mathbf{w}}]$*, such that* $\mathcal{V}'''_{\tilde{w}} > 0$ *over* $(\varsigma, \hat{\mathbf{w}}]$ *and* $\mathcal{V}'''_{\hat{\mathbf{w}}} < 0$ *over* $(\breve{w}(\hat{\mathbf{w}}), \varsigma]$*; otherwise,* $\mathcal{V}'''_{\hat{\mathbf{w}}} \leqslant 0$ *over* $(\breve{w}(\hat{\mathbf{w}}), \hat{\mathbf{w}})$.

By investigating the proof, it is clear to see that Lemma 19 actually holds for any $\widetilde{w} \in [0, \bar{w})$, rather than a specific value $\hat{\mathbf{w}}$. Here, we only need the result at $\hat{\mathbf{w}}$ to prove Lemmas 16 and 18.

The remaining part of this section is devoted to the proofs of Lemmas 15–19. To proceed, we need some preliminary results of $\mathsf{V_c}$. Recall (23). Obviously, $\mathsf{V_c}$ is strictly concave on $[0, \infty)$, and

$$\mathsf{V}'''_{\mathsf{c}}(w) = \mathsf{c}r/\rho \cdot (r/\rho - 1)(r/\rho - 2) w^{r/\rho - 3} > 0, \tag{102}$$

$$\mathsf{V}''''_{\mathsf{c}}(w) = \mathsf{c}r/\rho \cdot (r/\rho - 1)(r/\rho - 2)(r/\rho - 3) w^{r/\rho - 4} < 0, \tag{103}$$

for all $w \in (0, \infty)$. Therefore, $\mathsf{V}'_{\mathsf{c}}$ is strictly convex and $\mathsf{V}''_{\mathsf{c}}$ is strictly concave, which further implies that

$$\big(\mathsf{V}'_{\mathsf{c}}(w) - \mathsf{V}'_{\mathsf{c}}(w+\beta)\big) + \beta \mathsf{V}''_{\mathsf{c}}(w) < 0, \text{ and} \tag{104}$$

$$\big(\mathsf{V}''_{\mathsf{c}}(w) - \mathsf{V}''_{\mathsf{c}}(w+\beta)\big) + \beta \mathsf{V}'''_{\mathsf{c}}(w) > 0 \tag{105}$$

for any $w \in (0, \infty)$.

*Proof of Lemma 15.* Following the fact that $V_{\mathsf{l}}(w) = \mathsf{V_c}(w)$ for $w \in [0, \underline{\vartheta}]$, we have

$$\begin{aligned} g'_{\mathsf{l}}(w) &= \mu(V'_{\mathsf{l}}(w) - V'_{\mathsf{l}}(w+\beta)) + \mu\beta \mathsf{V}''_{\mathsf{c}}(w) \\ &\leqslant \mu(\mathsf{V}'_{\mathsf{c}}(w) - \mathsf{V}'_{\mathsf{c}}(w+\beta)) + \mu\beta \mathsf{V}''_{\mathsf{c}}(w) < 0, \end{aligned}$$

where the first inequality follows from $V'_{\mathsf{l}}(w+\beta) \geqslant \mathsf{V}'_{\mathsf{c}}(w+\beta)$, which holds because $w + \beta \leqslant \bar{\vartheta}$ and (89) with $\bar{\vartheta} = \bar{\theta}(\underline{\vartheta})$, and the second inequality follows from (104). ▫

*Proof of Lemma 16.* Define $\phi(w) = \mathcal{V}'_{\hat{\mathbf{w}}}(w) - \mathsf{V}'_{\mathsf{c}}(w)$. Since $\bar{\vartheta} = \inf\{w > \underline{\vartheta} : \phi(w) = 0\}$ and $\phi > 0$ over $(\underline{\vartheta}, \bar{\vartheta})$, we have $\phi'(\bar{\vartheta}) \leqslant 0$. It follows from $\mathcal{V}'''_{\hat{\mathbf{w}}}(w+\beta) \leqslant 0$ and Lemma 19 that $\mathcal{V}'''_{\hat{\mathbf{w}}} < 0$ over $(\breve{w}(\hat{\mathbf{w}}), w+\beta)$.

For any $w \in (\bar{\vartheta} - \beta, \underline{\vartheta})$, we have

$$\begin{aligned} g''_{\mathsf{l}}(w) &= \mu(\mathsf{V}''_{\mathsf{c}}(w) - V''_{\mathsf{l}}(w+\beta)) + \mu\beta \mathsf{V}'''_{\mathsf{c}}(w) \\ &= \mu(\mathsf{V}''_{\mathsf{c}}(w) - \mathcal{V}''_{\hat{\mathbf{w}}}(w+\beta)) + \mu\beta \mathsf{V}'''_{\mathsf{c}}(w) \text{ need } w + \beta > \underline{\vartheta} \text{ here} \\ &= \mu\big[(\mathsf{V}''_{\mathsf{c}}(w) - \mathsf{V}''_{\mathsf{c}}(w+\beta)) + \beta \mathsf{V}'''_{\mathsf{c}}(w)\big] + \mu\big(\mathsf{V}''_{\mathsf{c}}(w+\beta) - \mathcal{V}''_{\hat{\mathbf{w}}}(w+\beta)\big) \\ &> \mu\big(\mathsf{V}''_{\mathsf{c}}(w+\beta) - \mathcal{V}''_{\hat{\mathbf{w}}}(w+\beta)\big) \\ &= \mu(\mathsf{V}''_{\mathsf{c}}(\bar{\vartheta}) - \mathcal{V}''_{\hat{\mathbf{w}}}(\bar{\vartheta})) + \mu \int_{\bar{\vartheta}}^{w+\beta} (\mathsf{V}'''_{\mathsf{c}}(y) - \mathcal{V}'''_{\hat{\mathbf{w}}}(y)) \mathrm{d}y \\ &> 0, \end{aligned}$$

where the first inequality follows from (105), the last equality follows from that $\mathcal{V}_{\hat{\mathbf{w}}} \in C^3(\mathbb{R}_+ \backslash \{\hat{\mathbf{w}}, \hat{\mathbf{w}} - \beta, \breve{w}(\hat{\mathbf{w}})\})$ and the second inequality follows from $\mathsf{V}_{\mathsf{c}}''(\bar{\vartheta}) - \mathcal{V}_{\hat{\mathbf{w}}}''(\bar{\vartheta}) = -\phi'(\bar{\vartheta}) \geqslant 0$, $w + \beta > \bar{\vartheta}$, and $\mathsf{V}_{\mathsf{c}}''' > 0$ (see (102)), $\mathcal{V}_{\hat{\mathbf{w}}}''' < 0$ over $(\breve{w}(\hat{\mathbf{w}}), w + \beta)$. The proof is complete. ▫

*Proof of Lemma 17.* Following from (100), we have

$$\begin{aligned} g_{\mathsf{l}}''(w) &= \mathsf{c}r/\rho \cdot (r/\rho - 1) w^{r/\rho - 3}[\mu w + \rho \bar{w}(r/\rho - 2)] - \mu V_{\mathsf{l}}''(w + \beta) \\ &= \mathsf{c}r/\rho \cdot (r/\rho - 1) w^{r/\rho - 3}\mu[w + \beta(r/\rho - 2)] - \mu V_{\mathsf{l}}''(w + \beta) \\ &\geqslant -\mu V_{\mathsf{l}}''(w + \beta) \geqslant 0, \end{aligned}$$

where the first inequality follows from $w \leqslant (2 - r/\rho)\beta$ and the last inequality from the concavity of $V_{\mathsf{l}}$. ▫

*Proof of Lemma 18.* We prove this result by contradiction. Suppose there exists a $w^\dagger \in [(1 - r/\rho)\beta, \underline{\vartheta})$ such that $\mathcal{V}_{\hat{\mathbf{w}}}'''(w^\dagger + \beta) > 0$ and $g_{\mathsf{l}}'(w^\dagger) > 0$. Following Lemma 13, we know that there must exist a number $w^\ddagger \in (w^\dagger, \underline{\vartheta})$ such that

$$g_{\mathsf{l}}'(w^\ddagger) = 0 \text{ and } g_{\mathsf{l}}''(w^\ddagger) \leqslant 0. \tag{106}$$

