Performance Cycles*

David Lagziel† and Ehud Lehrer‡

July 24, 2023

Abstract:

A decision maker repeatedly exerts efforts to produce new outputs, while being compensated based on his past and current production levels. We show that the decision maker's optimal strategy dictates a cyclic oscillatory performance whenever the compensation depends on recent past performance. We apply our model to various economic settings such as the delegated portfolio-managers problem, an R&D investment problem, and a dynamic advertising problem.

Journal of Economic Literature classification numbers: C61, C62, C79.

Keywords: Performance cycles; Performance persistence; Dynamic Optimization.

*The authors wish to thank Eddie Dekel, Zvika Neeman, Dotan Persitz, David Schmeidler, Eilon Solan, and Avi Wohl for their valuable comments; as well as the participants of the ninth edition of the UECE Lisbon Meetings, the 29th Stony Brook Game-Theory conference, the Bar-Ilan University Economics Seminar, and the 10th annual Israeli Game-Theory conference. Lagziel acknowledges the support of the Israel Science Foundation, Grant #513/19.

†Ben-Gurion University of the Negev, Beer-Sheva 8410501, Israel. e-mail: davidlag@bgu.ac.il.

‡Department of Economics, Durham University, UK. E-mail: ehud.m.lehrer@durham.ac.uk.

1 Introduction

The phrase 'buy low, sell high' is presumably the timeless most simplest motto to the world of trading. Easy to explain and easy to understand, but quite difficult to apply since people are not too eager to buy a badly performing low-reputation product with the hope it would blossom in due time. Now consider the portfolio-management industry in which past returns and reputation coincide. Is it reasonable for an investor to invest through a low-return, potentially losing, investment firm? A recent study by Cornell et al. (2017) provides some indications by showing that returns oscillate in adjacent time periods as top-ranked investment firms, in terms of realized returns, become bottom-ranked firms and vice versa. Similar indications follow from Bessler et al. (2018) portraying limited reversions in realized returns as time progresses (see, e.g., Figure 2 of the relevant study).

In this paper, we try to explain such phenomena by studying the basic tension between past performance and incentives. We pursue this goal through a dynamic model in which a strategic decision maker (DM) repeatedly exerts effort to produce output. On the one hand, the DM is subject to instantaneous convexly increasing costs that limit his ability to exert effort. On the other hand, the DM is compensated according to his past and current production levels, so any additional effort at a given stage becomes beneficial in the stages to follow. The interaction between these two forces requires the DM to balance current costs with future earnings which ultimately depend on aggregate recent production, thus endogenously inducing a cyclic performance.

A key attribute of our model is the method by which the DM is compensated. The potential evaluation methods range from the sole last performance to the DM's accumulated infinite track record. We capture these alternatives through two distinct mechanisms, the *Transient model* and the *Persistent model*, that respectively represent short- and long-term averaging.

Under the Transient model, the DM is evaluated and compensated based on his last 2-stage aggregate (or average) output levels. This assessment ensures that any cost cut, at a given stage, immediately limits the DM's payoff at the subsequent stage, since one period's output is next period's benchmark position. This intuition establishes our first main result where, *even in a deterministic set-up*, the DM's optimal policy dictates that production oscillates around a certain stable level, while converging towards it asymptotically.

In the Persistent model, to differ, the DM is rewarded based on a discounted sum of all past production levels. The main focus of this part is to study how changes in the evaluation process affect the DM's incentives. It appears that a DM with an a-priori high evaluation profits from a higher weight on past performance maintaining his elite status at lower costs, while a low-evaluation DM benefits from myopic assessments of past production for the opposite reasons. We also prove the existence of a *basic tension between incentives and past evaluation* by showing that optimal incentives are reached only in case past performance is completely ignored. In addition and similarly to the Transient model, we show that a high-evaluation position carries similar adverse effects over incentives, whereas the convergence to a stable state is monotonic rather than cyclic.

1.1 Related literature & main contribution

Our model and analysis combine several well-documented models and policies in the economic world. The oscillatory optimal strategy carries some resemblance to the optimal (S,s)-policies in inventory problems, where an agent allows his inventory to fall until is reaches a low level s, only to be imminently increased to a high level S. Such policies were vastly studied in the context of the Pricing Problem (price adjustments and inflation), the Technology-Update Problem, and the Capital Stock Adjustment Problem.[1] In the delegated portfolio-manager context, the rational model of Berk and Green (2004) contributes one of the key ingredients for non-persistent performance. Berk and Green propose a non-strategic model where non-persistence and related phenomenons are attributed to decreasing returns to scale, as firms that recently outperformed suffer from the positive inflow of funds. Similar non-persistence arises in the seminal work of Holmstrom (1999), where the market's inference about ability is a random walk with a declining variance.