From Lemma 19, we have $\mathcal{V}_{\hat{\mathbf{w}}}'''(w^\ddagger + \beta) > 0$. Furthermore, (85) at $w^\ddagger + \beta$ gives

$$(\mu + r)\mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta) - \mu \mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + 2\beta) + \rho(\bar{w} - w^\ddagger - \beta)\mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta) - \rho \mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta) + \rho - r = 0.$$

Since $\mathcal{V}_{\hat{\mathbf{w}}}'''(w^\ddagger + \beta) > 0$, we have $\mathcal{V}_{\hat{\mathbf{w}}}''' > 0$ over $[w^\ddagger + \beta, \hat{\mathbf{w}})$ in view of Lemma 19. According to the Lagrange's mean value theorem, $\mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + 2\beta) - \mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta) = \beta \mathcal{V}_{\hat{\mathbf{w}}}''(w^b)$ for some $w^b \in (w^\ddagger + \beta, w^\ddagger + 2\beta)$. Hence, we have

$$\begin{aligned} \rho(\bar{w} - w^\ddagger - \beta)\mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta) &= (\rho - r)\big(\mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta) - 1\big) + \mu\big(\mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + 2\beta) - \mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta)\big) \\ &= (\rho - r)\big(\mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta) - 1\big) + \mu\beta \mathcal{V}_{\hat{\mathbf{w}}}''(w^b) \\ &> (\rho - r)\big(\mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta) - 1\big) + \mu\beta \mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta), \end{aligned}$$

where the inequality follows from $\mathcal{V}_{\hat{\mathbf{w}}}''' > 0$ over $[w^\ddagger + \beta, w^b)$. Using $\rho\bar{w} = \mu\beta$, the above inequality can be rewritten as

$$(\rho - r)(1 - \mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta)) > \rho(w^\ddagger + \beta)\mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta). \tag{107}$$

Since $g_{\mathsf{l}}'(w^\ddagger) = 0$, using (99) we have

$$1 - \mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta) = -\mathsf{c}r/\rho \cdot (w^\ddagger)^{r/\rho - 2}\left[w^\ddagger - (1 - r/\rho)\beta\right]. \tag{108}$$

Following (100), we have

$$\begin{aligned} g_{\mathsf{l}}''(w^\ddagger)/\mu &= \mathsf{c}r/\rho \cdot (r/\rho - 1)(w^\ddagger)^{r/\rho - 3}[w^\ddagger + \beta(r/\rho - 2)] - \mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta) \\ &= \frac{(\rho - r)(1 - \mathcal{V}_{\hat{\mathbf{w}}}'(w^\ddagger + \beta))}{w^\ddagger - (1 - r/\rho)\beta} \cdot \frac{w^\ddagger + \beta(r/\rho - 2)}{\rho w^\ddagger} - \mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta) \\ &> \left[\frac{\rho(w^\ddagger + \beta)[w^\ddagger + \beta(r/\rho - 2)]}{(w^\ddagger - (1 - r/\rho)\beta) \cdot \rho w^\ddagger} - 1\right]\mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta) \\ &= -\frac{(2\rho - r)\beta^2}{(w^\ddagger - (1 - r/\rho)\beta) \cdot \rho w^\ddagger}\mathcal{V}_{\hat{\mathbf{w}}}''(w^\ddagger + \beta) > 0, \end{aligned}$$

where the second equality follows from (108), and the first inequality follows from (107) and uses the fact that $w^\ddagger > w^\dagger \geqslant (1 - r/\rho)\beta$. This reaches a contradiction with (106). ▫

*Proof of Lemma 19.* Using (36) in Cao et al. (2021), we have

$$\mathcal{V}'''_{\hat{\mathbf{w}}}(w) = -b_{\hat{\mathbf{w}}}\frac{r+\mu}{\rho}\cdot\frac{r+\mu-\rho}{\rho}\cdot\frac{r+\mu-2\rho}{\rho}(\bar{w}-w)^{(r+\mu-2\rho)/\rho} \tag{109}$$

for $w\in(\hat{\mathbf{w}}-\beta,\hat{\mathbf{w}})$ with $b_{\hat{\mathbf{w}}} := \frac{r-\rho}{r+\mu-\rho}\cdot\frac{\rho}{r+\mu}(\bar{w}-\hat{\mathbf{w}})^{(\rho-r-\mu)/\rho} < 0$.

Besides, for $w\in(\breve{w}(\hat{\mathbf{w}}),\hat{\mathbf{w}}-\beta)$, we have

$$\rho(\bar{w}-w)\mathcal{V}'''_{\hat{\mathbf{w}}}(w) = \mu(\mathcal{V}''_{\hat{\mathbf{w}}}(w+\beta)-\mathcal{V}''_{\hat{\mathbf{w}}}(w)) + (2\rho-r)\mathcal{V}''_{\hat{\mathbf{w}}}(w) \tag{110}$$

and thus

$$\rho(\bar{w}-w)\mathcal{V}''''_{\hat{\mathbf{w}}}(w) = \mu(\mathcal{V}'''_{\hat{\mathbf{w}}}(w+\beta)-\mathcal{V}'''_{\hat{\mathbf{w}}}(w)) + (3\rho-r)\mathcal{V}'''_{\hat{\mathbf{w}}}(w). \tag{111}$$

We mention that the left-3rd-order-derivative of the function $\mathcal{V}_{\hat{\mathbf{w}}}$ at $\hat{\mathbf{w}}-\beta$ is not equal to its right-3rd-order-derivative. In fact, the right-3rd-order-derivative$\mathcal{V}'''_{\hat{\mathbf{w}}}((\hat{\mathbf{w}}-\beta)+)$ can be obtained from (109) by replacing $w$ with $\hat{\mathbf{w}}-\beta$. Moreover, it follows from (110) at $\hat{\mathbf{w}}-\beta$ that the left-3rd-order-derivative is given by

$$\begin{aligned}\mathcal{V}'''_{\hat{\mathbf{w}}}((\hat{\mathbf{w}}-\beta)-) &= \mathcal{V}'''_{\hat{\mathbf{w}}}((\hat{\mathbf{w}}-\beta)+) + \frac{\mu\mathcal{V}''_{\hat{\mathbf{w}}}(\hat{\mathbf{w}}-)}{\rho(\bar{w}-(\hat{\mathbf{w}}-\beta))}\\ &= \mathcal{V}'''_{\hat{\mathbf{w}}}((\hat{\mathbf{w}}-\beta)+) - \frac{(\rho-r)\mu}{\rho^2(\bar{w}-(\hat{\mathbf{w}}-\beta))(\bar{w}-\hat{\mathbf{w}})} < \mathcal{V}'''_{\hat{\mathbf{w}}}((\hat{\mathbf{w}}-\beta)+).\end{aligned} \tag{112}$$

We proceed to prove Lemma 19. First, we consider the case that $2\rho < r+\mu$ or $\rho > r+\mu$. Hence, (109) implies that $\mathcal{V}'''_{\hat{\mathbf{w}}} > 0$ over $(\hat{\mathbf{w}}-\beta,\hat{\mathbf{w}})$. (If $\hat{\mathbf{w}}-\beta \leqslant \breve{w}(\hat{\mathbf{w}})$, the desired result is obtained by setting $\varsigma = \breve{w}(\hat{\mathbf{w}})$.) From (112), $\mathcal{V}'''_{\hat{\mathbf{w}}}(\hat{\mathbf{w}}-\beta)$ may not be larger than 0 (recall the convention that we use $\mathcal{V}'''_{\hat{\mathbf{w}}}$ to denote the left-3rd-order-derivative of function $\mathcal{V}_{\hat{\mathbf{w}}}$), which motivates us to consider the following two subcases.

*Subcase 1:* $\mathcal{V}'''_{\hat{\mathbf{w}}}(\hat{\mathbf{w}}-\beta) > 0$. Let $w^c := \sup\{w\in(\breve{w}(\hat{\mathbf{w}}),\hat{\mathbf{w}}-\beta) : \mathcal{V}'''_{\hat{\mathbf{w}}}(w)\leqslant 0\}$. If the set is empty, we set $w^c = \breve{w}(\hat{\mathbf{w}})$, in which case we have $\mathcal{V}'''_{\hat{\mathbf{w}}} > 0$ over $(\breve{w}(\hat{\mathbf{w}}),\hat{\mathbf{w}})$. Consequently, the desired result is obtained with $\varsigma = \breve{w}(\hat{\mathbf{w}})$.

If the set is nonempty, then $\breve{w}(\hat{\mathbf{w}}) < w^c < \hat{\mathbf{w}}-\beta$. Since $\mathcal{V}_{\hat{\mathbf{w}}}\in C^3((\breve{w}(\hat{\mathbf{w}}),\hat{\mathbf{w}}-\beta))$, we have $\mathcal{V}'''_{\hat{\mathbf{w}}}(w^c) = 0$ and $\mathcal{V}'''_{\hat{\mathbf{w}}} > 0$ over $(w^c,\hat{\mathbf{w}})$. Now we prove that

$$\mathcal{V}'''_{\hat{\mathbf{w}}} < 0 \text{ over } (\breve{w}(\hat{\mathbf{w}}),w^c). \tag{113}$$

It follows from (111) at $w^c$ and $\mathcal{V}'''_{\hat{\mathbf{w}}}(w^c) = 0$ that

$$\rho(\bar{w}-w^c)\mathcal{V}''''_{\hat{\mathbf{w}}}(w^c) = \mu\mathcal{V}'''_{\hat{\mathbf{w}}}(w^c+\beta) > 0,$$

which implies that $\mathcal{V}'''_{\hat{\mathbf{w}}} < 0$ over $(w^c-\epsilon,w^c)$ for some small $\epsilon > 0$.

If (113) fails to hold, then $w^d := \sup\{w\in(\breve{w}(\hat{\mathbf{w}}),w^c) : \mathcal{V}'''_{\hat{\mathbf{w}}}(w)\geqslant 0\}$ is well-defined and $w^d\in(\breve{w}(\hat{\mathbf{w}}),w^c)$. Hence, we have $\mathcal{V}'''_{\hat{\mathbf{w}}}(w^d) = 0$ and $\mathcal{V}''''_{\hat{\mathbf{w}}}(w^d) < 0$. It follows from (111) at $w^d$ that $\rho(\bar{w}-w^d)\mathcal{V}''''_{\hat{\mathbf{w}}}(w^d) = \mu\mathcal{V}'''_{\hat{\mathbf{w}}}(w^d+\beta) < 0$. By the definition of $w^c$, we have $w^d+\beta < w^c$. Hence, $\mathcal{V}'''_{\hat{\mathbf{w}}} < 0$ over $(w^d,w^d+\beta]$ and thus $\mathcal{V}''_{\hat{\mathbf{w}}}(w^d+\beta) < \mathcal{V}''_{\hat{\mathbf{w}}}(w^d)$. Furthermore, it follows from (110) at $w^d$ that

$$\mu\mathcal{V}''_{\hat{\mathbf{w}}}(w^d+\beta) = (\mu+r-2\rho)\mathcal{V}''_{\hat{\mathbf{w}}}(w^d),$$

which contradicts with $\mathcal{V}''_{\hat{\mathbf{w}}}(w^d+\beta) < \mathcal{V}''_{\hat{\mathbf{w}}}(w^d) < 0$, under either $2\rho < \mu+r$ or $\rho > r+\mu$. Hence, (113) holds in this case. That is, the desired result is obtained by letting $\varsigma = w^c$.

*Subcase 2:* $\mathcal{V}_{\hat{\mathbf{w}}}'''(\hat{\mathbf{w}}-\beta)\leqslant 0$. We show that

$$\mathcal{V}_{\hat{\mathbf{w}}}'''<0 \text{ over } (\breve{w}(\hat{\mathbf{w}}),\hat{\mathbf{w}}-\beta). \tag{114}$$

Note that if $\mathcal{V}_{\hat{\mathbf{w}}}'''(\hat{\mathbf{w}}-\beta)=0$, then $\mathcal{V}_{\hat{\mathbf{w}}}''''(\hat{\mathbf{w}}-\beta)>0$ in view of (111). Hence, we have $\mathcal{V}_{\hat{\mathbf{w}}}'''<0$ over $(\hat{\mathbf{w}}-\beta-\epsilon,\hat{\mathbf{w}}-\beta)$ for some $\epsilon>0$. If (114) fails to hold, then $w^e:=\sup\{w\in[\breve{w}(\hat{\mathbf{w}}),\hat{\mathbf{w}}-\beta):\mathcal{V}_{\hat{\mathbf{w}}}'''(w)\geqslant 0\}$ is well-defined and $w^e\in[\breve{w}(\hat{\mathbf{w}}),\hat{\mathbf{w}}-\beta)$. Moreover, we have $\mathcal{V}_{\hat{\mathbf{w}}}'''(w^e)=0$ and $\mathcal{V}_{\hat{\mathbf{w}}}''''(w^e)<0$. Hence, it follows from (111) at $w^e$ that $\rho(\bar{w}-w^e)\mathcal{V}_{\hat{\mathbf{w}}}''''(w^e)=\mu\mathcal{V}_{\hat{\mathbf{w}}}'''(w^e+\beta)<0$. Since $\mathcal{V}_{\hat{\mathbf{w}}}'''>0$ over $[\hat{\mathbf{w}}-\beta,\hat{\mathbf{w}})$, we have $w^e+\beta\leqslant\hat{\mathbf{w}}-\beta$. Hence, $\mathcal{V}_{\hat{\mathbf{w}}}'''<0$ over $(w^e,w^e+\beta)$, which gives $\mathcal{V}_{\hat{\mathbf{w}}}''(w^e+\beta)<\mathcal{V}_{\hat{\mathbf{w}}}''(w^e)$. Furthermore, it follows from (110) at $w^e$ that

$$\mu\mathcal{V}_{\hat{\mathbf{w}}}''(w^e+\beta)=(\mu+r-2\rho)\mathcal{V}_{\hat{\mathbf{w}}}''(w^e),$$

which contradicts with $\mathcal{V}_{\hat{\mathbf{w}}}''(w^e+\beta)<\mathcal{V}_{\hat{\mathbf{w}}}''(w^e)<0$. Hence, (114) holds. The desired result is obtained by noting (109) and letting $\varsigma=\hat{\mathbf{w}}-\beta$.

Next, we turn to study the case that $\rho<r+\mu<2\rho$. In this case, (109) implies that $\mathcal{V}_{\hat{\mathbf{w}}}'''<0$ over $(\hat{\mathbf{w}}-\beta,\hat{\mathbf{w}})$. Hence, we must have $\mathcal{V}_{\hat{\mathbf{w}}}'''(\hat{\mathbf{w}}-\beta)<0$ in view of (112). The argument for Subcase 2 is valid, which leads us to the desired result. ◻

### B.4. Proofs of the Results in Section 5

#### B.4.1. Proofs of the Results in Section 5.1

*Proof of Theorem 5.* The proof consists of three parts.