In view of previous studies, we can underline the first contribution of the current work: accommodating for endogenous cyclic performance in a general framework. Our results indicate that a cyclic performance, rather than non-persistence, could be attributed to incentives whenever payoffs depend on the DM's past performance. Such oscillations were recently observed in the empirical studies of Cornell et al. (2017) and Bessler et al. (2018), both studying the portfolio management industry. These studies show that recent performance, namely the short-term moving-average of the DM's yields, is negatively correlated with forthcoming ones. Interestingly, their evaluation methods and results are in-line with our theoretical predictions.[2] Additionally, our findings complement the recent study by Ishiguro (2022), which identifies endogenous and persistent management cycles under a different dynamic setting.

The second key contribution of our work relates to a more theoretical strand in the literature,

[1]See, e.g., Arrow et al. (1951); Dvoretzky et al. (1952, 1953); Bellman et al. (1955); Bailey (1956); Arrow et al. (1958); Scarf (1959); Barro (1972); Sheshinski and Weiss (1977), and Sheshinski and Weiss (1993) for a general survey.

[2]The public and academic debate over the performance of fund managers is rather extensive. On the one hand, papers such as Grinblatt and Titman (1992), Elton et al. (1993), Hendricks et al. (1993), Goetzmann and Ibbotson (1994), Brown and Goetzmann (1995), Carhart (1997), Bollen and Busse (2005), and Busse et al. (2010), indicate that either long- or short-term performance persist, while many others, such as Goyal and Wahal (2008) and Barras et al. (2010), claim otherwise.

regarding the problem of global stability in discounted problems.[3] Our contribution to this line of research follows from the identification of the process and method by which the production converges to a stable state. That is, we do not limit ourselves to proving the existence of a unique optimal strategy and state; rather we show how they systematically converge due to the evaluation process. In this respect our results go beyond the fundamental work of Blackwell (1965), which guarantees convergence to an optimal policy by showing that *the optimal policy itself has a fixed point.* We go further and prove that the same optimal policy induces a cyclic convergence towards that fixed point. In other words, we prove the existence of *a fixed point within a fixed point*, and show how the process endogenously converges to both.

Another contribution is attributed to the inclusion of information frictions between the DM and his evaluator. In our model, the evaluation process follows simple heuristics that dictate the DM's payoff. There is thus no need for a common prior, Bayesian updating, or even common knowledge of distributions over abilities. The stochastic elements in our model need not be normally distributed, i.i.d., or even ergodic. This generality opens the door to a broad analysis of the problem from a designer's viewpoint.

1.2 Structure of the paper

The paper is organized as follows. Section 2 depicts the Transient model and is divided into two parts, one is devoted to the deterministic case and the other to the stochastic one. In Section 3 we revert to the Persistent model, and in Section 4 we focus on the way changes in the evaluation process affect incentives. Concluding remarks and comments are given in Section 5.

2 The Transient model

Consider a decision maker (DM) in an infinitely repeated set-up. At every stage, the DM strategically exerts efforts that translate into outputs, while his per-period payoff depends on either the aggregate, or the average of his last and current output levels. The DM's main goal is to maximize the infinite discounted-sum of the effort-deducted payoffs.

Formally, let $E = [e_{\min}, e_{\max}] \subseteq \mathbb{R}_+$ be a non-empty compact interval denoting the DM's single-period effort choice. For every effort level $e \in E$, the single-stage output $Q(e)$ is determined by *the output function* $Q : E \to \mathbb{R}_+$. The DM's reward is defined by *the reward function* $R : Q(E) \to \mathbb{R}_+$. Specifically, the last and current output levels are reduced to a single factor q, referred to as *past*

[3] See Scheinkman (1976); Rockafellar (1976); Cass and Shell (1976); Brock and Scheinkman (1976), among many others.

performance such that the DM receives the single-stage reward of $R(q)$. Due to diminishing marginal returns, both R and Q are assumed to be strictly increasing, strictly concave, and continuously differentiable functions.

The problem begins at stage $t = 1$ with an initial output of $Q_0 = Q(e_0)$.[4] The DM chooses an effort e_1 to generate an output of $Q_1 = Q(e_1)$. Once Q_0 and Q_1 are realized, the DM is rewarded a payoff of $R(Q_0 + Q_1)$. Continuing inductively, at every stage $t > 1$, and given past output $Q_0, Q_1, \ldots, Q_{t-1}$ where $Q_{t-1} = Q(e_{t-1})$, the DM exerts effort e_t, which generates an output of $Q_t = Q(e_t)$, and provides a reward of $R(Q_{t-1} + Q_t)$. The DM's profit in the repeated process is given by the β-discounted sum

$$\pi(\underline{e}) = \sum_{t=1}^{\infty} \beta^{t-1}[-e_t + R(Q_{t-1} + Q_t)], \tag{1}$$

where $\underline{e} = (e_1, e_2, \ldots)$ is the DM's infinite-horizon realized actions and $\beta \in (0, 1)$.

Before we proceed, let us comment on two assumptions concerning the output function and stage payoffs. First, the reward function's concavity assumption could be weakened as long as the composition $R \circ Q$ remains concave. Second, an equivalent way to present the stage payoff is by letting the DM directly choose the output level Q_t, so that $-e_t + R(Q_{t-1} + Q_t) = -Q^{-1}(Q_t) + R(Q_{t-1} + Q_t)$. Though the two representations are completely equivalent, in the notations on the right the DM directly chooses output, which translates to an effective cost through a standard increasing and convex *cost function* $Q^{-1} : \mathbb{R}_+ \to E$. Hence, the output function's concavity is basically an alternative manifestation of convexly increasing costs.