*Part 1.* First, we show that under Condition 2 and (H2), functions $V_{\mathsf{I}}(w)$ as defined in (39) and $V_{\emptyset}(w)=\underline{v}$ satisfy the optimality condition (20)–(22). Note that the first inequality in Condition 2 implies $\bar{w}<\beta$. If $w\in[0,\bar{w}]$, then we have

$$\begin{aligned}(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w)=&(\mu+r)V_{\mathsf{I}}(w)-\mu V_{\mathsf{I}}(w+\beta)+\rho(\bar{w}-w)V_{\mathsf{I}}'(w)-(\mu R-c)+(\rho-r)w\\ =&(\mu+r)\left(\underline{v}+\frac{\bar{V}-\underline{v}}{\bar{w}}w\right)-\mu\bar{V}+\rho(\bar{w}-w)\frac{\bar{V}-\underline{v}}{\bar{w}}-(\mu R-c)+(\rho-r)w\\ =&(\bar{w}-w)\left[\frac{\bar{V}-\underline{v}}{\bar{w}}(\rho-r-\mu)-(\rho-r)\right]\geqslant 0.\end{aligned}$$

Here, it is worth mentioning that although at $\bar{w}$, $V_{\mathsf{I}}$ is not differentiable, its left-derivative exists and is $(\bar{V}-\underline{v})/\bar{w}$.

If $w\in(\bar{w},\infty)$, then

$$\begin{aligned}(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w)&=(\mu+r)\bar{V}-\mu\bar{V}-(\mu R-c)+(\rho-r)w\\ &=(\rho-r)(w-\bar{w})>0.\end{aligned}$$

Combining the above two cases yields $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w)\geqslant 0$ for any $w\in\mathbb{R}_+$. Besides, for any $w\in\mathbb{R}_+$, $(\mathcal{A}_{\emptyset}V_{\emptyset})(w)=(\rho-r)w\geqslant 0$. Hence, (20) holds.

It is straightforward to see that $V_{\mathsf{I}}(w)-V_{\emptyset}(w)\geqslant 0$, and $V_{\emptyset}(w)=\underline{v}\geqslant\bar{V}-K\geqslant V_{\mathsf{I}}(w)-K$. Hence, (21) holds. Obviously, $V_{\mathsf{I}}(0)=V_{\emptyset}(0)=\underline{v}$, which implies (22).

*Part 2.* Second, we show that under Condition 2 and (M2), functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ as defined in (39) and (40), respectively, satisfy (20)–(22). According to the proof for the case under Condition 2 and (H2), we have $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) \geqslant 0$ for any $w \in \mathbb{R}_+$. Below we show that $(\mathcal{A}_{\emptyset}V_{\emptyset})(w) \geqslant 0$ for all $w \in \mathbb{R}_+$.

If $w \in [0, \bar{w}]$, then

$$\begin{aligned}(\mathcal{A}_{\emptyset}V_{\emptyset})(w) &= rV_{\emptyset}(w) - \rho w V_{\emptyset}'(w) + (\rho - r)w - R\underline{\mu} \\ &= (\rho - r)\left(1 - \frac{\bar{V} - \underline{v} - K}{\bar{w}}\right) w \geqslant 0.\end{aligned}$$

Here, we mention that although at $\bar{w}$, $V_{\emptyset}$ is not differentiable, its left-derivative exists and is $(\bar{V} - \underline{v} - K)/\bar{w}$.

If $w \in (\bar{w}, \infty)$, then we have

$$\begin{aligned}(\mathcal{A}_{\emptyset}V_{\emptyset})(w) &= rV_{\emptyset}(w) - \rho w V_{\emptyset}'(w) + (\rho - r)w - R\underline{\mu} \\ &= r(\bar{V} - K) + (\rho - r)w - R\underline{\mu} > r(\bar{V} - K - \underline{v}) > 0.\end{aligned}$$

Therefore, (20) holds. Note that $V_{\mathsf{I}}(w) - V_{\emptyset}(w) = K$ for $w \in [\bar{w}, \infty)$ and $V_{\mathsf{I}}(w) - V_{\emptyset}(w) = K/\bar{w} \cdot w$ for $w \in [0, \bar{w})$. Hence, (21) holds. Besides, $V_{\mathsf{I}}(0) = V_{\emptyset}(0) = \underline{v}$ and thus (22) holds.

*Part 3.* Next, we show (41) and (42). First, by the definition of $\bar{\Gamma}$ as in (15) and the definition of $\underline{\Gamma}$ as in (16), we have $U(\underline{\Gamma}, \emptyset) = \underline{v}$ and $U(\bar{\Gamma}, \mathsf{I}) = \bar{V} - \bar{w}$. The remaining inequalities can be easily verified using (39), (40) and (H2) (or (M2)). ▫

### B.4.2. Proofs of the Results in Section 5.2

*Proof of Lemma 5.* For any $\underline{\theta} \in [0, \bar{w}]$, it is straightforward to verify that

$$C^1(\underline{\theta}) := \frac{\bar{V} - \underline{v} - \bar{w}}{r/\rho \cdot \underline{\theta}^{r/\rho - 1}[(\rho/r - 1)\underline{\theta} + \bar{w}]} \text{ and } m(\underline{\theta}) := \frac{(\rho/r - 1)\underline{\theta} + \bar{V} - \underline{v}}{(\rho/r - 1)\underline{\theta} + \bar{w}} \tag{115}$$

satisfies (43) and (44), with $\underline{\theta}$ replacing $\underline{\theta}_K$, $C^1(\underline{\theta})$ replacing $c_K$ and $m(\underline{\theta})$ replacing $m_K$. Moreover, it follows from Conditions 2 and (L2) that $C^1(\underline{\theta}) > 0$. Since

$$\frac{\mathrm{d}}{\mathrm{d}x}\left(x^{r/\rho - 1}[(\rho/r - 1)x + \bar{w}]\right) = (1 - r/\rho)x^{r/\rho - 2}(x - \bar{w}) < 0$$

for $x \in (0, \bar{w})$, $C^1(\underline{\theta})$ is increasing in $\underline{\theta}$ on $[0, \bar{w}]$. It is evident that $m(\underline{\theta})$ is strictly decreasing in $\underline{\theta}$ on $[0, \bar{w}]$.

We have the following result, which is stated as a lemma for ease of reference. Its proof is provided later.

Lemma 20. *Under Conditions 2 and (L2), function $\psi_1(\underline{\theta})$, defined by*

$$\psi_1(\underline{\theta}) = \bar{V} - \underline{v} - \bar{w} - C^1(\underline{\theta}) \cdot (\bar{w})^{r/\rho},$$

*is continuous and decreasing in $\underline{\theta}$ on $[0, \bar{w}]$. Moreover, $\psi_1(\bar{w}) = 0$ and $\psi_1(0) = \bar{V} - \underline{v} - \bar{w} > K$. Consequently, there exists $\underline{\theta}_K \in (0, \bar{w})$ such that $\psi_1(\underline{\theta}_K) = K$. Furthermore, $\underline{\theta}_K$ is decreasing in $K$ with $\lim_{K \downarrow 0} \underline{\theta}_K = \bar{w}$.*

Lemma 20 immediately implies that $(\underline{\theta}_K, c_K, m_K)$ with $c_K = C^1(\underline{\theta}_K)$, $m_K = m(\underline{\theta}_K)$ satisfies (43)–(45). The monotonicity of $\underline{\theta}_K$ has already been obtained in Lemma 20. The monotonicity of $c_K$ and $m_K$ in $K$ follows from that of $C^1(\underline{\theta})$ and $m(\underline{\theta})$ in $\underline{\theta}$. ▫

*Proof of Lemma 20.* Since $C^1(\underline{\theta})$ is continuous and increasing in $\underline{\theta}$ on $[0,\bar{w}]$, $\psi_1(\underline{\theta})$ is continuous and decreasing in $\underline{\theta}$ on $[0,\bar{w}]$. Note that $C^1(0)=0$. Hence, $\psi_1(0)=\bar{V}-\underline{v}-\bar{w}>K$, which follows from (L2). Besides, $C^1(\bar{w})=(\bar{V}-\underline{v}-\bar{w})\bar{w}^{-r/\rho}$, which implies that $\psi_1(\bar{w})=0<K$. Consequently, there exists $\underline{\theta}_K\in(0,\bar{w})$ such that $\psi_1(\underline{\theta}_K)=K$. This, together with the fact that $\psi_1(\underline{\theta})$ is decreasing in $\underline{\theta}$ on $[0,\bar{w}]$, yields that $\underline{\theta}_K$ is decreasing in $K$. $\lim_{K\downarrow 0}\underline{\theta}_K=\bar{w}$ follows by the continuity of $\psi_1(\cdot)$ and $\psi_1(\bar{w})=0$. ◻

*Proof of Theorem 6.* The proof consists of two parts.

*Part 1.* First, we show that under Condition 2, (L2), and (mH), functions $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$, as defined in (47) and (48), respectively, satisfy the optimality condition (20)–(22).

Obviously, (22) holds by noting that $V_{\mathsf{I}}(0)=V_{\emptyset}(0)=\underline{v}$. We proceed to verify that $V_{\mathsf{I}}(w)$ and $V_{\emptyset}(w)$ satisfy (20) and (21).

We show that $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w)\geqslant 0$ for all $w\in\mathbb{R}_+$ by considering the following cases.

*Case 1:* $w\in[\underline{\theta}_K,\bar{w}]$. We have

$$
\begin{aligned}
(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) &= (\mu+r)(\bar{V}+m_K\cdot(w-\bar{w}))-\mu\bar{V}+\rho(\bar{w}-w)\cdot m_K-(\mu R-c)+(\rho-r)w\\
&=(\bar{w}-w)[m_K\cdot(\rho-r-\mu)-(\rho-r)]\geqslant 0.
\end{aligned}
$$

Here, we mention that although at $\bar{w}$, $V_{\mathsf{I}}$ is not differentiable, its left-derivative exists and is $m_K$.

*Case 2:* $w\in[0,\underline{\theta}_K)$. We have

$$
\begin{aligned}
(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w) &= (\mu+r)V_{\mathsf{I}}(w)-\mu V_{\mathsf{I}}(w+\beta)+\rho(\bar{w}-w)V_{\mathsf{I}}'(w)-(\mu R-c)+(\rho-r)w\\
&=\mu V_{\mathsf{I}}(w)-\mu V_{\mathsf{I}}(w+\beta)+\rho\bar{w}V_{\mathsf{I}}'(w)-\Delta\mu\cdot R+c\\
&=\mu(\underline{v}+w+c_Kw^{r/\rho})-\mu\bar{V}+\mu\beta(1+c_Kw^{r/\rho-1}r/\rho)-\Delta\mu\cdot R+c=:g_{\mathsf{I}}(w),
\end{aligned}
$$

where the second equality follows from $(\mathcal{A}_{\emptyset}V_{\mathsf{I}})(w)=0$ on $[0,\underline{\theta}_K)$, and the third equality follows from $\beta>\mu\beta/\rho=\bar{w}$ due to Condition 2.

Since $V_{\mathsf{I}}$ is continuously differentiable on $[0,\bar{w})$, $(\mathcal{A}_{\mathsf{I}}V_{\mathsf{I}})(w)$ is also continuous in $w$ on $[0,\bar{w})$, which implies that $g_{\mathsf{I}}(\underline{\theta}_K)\geqslant 0$. Hence, it suffices to show that $g_{\mathsf{I}}(w)$ is decreasing in $w$. Note that

$$
g_{\mathsf{I}}'(w)=\mu+\mu r/\rho\cdot c_Kw^{r/\rho-2}\big(w+(r-\rho)/\rho\cdot\beta\big),
$$

and

$$
\begin{aligned}
g_{\mathsf{I}}'(\underline{\theta}_K) &= \mu+\mu r/\rho\cdot c_K\cdot(\underline{\theta}_K)^{r/\rho-2}\big(\underline{\theta}_K+(r-\rho)/\rho\cdot\beta\big)\\
&=\mu+\mu\cdot\frac{m_K-1}{\underline{\theta}_K}\cdot\big(\underline{\theta}_K+(r-\rho)/\rho\cdot\beta\big)\\
&<\mu+\mu\left(\frac{\rho-r}{\rho-r-\mu}-1\right)\left(1+\frac{(r-\rho)\cdot\beta}{\rho\underline{\theta}_K}\right)\\
&<\mu+\mu\frac{\mu}{\rho-r-\mu}\left(1+\frac{(r-\rho)\cdot\beta}{\rho\bar{w}}\right)=0,
\end{aligned}
$$

where the second equality follows from (44), the first inequality follows from (mH) and the fact that $\underline{\theta}_K + (r-\rho)/\rho \cdot \beta < \bar{w} + (r-\rho)/\rho \cdot \beta = (\mu + r - \rho)/\rho \cdot \beta < 0$, and the last equality follows from $\bar{w} = \mu\beta/\rho$. Besides, we have

$$\begin{aligned} g_{\mathsf{I}}''(w) &= r/\rho(r/\rho - 1) \cdot c_K w^{r/\rho - 3}[\mu w + \rho\bar{w}(r/\rho - 2)] \\ &\geqslant r/\rho(r/\rho - 1) \cdot c_K w^{r/\rho - 3}[\mu\bar{w} + \rho\bar{w}(r/\rho - 2)] \\ &= r/\rho(r/\rho - 1) \cdot c_K w^{r/\rho - 3}(\mu + r - 2\rho)\bar{w} > 0, \end{aligned}$$

where the last inequality follows from $\rho > r + \mu$. Therefore, $g_{\mathsf{I}}'(w) < 0$ for $w \in [0, \underline{\theta}_K]$.

*Case 3: $w \in (\bar{w}, \infty)$.* We have

$$\begin{aligned} (\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) &= (\mu + r)\bar{V} - \mu\bar{V} - (\mu R - c) + (\rho - r)w \\ &= (\rho - r)(w - \bar{w}) > 0. \end{aligned}$$

Combining the above three cases yields $(\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}})(w) \geqslant 0$ for any $w \in \mathbb{R}_+$.

Next, we establish $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) \geqslant 0$ for all $w \in \mathbb{R}_+$. Obviously, we have $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) = 0$ for $w \in [0, \bar{w}]$. (We mention that $V_{\emptyset}$ is not differentiable at $\bar{w}$, but its left-derivative exists.) If $w \in (\bar{w}, \infty)$, then

$$\begin{aligned} (\mathcal{A}_{\emptyset} V_{\emptyset})(w) &= r(\bar{V} - K) + (\rho - r)w - R\underline{\mu} \\ &> r(\bar{V} - K) - R\underline{\mu} = r(\bar{V} - K - \underline{v}) > 0. \end{aligned}$$

Hence, (20) is proved.

Below we establish (21). If $w \in [0, \underline{\theta}_K]$, we have $V_{\mathsf{I}}(w) - V_{\emptyset}(w) = 0$, and if $w \in [\bar{w}, \infty)$, we have $V_{\mathsf{I}}(w) - V_{\emptyset}(w) = K$. If $w \in (\underline{\theta}_K, \bar{w})$, we have

$$V_{\mathsf{I}}'(w) - V_{\emptyset}'(w) = m_K - V_{\emptyset}'(w) \geqslant m_K - V_{\emptyset}'(\underline{\theta}_K) = 0,$$

which implies that $V_{\mathsf{I}} - V_{\emptyset}$ is increasing on $[\underline{\theta}_K, \bar{w}]$. Consequently, we have $0 \leqslant V_{\mathsf{I}}(w) - V_{\emptyset}(w) \leqslant K$ for $w \in (\underline{\theta}_K, \bar{w})$.