A *strategy* σ is a function from all past realized outputs (histories) $\bigcup_{t \in \mathbb{N}} Q(E)^t$ to the effort set E. A *stationary strategy* σ is a function from the set $Q(E)$ of single-period output to the effort set, or equivalently, a function from the effort set to itself. In general, the process summarized in Eq. (1) is an MDP, where the state variables are either E, or equivalently, $Q(E)$.[5] So, a stationary strategy is a time-invariant one-stage strategy that is played repeatedly throughout the process, and depends solely on the current state variable rather than on the entire history.

Given any strategy σ and an initial output level $Q_0 = Q(e_0)$, derived from an initial effort level e_0, denote the DM's payoff by $\pi(e_0|\sigma)$, where all effort levels $\{e_t\}_{t \in \mathbb{N}}$ are determined according to σ. A strategy is considered *optimal* if it solves the optimization problem

$$\pi(e_0|\sigma) = \sup_{\sigma} \sum_{t=1}^{\infty} \beta^{t-1}[-\sigma_t + R(Q(\sigma_{t-1}) + Q(\sigma_t))],$$

[4] The initial condition could be exogenously fixed or randomly chosen. In any case, it will become redundant under the generalization to a stochastic set-up.

[5] The equivalence follows from the fact that the output function is continuous and strictly increasing and the effort set is convex, thus there exists a one-to-one mapping between the two sets, E and $Q(E)$.

where $\sigma_1 = \sigma(Q_0)$, and $\sigma_k = \sigma(Q_0, Q(\sigma_1), \ldots, Q(\sigma_{k-1}))$. That is, a strategy is optimal if it produces the maximal payoff, denoted $\hat{\pi}(e_0)$, given an initial effort level e_0.

The interior-solution property. To simplify the analysis, we require an additional technical assumption stating the the optimal solution is not trivial. Namely, we fix the parameters such that the extreme points $\{e_{\min}, e_{\max}\}$ cannot be the DM's optimal action. One can clearly weaken this assumption by restricting the initial condition to a subset which ensures the end-point solutions are suboptimal.

Remark 1. *In the Transient model we assume that the reward function is deterministic. To differ, in typical principal-agent models, the stage-output Q_t depends stochastically on the chosen effort level e_t, i.e., $Q_t \sim P(\cdot|e_t)$ for some probability measure P. The stochastic framework described in* Section 2.3 *accounts for this possibility.*

2.1 Economic applications

Though we use a generic decision-problem terminology, our model applies to several economic scenarios. For example, consider the portfolio-management industry. Any investment firm typically exerts per-period "effort", either through the accumulation of information as in Stoughton (1993) and Admati and Pfleiderer (1997) or through managerial replacements as in Lynch and Musto (2003) and Dangl et al. (2008). This single-stage effort of e_t translates into the return $Q(e_t)$.

Now investors typically cannot a priori determine the ability of a fund manager or an investment firm. For that purpose, the market observes past returns and averages recent performance to determine expected ability.[6] The dependence on recent past performance is crucial in this context, since it enables a more accurate evaluation of one's abilities. For tractability, we assume that investors observe the last two stages average performance $\frac{1}{2}(Q_{t-1} + Q_t)$ and allocate funds accordingly. Note that the use of running averages implies that current and recent performance are perfect substitutes. Due to assets-under-management fees, the recent average return becomes lucrative as funds flow towards the firm in stages to follow. The latter interaction is captured through the reward function $R(Q_{t-1} + Q_t)$. Interestingly, recent empirical studies (see Cornell et al. (2017) and Figure 2 of Bessler et al. (2018)) portrayed a limited cyclic performance at the firm level, whenever firms were evaluated through a moving average of recent performance.

Another application of our model is evidenced in any trading process involving credit, such that payments are distributed along sequential time periods. A firm constantly produces output

[6] For example, the New York Times regularly publishes the total returns of various funds according to the previous year, three years, and five years.

$(Q_1, Q_2, \dots)$, but payments are partitioned over two adjacent stages, namely the payoff at stage t is $R\left(\frac{1}{2}(Q_{t-1} + Q_t)\right)$.[7] Thus, the single-stage payoff depends partially on recent past performance due to the postponed payments and depends partially on current performance, so the services and costs are instantaneous whereas payments are postponed. Given such interpretation, past performance is translated into direct monetary transfers, and the firm would strive to balance its recent average production, by "smoothing" its income and production.

A recent study by Chen et al. (2016) provides evidence of negative autocorrelation in decision making in the context of asylum judges, loan officers, and baseball umpires. Namely, they find significant negative autocorrelation under which loan officers are less likely to approve a loan if they approved the previous loan, and vice versa. Though their primary explanation is the gambler's fallacy, one can also think of loan officers as agents that are compensated based on their aggregate performance. Thus, it becomes optimal for them to smooth their performance over time such that a substantial extraction of effort, which leads to credit denial, is followed by an under-investment of effort and a loan approval. This becomes even more evident given our analysis of the stochastic case in Section 2.3 below, which accounts for exogenous variations in output, and the inverted strategic reaction of the DM towards them.

One could also adapt our model to a semi optimal-growth model with post-generational transfers, as a portion of one's wealth is transferred to subsequent generations. By and large, any strategic interaction that combines the two previously mentioned key components of marginally decreasing output and history-based payoffs will be closely related to our framework and, therefore, to our conclusions *regarding cyclic performance to incentives.*

2.2 Main result - the deterministic case

The payoff function given in Eq. (1) presents the basic tension under which the DM operates. Over (or under) investing in effort at one stage, generates a balancing counter-reaction at the subsequent stage to invest less (or more) effort. This balancing effect motivates Theorem 1 which follows.

Theorem 1 states that there exists an *stationary* effort level e^*, such that the DM balances his performance relative to e^* at every two adjacent stages. Namely, in case the current evaluation level is higher (or lower) than the absorbing level, the DM will invest less (or more) effort relative to e^*, to balance the performance at the subsequent stage. These alternating effort levels continue to fluctuate around the stationary level, while converging towards it asymptotically. By their technical nature, we defer all proofs to the appendix.

[7] We simplify the exposition by considering an evenly partitioned payment over two stages.

Theorem 1. *There exists a unique, stationary and continuous optimal-strategy $\sigma : E \to E$. Given σ, the payoff function $\hat{\pi}(e_0) = \pi(e_0|\sigma)$ is a strictly concave, and continuously increasing function of e_0. In addition, if the interior-solution property holds, then:*

- *the optimal strategy σ is strictly decreasing with a single fixed point $e^* \in (e_{\min}, e_{\max})$;*
- *the sequences $\left(\sigma^{2n}(e_0)\right)_{n\in\mathbb{N}}$ and $\left(\sigma^{2n+1}(e_0)\right)_{n\in\mathbb{N}}$ monotonically converge to e^*;*
- *the fixed point e^* is bounded between $\sigma^n(e_0)$ and $\sigma^{n+1}(e_0)$ for every $n \in \mathbb{N}$.*

Though the proof is given in Appendix A, we wish to provide a brief description of its main elements. The first part of the proof follows standard dynamic-optimization techniques, based closely on the results in Chapter 4 of Stokey et al. (1989). We use their analysis, which originates from Bellman's principle of optimality and Blackwell's Contraction Mapping Theorem, to establish that a unique, stationary, and continuous optimal-strategy exists, and that the payoff function is a strictly concave, continuously increasing function, where the concavity follows from the concavity of $R(Q(\cdot))$.

The second, and more original, part of the proof relates to our key insight concerning the cyclic performance. It is based on the idea that the optimal strategy σ is a strictly decreasing function from E to itself, hence it maintains a fixed point e^*. We establish this monotonicity by taking the standard Bellman equation and differentiating both sides w.r.t. the optimal action, given e_0. Then, we perform a comparative statics by studying an incremental increase in e_0 to show that σ is a strictly decreasing function (this relies on the concavity of all the aforementioned functions). Once monotinicity is established, the fixed point e^* follows directly, and every initial position $e_0 < e^*$ must be accompanied with an optimal action $\sigma(e_0) > e^*$. The action $\sigma(e_0)$, which exceeds e^*, is also the initial condition of the subsequent stage, thus generating an oscillating performance.

The fact that the fixed point e^* is bounded between any two subsequent effort levels suggests that the DM's performance is of a cyclic nature. Assume, e.g., that $\sigma^k(e_0) > \sigma^{k+1}(e_0)$ for some given stage k. It must be the case that $\sigma^k(e_0) > e^* > \sigma^{k+1}(e_0)$, so in order to maintain the same e^*-interior-condition between stages $k+1$ and $k+2$, it follows that $\sigma^{k+2}(e_0) > e^* > \sigma^{k+1}(e_0)$. This implies that output oscillates around e^* throughout the process.

In other words, Theorem 1 suggests that a cyclic performance, monotonically and systematically converging to equilibrium, is natural when dealing with a DM concerned with current and recent performance. This outcome captures two important aspects of the current work. First, depicting a specific path and method of converging to a stable state in a dynamic-optimization problem. Second, linking aggregated performance, and therefore incentives, to a cyclic-performance phenomenon. This result somewhat resembles the outcome in Holmstrom (1999), where the market's evaluation of an agent's ability is a random walk with decreasing volatility. However, note that we reached this conclusion in a deterministic framework and, in the following section, we extend it to a *general* Markov process.

2.3 Main result - the stochastic case

The first extension of the Transient model concerns the introduction of randomness to the output function. The randomness that we impose need not be i.i.d or even ergodic. Rather, we assume that the output function depends on a randomly chosen state of the world, dictated by a Markov process, along with prior dependence on the DM's strategic effort. Though its general nature, this extension does not impair previous results.