*Part 2.* Next, we show (49) and (50). By the definition of contract $\bar{\Gamma}$, it is clear that $U(\bar{\Gamma}, \mathsf{I}) = \bar{V} - \bar{w}$ and $U(\bar{\Gamma}, \emptyset) = \bar{V} - \bar{w} - K$. The remaining inequalities hold by noting that $\max_{w \geqslant 0}\{V_{\mathsf{I}}(w) - w\} = V_{\mathsf{I}}(\bar{w}) - \bar{w}$ under (mH), and $\max_{w \geqslant 0}\{V_{\emptyset}(w) - w\} = V_{\emptyset}(\bar{w}) - \bar{w}$. ▫

*Proof of Theorem 7.* We only show that Proposition 4 hold under Conditions 2, (L2), and (mL), as the proofs of Proposition 5 and Theorem 4 only rely on Proposition 4 and thus hold naturally. Since most arguments are exactly the same as those for Proposition 4 under Condition 1, we only provide a sketch here. To start, we observe that $\breve{\underline{\theta}} := \frac{(\bar{V} - \underline{v})(\rho - r - \mu) - (\rho - r)\bar{w}}{\mu(\rho/r - 1)} \in (0, \underline{\theta}_K)$ satisfies $m(\breve{\underline{\theta}}) = \frac{\rho - r}{\rho - r - \mu}$ by (115). Moreover, we have

$$\bar{V} - \underline{v} - \bar{w} - C^1(\breve{\underline{\theta}}) \cdot \bar{w}^{r/\rho - 1} > K \tag{116}$$

since $\psi_1(\breve{\underline{\theta}}) > \psi_1(\underline{\theta}_K) = 0$.

Note that under Condition 2, $\breve{w}(\tilde{w}) = 0$, which follows from property (i) in Lemma 2. Hence, we will use $V_{\tilde{w}}$ instead of $\mathcal{V}_{\tilde{w}}$ in the proof. Next, we will show the desired result by the following four lemmas, which parallel Lemmas 7–10 used to prove Proposition 4. These lemmas' proofs are provided at the end of this section.

Lemma 21. *For any* $\underline{\theta} \in (\breve{\underline{\theta}}, \bar{w})$*, there exists unique values* $\widetilde{w}(\underline{\theta}) \in (\underline{\theta}, \bar{w})$ *and* $C(\underline{\theta})$*, in place of* $\tilde{\mathrm{w}}$ *and* $\mathsf{c}$*, such that* (30)–(31) *are satisfied at* $\underline{\vartheta} = \underline{\theta}$*.*

Lemma 22. *Value* $\underline{\theta}^0 := \inf\{\underline{\theta} \in (\breve{\underline{\theta}}, \bar{w}) : \widetilde{w}'(\underline{\theta}^0) \geqslant 0\}$ *is well-defined. We have* $\widetilde{w}(\underline{\theta})$ *is strictly decreasing in* $\underline{\theta}$*, and* $C(\underline{\theta})$ *is strictly increasing in* $\underline{\theta}$ *on* $(\breve{\underline{\theta}}, \underline{\theta}^0)$ *with* $\widetilde{w}'(\underline{\theta}^0) = 0$*. Moreover,* $C(\underline{\theta}) > 0$ *for any* $\underline{\theta} \in (\breve{\underline{\theta}}, \underline{\theta}^0)$*.*

Lemma 23. *For any* $\underline{\theta} \in (\breve{\underline{\theta}}, \underline{\theta}^0)$*, the threshold* $\bar{\theta}(\underline{\theta})$

$$\bar{\theta}(\underline{\theta}) := \inf\left\{w > \underline{\theta} : V'_{\widetilde{w}(\underline{\theta})}(w) \leqslant 1 + C(\underline{\theta}) r/\rho \cdot w^{r/\rho - 1}\right\}$$

*is well-defined. As a function of* $\underline{\theta}$*, threshold* $\bar{\theta}(\underline{\theta})$ *is decreasing in* $\underline{\theta}$ *on* $[0, \underline{\theta}^0)$*;* $\lim_{\underline{\theta} \uparrow \underline{\theta}^0} \bar{\theta}(\underline{\theta}) = \underline{\theta}^0$ *and* $\lim_{\underline{\theta} \downarrow \breve{\underline{\theta}}} \bar{\theta}(\underline{\theta}) = \bar{w}$*.*

Lemma 24. *Function* $\psi(\underline{\theta})$*, defined as in* (90)*, is continuous and decreasing in* $\underline{\theta}$ *on* $(\breve{\underline{\theta}}, \underline{\theta}^0)$*, and satisfies*

$$\lim_{\underline{\theta} \uparrow \underline{\theta}^0} \psi(\underline{\theta}) = 0, \ \text{and} \ \lim_{\underline{\theta} \downarrow 0} \psi(\underline{\theta}) > K.$$

*Consequently, there exists a value* $\underline{\vartheta} \in (\breve{\underline{\theta}}, \underline{\theta}^0)$ *such that* $\psi(\underline{\vartheta}) = K$*.*

Now we are ready to complete the proof. According to Lemmas 21–24, $(\hat{\mathbf{w}}, \mathsf{c}, \underline{\vartheta}, \bar{\vartheta})$ defined by $\hat{\mathbf{w}} = \widetilde{w}(\underline{\vartheta})$, $\mathsf{c} = C(\underline{\vartheta})$ and $\bar{\vartheta} = \bar{\theta}(\underline{\vartheta})$ satisfies (30)–(33). Besides, it follows from $\bar{\theta}(\underline{\theta}) < \widetilde{w}(\underline{\theta})$ for $\underline{\theta} \in [\breve{\underline{\theta}}, \underline{\theta}^0)$ that $\bar{\vartheta} < \hat{\mathbf{w}}$, which implies $\hat{\mathbf{w}} = \widetilde{w}(\underline{\vartheta}) < \widetilde{w}(\underline{\theta}^0) = \bar{w}$ by noting that $\widetilde{w}(\underline{\theta})$ is decreasing in $\underline{\theta}$ on $[\breve{\underline{\theta}}, \underline{\theta}^0)$. ▫

*Proof of Lemma 21.* We will use functions $C_1(\widetilde{w}, \underline{\theta})$ and $C_2(\widetilde{w}, \underline{\theta})$ defined as in the proof of Lemma 7 to derive the desired result. In the proof of Lemma 7, we established that $C_1(\widetilde{w}, \underline{\theta}) \to -\infty$ and $C_2(\widetilde{w}, \underline{\theta}) \to \infty$ as $\widetilde{w} \uparrow \bar{w}$ and thus

$$\lim_{\widetilde{w} \uparrow \bar{w}} C_1(\widetilde{w}, \underline{\theta}) < \lim_{\widetilde{w} \uparrow \bar{w}} C_2(\widetilde{w}, \underline{\theta}) \tag{117}$$

for any $\underline{\theta} \in (0, \bar{w})$. Now, we claim that (117) also holds under the conditions stated in Theorem 7 for any $\underline{\theta} \in (\breve{\underline{\theta}}, \bar{w})$. In fact, it follows from property (v) in Lemma 6 that

$$\begin{aligned}
\lim_{\widetilde{w} \uparrow \bar{w}} C_1(\widetilde{w}, \underline{\theta}) &= \left(\bar{V} - \frac{\rho - r}{\rho - r - \mu}(\bar{w} - \underline{\theta}) - \underline{v} - \underline{\theta}\right) \underline{\theta}^{-r/\rho} \\
&= \left(\bar{V} - \frac{\rho - r}{\rho - r - \mu}\bar{w} - \underline{v} + \frac{\mu}{\rho - r - \mu}\underline{\theta}\right) \underline{\theta}^{-r/\rho}, \text{ and} \\
\lim_{\widetilde{w} \uparrow \bar{w}} C_2(\widetilde{w}, \underline{\theta}) &= \frac{\rho}{r}\left(\frac{\rho - r}{\rho - r - \mu} - 1\right) \underline{\theta}^{1 - r/\rho} = \frac{\rho}{r} \frac{\mu}{\rho - r - \mu} \underline{\theta} \cdot \underline{\theta}^{-r/\rho}.
\end{aligned}$$

It is clear that

$$\frac{\lim_{\widetilde{w} \uparrow \bar{w}} C_1(\widetilde{w}, \underline{\theta})}{\lim_{\widetilde{w} \uparrow \bar{w}} C_2(\widetilde{w}, \underline{\theta})} = \frac{\bar{V} - \frac{\rho - r}{\rho - r - \mu}\bar{w} - \underline{v} + \frac{\mu}{\rho - r - \mu}\underline{\theta}}{\frac{\rho}{r}\frac{\mu}{\rho - r - \mu}\underline{\theta}}$$

is decreasing in $\underline{\theta}$ and takes value 1 at $\breve{\underline{\theta}}$. Hence, (117) holds for any $\underline{\theta} \in (\breve{\underline{\theta}}, \bar{w})$. The remaining proof is exactly the same as that for Lemma 7 and thus omitted. Moreover, we have the following by-product:

$$\widetilde{w}(\breve{\underline{\theta}}) := \lim_{\underline{\theta} \downarrow \breve{\underline{\theta}}} \widetilde{w}(\underline{\theta}) = \bar{w} \text{ and } C(\breve{\underline{\theta}}) = \frac{\rho\mu}{r(\rho - r - \mu)}(\breve{\underline{\theta}})^{1 - r/\rho}. \tag{118}$$

▫

*Proof of Lemma 22.* We only list the differences between this proof and that of Lemma 8 as follows:

(i) Show $\widetilde{w}'(\breve{\underline{\theta}}) < 0$ instead of $\widetilde{w}'(0) < 0$. This holds by noting that $h_2(\widetilde{w}(\breve{\underline{\theta}}), \breve{\underline{\theta}}) = \lim_{\widetilde{w}\uparrow\bar{w}} h_2(\widetilde{w}, \breve{\underline{\theta}}) = (\rho - r - (\rho - r)\cdot(\rho - r)/(\rho - r - \mu))/r < 0$.

(ii) Use the result $\lim_{\underline{\theta}\uparrow\bar{w}} \widetilde{w}(\underline{\theta}) = \lim_{\underline{\theta}\downarrow\breve{\underline{\theta}}} \widetilde{w}(\underline{\theta}) = \bar{w}$ instead of $\lim_{\underline{\theta}\uparrow\bar{w}} \widetilde{w}(\underline{\theta}) = \bar{w} > \hat{\mathrm{w}} = \widetilde{w}(0)$ to establish the existence of $\underline{\theta}^0$.

(iii) Use $V'_{\widetilde{w}(\breve{\underline{\theta}})}(\breve{\underline{\theta}}) = \lim_{\widetilde{w}\uparrow\bar{w}} V'_{\widetilde{w}}(\breve{\underline{\theta}}) = (\rho - r)/(\rho - r - \mu) > 1$ to characterize the monotonicity of $C(\cdot)$ near $\breve{\underline{\theta}}$, instead of using $\mathcal{V}'_{\hat{\mathrm{w}}}(0) > 1$ to characterize the monotonicity of $C(\cdot)$ near 0.

□

*Proof of Lemma 23.* The proof is the same as that for Lemma 9, with the range of $\underline{\theta}$ changed from $(0, \underline{\theta}^0)$ to $(\breve{\underline{\theta}}, \underline{\theta}^0)$. One exception is that we need to show that $\lim_{\underline{\theta}\downarrow\breve{\underline{\theta}}} \bar{\theta}(\underline{\theta}) = \bar{w}$. To show this, we first obtain that for any $w \in (\breve{\underline{\theta}}, \bar{w})$, we have

$$\begin{aligned}\lim_{\underline{\theta}\downarrow\breve{\underline{\theta}}} \Psi(w, \underline{\theta}) &= \lim_{\widetilde{w}\uparrow\bar{w}} \left\{ V'_{\widetilde{w}}(w) - 1 - C(\breve{\underline{\theta}})\cdot r/\rho \cdot w^{r/\rho - 1} \right\} \\ &= \frac{\rho - r}{\rho - r - \mu} - 1 - \frac{\rho\mu}{r(\rho - r - \mu)} (\breve{\underline{\theta}})^{1 - r/\rho} \cdot r/\rho \cdot w^{r/\rho - 1} \\ &= \frac{\mu}{\rho - r - \mu}\Big[1 - \big(\frac{w}{\breve{\underline{\theta}}}\big)^{r/\rho - 1}\Big] > 0,\end{aligned}$$

where the first equality follows from (118) and property (iii) in Lemma 6, the second equality follows from property (iv) in Lemma 6. This, combining with $\bar{\theta}(\underline{\theta}) < \widetilde{w}(\underline{\theta})$, yields that $\lim_{\underline{\theta}\downarrow\breve{\underline{\theta}}} \bar{\theta}(\underline{\theta}) = \bar{w}$. □

*Proof of Lemma 24.* The proof is exactly the same as that for Lemma 10 (also with the range of $\underline{\theta}$ changed from $(0, \underline{\theta}^0)$ to $(\breve{\underline{\theta}}, \underline{\theta}^0)$), except that we will show $\lim_{\underline{\theta}\downarrow\underline{\theta}^0} \psi(\underline{\theta}) > K$ rather than $\lim_{\underline{\theta}\downarrow 0} \psi(\underline{\theta}) > K$. In fact, we have

$$\begin{aligned}\lim_{\underline{\theta}\downarrow\underline{\theta}^0} \psi(\underline{\theta}) &= \lim_{\widetilde{w}\uparrow\bar{w}} V_{\widetilde{w}}(\bar{w}) - [\underline{v} + \bar{w} + C(\breve{\underline{\theta}})\,(\bar{w})^{r/\rho}] \\ &= \bar{V} - \underline{v} - \bar{w} - C(\breve{\underline{\theta}})\,(\bar{w})^{r/\rho} > K,\end{aligned}$$

where the second equality follows from property (v) in Lemma 6, and the inequality follows from (116). □

*Proof of Proposition 6.* It follows from (115) and $\lim_{K\downarrow 0} \underline{\theta}_K = \bar{w}$ that

$$\lim_{K\downarrow 0} c_K = \frac{\bar{V} - \underline{v} - \bar{w}}{\bar{w}^{-r/\rho}} \text{ and } \lim_{K\downarrow 0} m_K = 1 + \frac{r(\bar{V} - \underline{v} - \bar{w})}{\rho\bar{w}} = \frac{R\Delta\mu - c}{\mu\beta}.$$

It is straightforward to verify that $\lim_{K\downarrow 0} m_K \geqslant (\rho - r)/(\rho - r - \mu)$ if and only if $R \geqslant \bar{R}$, where

$$\bar{R} := \left[\frac{c}{b} + \frac{(\rho - r)\mu}{\Delta\mu(\rho - r - \mu)}\right]\beta > \hat{R}.$$

Hence, by definition of $\breve{K}_2$ and the monotonicity of $m_K$ in $K$, it is clear that $\breve{K}_2 = 0$ if and only if $R \geqslant \bar{R}$.

If $R < \bar{R}$, we have $m_{\breve{K}_2} = (\rho - r)/(\rho - r - \mu)$. Hence, it follows from (115) with $\underline{\theta} = \underline{\theta}_{\breve{K}_2}$ that

$$\underline{\theta}_{\breve{K}_2} = \frac{\bar{V} - \underline{v} - (\rho - r)/(\rho - r - \mu)\bar{w}}{\mu/(\rho - r - \mu)\cdot(\rho/r - 1)} \text{ and } c_{\breve{K}_2} = \frac{\mu}{\rho - r - \mu}\frac{\rho}{r}\underline{\theta}_{\breve{K}_2}{}^{1 - r/\rho}.$$

Substituting these values into (45) with $K = \breve{K}_2$, we obtain the following closed-form expression of $\breve{K}_2$:

$$\breve{K}_2 = \bar{V} - \underline{v} - \bar{w} - \frac{\mu}{\rho - r - \mu} \frac{\rho}{r} \left[ \frac{\bar{V} - \underline{v} - (\rho - r)/(\rho - r - \mu) \cdot \bar{w}}{\mu/(\rho - r - \mu) \cdot (\rho/r - 1)} \right]^{1 - r/\rho}. \tag{119}$$

Finally, taking the first-order derivative of $\breve{K}_2$ with respect to $R$ in (119), investigating its sign and noting by Assumption 1 that $R > c/\Delta\mu$, we obtain that $\breve{K}_2$ is decreasing in $R$ on $(c/\Delta\mu, \bar{R})$.