Formally, consider a finite[8] set Ω of states and denote by $P = (P_{ij})_{1 \leqslant i,j \leqslant |\Omega|}$ the transition matrix where P_{ij} is the probability of moving from state i to state j in a single time period. Given the states and transition function, consider a generalization of the output function such that $Q : \Omega \times E \to \mathbb{R}_+$ depends on the realized state $\omega \in \Omega$ and on the DM's effort. We assume that the output function maintains its basic properties independently of the realized state. Namely, for every $\omega \in \Omega$, the output function's ω-section $Q_\omega : E \to \mathbb{R}_+$ is a strictly increasing, strictly concave, continuously differentiable function. Denote by S the convex hull of the compact set $Q(\Omega, E)$ of all possible realized outputs.

The stochastic decision problem evolves similarly to the deterministic one. At stage $t = 1$, with an initial state $\omega_0 \in \Omega$ and an initial output of Q_0, the DM chooses an effort level $e_1 \in E$. Next, a state ω_1 is realized according to P and ω_0, and the single-stage realized-output is $Q_1 = Q(\omega_1, e_1)$. Continuing inductively, at every stage $t > 1$ and given a *history* $h_{t-1} = (\omega_0, Q_0, \omega_1, Q_1, \ldots, \omega_{t-1}, Q_{t-1})$ of past realized outputs and states, the DM chooses an effort e_t. A state ω_t is realized according to P and ω_{t-1}, and the single-stage output is $Q(\omega_t, e_t)$. Therefore, a strategy σ applied by the DM is a function from the set $\bigcup_{t \in \mathbb{N}} (\Omega \times S)^t$ of all finite histories to E, such that $\sigma(h_{t-1}) = e_t$ is the strategy's realized action at stage t.[9] As before, a stationary strategy is a time-invariant one-stage strategy that is played repeatedly throughout the process, and maps every state variable (i.e., every (ω_t, Q_t) in this case) to an effort level e_t in every stage t. [10]

[8] In general, the use of a finite state space could be avoided by taking any compact Borel set in $\mathbb{R}$. In that case, the transition function must hold the Feller property, roughly stating that every bounded continuous function is mapped, under the expectation operator and given the transition function, to a bounded continuous function.

[9] Since Q and R are continuous, E is compact, and Ω is finite, measurability requirements are met.

[10] Following Remark 1, one can extend the given stochastic framework so that the state of the world is a pair (ω, ξ), chosen from a product space $\Omega \times \Xi$. While ω represents the random state of the economy, ξ is responsible for the stochastic aspect of the realized output (once ω and the effort level are determined). At stage t the pair (ω_t, ξ_t) is drawn according to $P(\cdot|\omega_{t-1})$, so that the output is stochastically determined by e_t and ω_t: $Q_t \sim Q(\cdot|\omega_t, e_t)$. In other words, as the stochastic model dictates, once (ω_t, ξ_t) is realized, Q_t is uniquely determined by the effort level e_t. Assuming that

Given a strategy σ and initial conditions (ω_0, Q_0), the DM's expected β-discounted payoff is

$$\pi(\omega_0, Q_0|\sigma) = \mathbf{E}_{\sigma,\omega_0}\left[\sum_{t=1}^{\infty}\beta^{t-1}(-e_t + R(Q_{t-1}+Q_t))\right], \tag{2}$$

where $\mathbf{E}_{\sigma,\omega_0}[\cdot]$ is the expectation operator with respect to the probability measure induced by the transition probabilities P, the initial state ω_0, and the strategy σ. Note that the strategy is a random variable since it depends on realized states. Thus, the expectation operator also relates to the strategy-induced effort levels throughout the stages.

By the randomness of the process, the DM's realized output might not accurately follow the same cyclic performance as in Theorem 1. However, in Theorem 2 which follows, we prove that current output decreases (in expectation) with respect to previously realized ones. We also show that the optimal strategy is a strictly decreasing function of a recently realized output. Thus, the oscillating process presented in the deterministic case remains valid.

Theorem 2. *There exists a unique, stationary and continuous optimal-strategy $\sigma : \Omega \times S \to E$. Given σ and ω_0, the payoff function $\hat{\pi}(\omega_0, Q_0) = \pi(\omega_0, Q_0|\sigma)$ is a strictly concave, continuously increasing function of Q_0. In addition, if the interior-solution property holds, then $\boldsymbol{E}_{\sigma,\omega_{t-1}}[Q_t|Q_{t-1}]$ is a strictly decreasing function of Q_{t-1}, and for every state ω visited infinitely many times, there exists an output level q_ω such that, w.p. 1, output oscillates around q_ω infinitely many times.*

The construction of the proof of Theorem 2, given in Appendix B, follows similar arguments as the proof of Theorem 1. We adapt the results of Chapter 9 of Stokey et al. (1989) to establish a unique, stationary, and continuous optimal strategy, and a strictly concave, increasing payoff function. Next, we differentiate the Bellman equation (as in the proof of Theorem 1), but with the expectation operator, to perform comparative statics and show that $\sigma(\omega_0, Q_0)$ is decreasing in Q_0. We then define the function $\psi_\omega(q) = \mathbf{E}_\omega\left[Q(\tilde{\omega}, \sigma(\omega, q))\right]$, where ω is fixed and $\tilde{\omega}$ is drawn according to P, and use the monotonicity of σ to establish a fixed point q_ω for $\psi_\omega(q)$. This fixed point, along with the monotonicity of σ, yields the oscillating performance stated in the theorem.