The statement on $\underline{K}_2$ follows directly from its definition. ▫

### B.5. Proof of the Results in Section 6

*Proof of Proposition 7.* We only consider the case under Conditions 1 and (L1), since the case under Conditions 2, (L2) and (mL) can be treated similarly. It follows from Lemma 10 that $\psi(\underline{\vartheta}) = K$ with $\psi$ being decreasing on $(0, \underline{\theta}^0)$. Consequently, $\underline{\vartheta}$ is decreasing in $K$. Recall that $\bar{\vartheta} = \bar{\theta}(\underline{\vartheta})$. It follows from part (ii) in Lemma 9 that $\bar{\vartheta} = \bar{\theta}(\underline{\vartheta})$ is increasing in $K$.

For the last assertion, we first note that under Condition 1 and $\underline{K}_1 > 0$, $\lim_{\underline{\theta} \uparrow \theta^0} \psi(\underline{\theta}) = 0$ by Lemma 10, which implies $\lim_{K \downarrow 0} \underline{\vartheta} = \underline{\theta}^0$. Then, using part (iii) in Lemma 9, we obtain $\lim_{K \downarrow 0} \bar{\vartheta} = \underline{\theta}^0$. Under Condition 2 and $\breve{K}_2 > 0$, we also have $\lim_{K \downarrow 0} \underline{\vartheta} = \underline{\theta}^0 = \lim_{K \downarrow 0} \bar{\vartheta}$, by a similar argument and using Lemmas 22 and 24. Hence, the desired result holds with $\theta_0 = \underline{\theta}^0$. ▫

*Proof of Theorem 8.* First, we consider the case under Condition 1 and $\underline{K}_1 > 0$. It follows from Proposition 7 that $\lim_{K \downarrow 0} \underline{\vartheta} = \lim_{K \downarrow 0} \bar{\vartheta} = \underline{\theta}^0$. Recall from the proof of Proposition 4 that $\hat{\mathbf{w}} = \widetilde{w}(\underline{\vartheta})$ and $\mathsf{c} = C(\underline{\vartheta})$. Hence, we have $\lim_{K \downarrow 0} \hat{\mathbf{w}} = \lim_{K \downarrow 0} \widetilde{w}(\underline{\vartheta}) = \widetilde{w}(\underline{\theta}^0)$ and $\lim_{K \downarrow 0} \mathsf{c} = \lim_{K \downarrow 0} C(\underline{\vartheta}) = C(\underline{\theta}^0)$. The conclusion (52) is obtained by setting $\theta_0 = \underline{\theta}^0$, $\hat{\mathbf{w}}_0 = \widetilde{w}(\underline{\theta}^0)$ and $\mathsf{c}_0 = C(\underline{\theta}^0)$. Moreover, $\mathcal{V}_{\hat{\mathbf{w}}}$ and $\mathsf{V}_{\mathsf{c}}$ converge uniformly to $\mathcal{V}_{\hat{\mathbf{w}}_0}$ and $\mathsf{V}_{\mathsf{c}_0}$, respectively, as $K$ approaches 0. Consequently, both value functions as defined in (34) converge to $\mathfrak{V}_{\theta_0}$ uniformly as $K$ approaches 0. Invoking Proposition 5 and sending $K$ to 0, we conclude that functions $V_{\mathsf{l}} = V_{\emptyset} = \mathfrak{V}_{\theta_0}$ satisfy the optimality conditions (20)–(22) for $K = 0$.

The argument for the case under Condition 2 and $\breve{K}_2 > 0$ is exactly the same, and thus is omitted. ▫

### B.6. Proofs of the Results in Appendix A

*Proof of Proposition 8.* These results have already been shown in the second part of the proof of Proposition 7. ▫

*Proof of Theorem 9.* The proof consists of three parts.

*Part 1.* Similar to the proof of Proposition 2, we can show that (i) $U(\Gamma^*(w, \mathsf{l}; 0, 0, \bar{w}, \bar{w}), \mathsf{l}) = V_e(w)$ and $U(\underline{\Gamma}, \emptyset) = \underline{v}$ for any $w \geqslant 0$; and (ii) under condition that $K < \bar{K}_e$ and $m^K > 1$, $U(\Gamma^*(w, \mathsf{l}; 0, 0, \bar{w}, \bar{w}), \emptyset) = V_{\emptyset}(w) - w$ for any $w \geqslant \bar{\theta}^K$ with $V_{\emptyset}$ as defined in (71).

*Part 2.* Next, we show that under condition $K \geqslant \bar{K}_e$, functions $V_{\mathsf{l}} = V_e$ and $V_{\emptyset} = \underline{v}$ satisfy (20)–(22). By the definition of $V_e$, it is clear that $\mathcal{A}_{\mathsf{l}} V_{\mathsf{l}} = 0$. Moreover, $(\mathcal{A}_{\emptyset} V_{\emptyset})(w) = r\underline{v} - \underline{\mu} R = 0$ for any $w \geqslant 0$. Hence, (20) holds.

Note that $V_e$ is increasing on $[0, \bar{w}]$ (see Lemma 3 of Sun and Tian (2018)). Hence, for any $w \geqslant 0$, we have $V_{\mathsf{I}}(w) - V_{\emptyset}(w) \geqslant V_e(0) - \underline{v} = 0$ and $V_{\mathsf{I}}(w) - V_{\emptyset}(w) \leqslant \bar{V}_e - \underline{v} = \bar{K}_e \leqslant K$. Hence, (21) holds. Finally, it is evident that (22) holds.

*Part 3.* Now we show that under condition $K < \bar{K}_e$, functions $V_{\mathsf{I}} = V_e$ and $V_{\emptyset}$ as defined in (71) satisfy (20)–(22). Obviously, $\mathcal{A}_{\mathsf{I}} V_{\mathsf{I}} = 0$. Moreover, we have

$$\begin{aligned}(\mathcal{A}_{\emptyset} V_{\emptyset})(w) &= r V_{\emptyset}(w) - r w V_{\emptyset}'(w) - \underline{\mu} R \\ &= r w \left( \frac{V_{\emptyset}(w) - V_{\emptyset}(0)}{w} - V_{\emptyset}'(w) \right) \geqslant 0,\end{aligned}$$

where the equality follows from $V_{\emptyset}(0) = \underline{v}$, and the inequality follows from the concavity of $V_{\emptyset}$. Hence, (20) holds.

If $w \geqslant \bar{\theta}^K$, then $V_{\mathsf{I}}(w) - V_{\emptyset}(w) = K$. If $w \in [0, \bar{\theta}^K]$, then $V_{\mathsf{I}}'(w) - V_{\emptyset}'(w) = V_e'(w) - V_e'(\bar{\theta}^K) \geqslant 0$ due to the concavity of $V_e$, which implies that $V_{\mathsf{I}}(w) - V_{\emptyset}(w) \geqslant V_{\mathsf{I}}(0) - V_{\emptyset}(0) = 0$ and $V_{\mathsf{I}}(w) - V_{\emptyset}(w) \leqslant V_{\mathsf{I}}(\bar{\theta}^K) - V_{\emptyset}(\bar{\theta}^K) = K$. Hence, (21) holds. It is straightforward to see that (22) holds. ◻

## References


Biais, B., T. Mariotti, J.-C. Rochet, and S. Villeneuve (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica 78*(1), 73–118.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer Series in Statistics. Springer.

Cao, P., F. Tian, and P. Sun (2021). Comment on "optimal contract to induce continued effort". *Management Science*. Forthcoming.

Protter, P. E. (2003). *Stochastic Integration and Differential Equations* (second edition ed.), Volume 21 of *Stochastic Modelling and Applied Probability.* Springer.

Sun, P. and F. Tian (2018). Optimal contract to induce continued effort. *Management Science 64*(9), 4193–4217.