One way to compare the monotonicity result of Theorem 2 to the one given in Theorem 1, is by taking a mass-point measure over Ω. In such a case, the probability space becomes trivial and the term $\mathbf{E}_{\sigma,\omega_{t-1}}[Q_t|Q_{t-1}]$ translates to $Q(\sigma_{t-1})$, where $\sigma_{t-1} = \sigma(Q_{t-1})$ and the state Q_{t-1} is fixed. Therefore, the result which states that $\mathbf{E}_{\sigma,\omega_{t-1}}[Q_t|Q_{t-1}]$ is strictly decreasing in Q_{t-1} implies that σ is strictly decreasing, taking into account the fact that Q is an increasing function.

Q_{ω_t,ξ_t} maintain its key properties w.r.t. e_t (i.e., a strictly increasing, strictly concave, and continuously differentiable function), then Theorem 2 below follows accordingly.

An immediate question that originates from Theorem 2 relates to the role of the DM's strategy in inducing oscillation, versus the natural oscillations that occur due to randomness. In the proof, given in Appendix B, we not only show that infinitely many oscillations occur, but also prove that these oscillations occur infinitely many times between extreme output levels, related to extreme states. These levels are not necessarily close. In this sense, the strategic reaction of the DM intensifies the oscillatory trends, even under the mean-reversion phenomenon. Therefore, Theorem 2 joins Theorem 1 to support the results of various empirical studies suggesting that DMs' performance patter tends to be cyclic. Our results and model explain such occurrences through a straightforward economic reasoning, which goes beyond the probabilistic properties. Namely, when DMs' payoffs are history dependent, and production is increasingly costly, DMs' strategies level their performance accordingly.

Note that the monotonicity result in the last statement of Theorem 2 relates strictly to previously realized output, and not to the different states. Therefore, we fix the state variable ω_{t-1}, and derive monotonicity through changes in the realized output R_{t-1}. Without further assumptions over transitions and states, the realized output in adjacent stages might actually increase, e.g., in case the fixed state variable is significantly worse (in terms of production) than all other states.

2.3.1 Accommodating for competition

The stochastic model analysed in Theorem 2 is more than just a technical extension of the deterministic set-up. In fact, it could accommodate several possible interpretations, specifically relating to competition between various DMs as in Lagziel and Lehrer (2018), where the interaction between DMs is mediated explicitly by performance-based relative compensation.

In a general strategic set-up, each DM should respond not only to the current state but also to other players' strategies. In the current work, an implicit interaction between DMs could occur through the reward function: the DM's performance can be interpreted as relative to, e.g., other DMs or some exogenous benchmark. One could also attribute the proposed randomness to fluctuations in demand, assuming that the DM represents a price-taker firm interacting with a general set of consumers.

Our approach has both an advantage and a disadvantage. On one hand, our analysis does not require that a DM would know the strategies employed by all his rivals - the aggregate effect of all other players is encapsulated in an exogenous stochastic process which determines (along with his effort) the DM's payoff. On the other hand, our model does not deal with situations where the DM is truly strategic and incorporates other players' behaviour in his decisions. Moreover, the DM's payoff in our model is independent of other opponents' performance. Thus, a full analysis of a general framework with multiple players is left for future research.

3 The Persistent model

In this section we extend the preceding analysis by conditioning the DM's compensation on a longer history path. Specifically, we study a process in which the reward function depends on a discounted sum of all past output levels. The purpose of this adjustment is to address the decision problem through a wider scope. Namely, we wish to study how changes in the evaluation process affect, in the long run, the DM's realized production and payoff. Therefore, instead of focusing on short-term effects as in Section 2, we now turn to inspect long-term effects due to adjustments in the evaluation mechanism.

Formally, fix a discount factor $\lambda \in (0,1)$ and consider an optimization problem where the DM's performance is evaluated by a λ-discounted sum of past output. That is, we track the following optimization problem

$$\hat{\pi}(Q_0) = \sup_{\sigma} \sum_{t=1}^{\infty} \beta^{t-1}[-\sigma_t + R((1-\lambda)\widehat{Q_{t-1}} + \lambda Q(\sigma_t))],$$

where $\widehat{Q_0} = Q_0$, $\widehat{Q_k} = (1-\lambda)\widehat{Q_{k-1}} + \lambda Q(\sigma_k)$ for every k, and $\sigma_k = \sigma(Q_0, Q(\sigma_1), \ldots, Q(\sigma_{k-1}))$. In words, this optimization problem is similar to the original Transient model, apart from the exchange of the two-stage evaluation $Q_{t-1} + Q_t$ with the discounted sum $(1-\lambda)^t Q_0 + \lambda \sum_{n=1}^{t} (1-\lambda)^{t-n} Q_n$. We refer to this problem as the *Persistent model*.

The evolution of the Persistent set-up is similar to that of the transient one. At the beginning of every stage t, and given an evaluation of $\widehat{Q_{t-1}}$, the DM exerts effort e_t to generate an output of $Q(e_t)$. The DM's updated evaluation is set to $\widehat{Q_t} = (1-\lambda)\widehat{Q_{t-1}} + \lambda Q(e_t)$ and the DM collects a reward of $R(\widehat{Q_t})$. The process continues indefinitely.

3.0.1 Economic applications

A straightforward application of the persistent model relates to an R&D investment problem, where a firm decides how much to invest in R&D at every time period. The decision and costs are instantaneous, whereas the payoff is accumulated and spread throughout the stages. Any breakthrough, either a technological advantage or superior marketing abilities, would give the firm a long-term edge over the competition; an advantage which depreciates over time. In this context, an empirical study by Artés (2009), which focuses on R&D decisions in Spanish firms, finds significant persistence in firms' decision to invest in R&D over time. This is consistent with our prediction of monotone convergence in the Persistent model, compared to the oscillating convergence in the Transitory model.

Another application is attributed to a dynamic advertising problem where performance is accumulated in the process of building a franchise. A firm typically invests in advertising and its payoff depends on the accumulated investment with some form of depreciation.

3.1 Main result - the deterministic case

A comparison of Section 2's results with the following Theorem 3 will certify that many previous results hold under the updated problem, while others change completely. Starting with the similarities, we show that there exists a unique, stationary and continuous optimal-strategy $\sigma : Q(E) \to E$, such that the optimal payoff function is a strictly concave, continuously increasing function.[11] In addition, we again prove that the optimal strategy is strictly decreasing, thus implying that high past evaluation carries the same adverse effect over incentives. On the other hand, the two models differ in the paths by which the systems converge to a stable production level. In particular, the Persistent model generates a monotonic convergence, rather than an oscillating one. We broadly relate to this aspect after formally presenting the results of Theorem 3. The proof is given in Appendix C.

Theorem 3. *In the Persistent model, there exists a unique, stationary, and continuous optimal-strategy $\sigma : Q(E) \to E$. Given the optimal strategy, the payoff function $\pi(Q_0|\sigma)$ is a strictly concave, continuously increasing function of Q_0. In addition, if the interior-solution property holds, then the function $Q(\sigma) : Q(E) \to Q(E)$ is strictly decreasing with a single, interior, fixed point $Q^* \in Q(E)$. Moreover, the sequence $(\widehat{Q_t})_{t \in \mathbb{N}}$ of realized discounted performance, generated by σ and Q_0, monotonically converges to Q^*.*

The proof of Theorem 3, given in Appendix 3 follows the same steps as the proof of Theorem 1, yet reaches a somewhat different result. Namely, there are two important aspects that arise from the comparison of Theorem 3 with Theorem 1. First, the monotonic convergence of the discounted performance towards Q^*, and second, the optimal strategy's monotonicity. Let us provide a short discussion regarding these aspects.

To understand the differences between the monotonic convergence in the Persistent model (Theorem 3) and the oscillating convergence in the Transient model (Theorem 1), one needs to think of the different dynamics concerning the state variable in each set-up. Assume, e.g., that the state variable in stage t is below the stable (fixed-point) level Q^*. Then, in both models, the DM optimally generates an output Q_{t+1} which exceeds Q^*. That is, the optimal strategy in both models dictates that $Q_{t+1} > Q^*$; a similarity between the two models and results.

[11] As in the Transient model, the one-to-one mapping between E and $Q(E)$ enables us to consider either one of the two sets as the state variables. Namely, for every state q, given in terms of the discounted output level, there exists a unique effort level e such that $Q(e) = q$. To simplify the notation in the current analysis, we choose to use $Q(E)$ as state variables, with no loss of generality.

Let us now understand how this affects the state variable in subsequent stages. In the Transient model, which is Markovian, the fact that $Q_{t+1} > Q^*$ implies that state variable in stage $t+1$ is above the fixed-point level Q^*, thus generating a cyclic performance along the stages. In the Persistent model, however, this is not the case because the state variable in stage $t+1$ is a discounted sum of all past outputs. So the fact that $Q_{t+1} > Q^*$ does not imply that the state variable in stage $t+1$ is above Q^*. The state variable $\widehat{Q_{t+1}}$ in stage $t+1$ would exceed Q^* only if the output Q_{t+1} is sufficiently high. In other words, moving from an initial state of $\widehat{Q_t} < Q^*$ to $\widehat{Q_{t+1}} > Q^*$ is much more costly when the state variable in stage $t+1$ is a convex combination of current and past outputs, rather than just the last one. So although the DM still strives to balance his current *state* and subsequent output to reach the stable effort level Q^*, only in the Persistent model it is feasible, cheaper, and therefore more efficient, to monotonically tend towards the stable level, thus discounting costs by exerting more effort in the future.

Remark 2. *Similarly to* Theorem 2*, one can extend the persistent model to a stochastic environment and prove that the results of* Theorem 3 *still hold, in expectation, under the stochastic extension. We omit this generalization due to its technical nature.*

4 Long- and short-term memory

In view of the two proposed evaluation mechanisms, one of the most interesting (and potentially most policy related) question is how much weight should be put on past performance. In this section we study how changes in the evaluation process affect the DM's optimal payoff, as well as the realized output.

To simplify the analysis, we consider the Persistent model, where the evaluation at stage t is given by $\widehat{Q_t} = (1-\lambda)\widehat{Q_{t-1}} + \lambda Q(e_t)$. The advantage of this set-up is its ability to summarize the trade-off between past and current performance through a single factor, λ. Namely, in Theorem 3 we showed that the DM's performance converges to a stable level Q^*. Now we can examine how changes in the evaluation process through λ, affect the long term output through the steady level Q^*. For example, if $\lambda = 0$, there is no value to future performance, and the steady level becomes Q_0 since the DM has no incentive to exert effort. That is, the system remains fixed to the initial condition and the DM produces the minimal feasible effort level. However, if $\lambda = 1$, then past performance is not taken into account during the evaluation process, and the DM repeatedly solves the optimization problem $\max_{e \in E}\{-e + R(Q(e))\}$.

Before formally stating the results, a few preliminary explanations and notations are needed. For every parameter $\lambda \in [0, 1]$, let Q^*_λ be the limit production level in the λ-discounted model under the optimal behavior described in Theorem 3. We assume that the DM acts optimally, using the optimal stationary strategy given λ, and production converges to Q^*_λ. In addition, denote the DM's optimal payoff by $\hat{\pi}_\lambda(Q_0)$, where Q_0 is the initial evaluation.

The first result of Theorem 4 relates to the monotonicity of Q^*_λ with respect to λ. We prove that the optimal strategy's stable output level, towards which production converges, is strictly increasing in λ. To put it another way, production increases as the evaluation becomes myopic of past output and the assessment depends more heavily on current performance.

On the one hand, from a strategic point of view this result is quite intuitive. When agents cannot rely on past performance, they constantly need to re-justify their abilities at every stage to come, exerting more effort in the process. The accumulated performance generates a certain *inheritance effect* where the ability to transfer value from one stage to another leads to less exertion of effort throughout the stages.

On the other hand, the same result also hints at an important economic observation. It implies that the first-best solution, where the DM exerts the maximal rational effort, is achievable only if the DM does not retain any past dependence. That is, the only possibility of exerting the optimal effort from the DM is by ignoring past results completely, at any given stage. This may pose a problem in various scenarios. For example, when such a process comes into play, the ability to screen low-level DMs is eliminated. Therefore, whenever uncertainty emerges between several DMs and their differential abilities (as in Holmstrom (1999)), the market needs to balance between the screening process (putting more weight on past performance) and optimal incentives (putting more weight on current ones).

The second part of Theorem 4 concerns the DM's payoff as a function of λ and Q_0. These results are best exemplified in Figure 1, showing inverse effects between the discount factor and the initial condition. Namely, fix a discount factor λ_0 and the appropriate limit production level $Q^*_{\lambda_0}$. Let us assume that the the initial output level Q_0 equals the limit level $Q^*_{\lambda_0}$, so exerting the same effort and output levels throughout the process is optimal. Now, ceteris paribus, consider a lower $\lambda < \lambda_0$ so that more weight is attributed to past performance. The DM evidently benefits from this adjustment since he can *free-ride* the initial condition, and reduce his effort levels. This suggests that the (limit) production level under λ is lower, while the DM's payoff increases! To see this more clearly, assume first that the DM does not adapt to the reduced discount factor λ, but maintains the optimal strategy σ_{λ_0}. Clearly, the DM's payoff does not change since the initial condition and output levels remain the same. However, this is not the DM's optimal payoff since he is exercising a sub-optimal strategy σ_{λ_0}, whereas the discount factor is $\lambda < \lambda_0$. Once the DM reverts to σ_λ, the expected payoff increases accordingly, while output levels decrease.

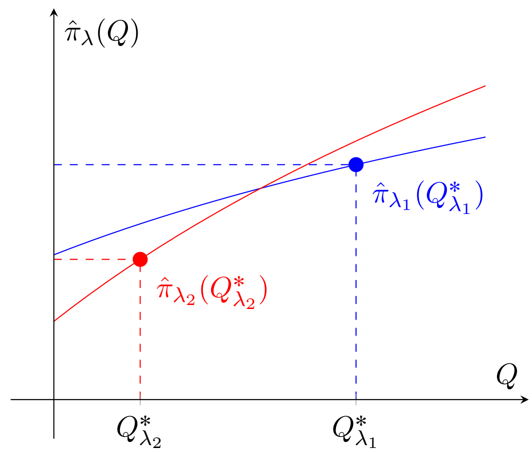

Figure 1: DM's expected payoff as a function of Q_0 and λ, where $\lambda_1 > \lambda_2$. Optimal-payoff functions are concave and continuously increasing in the initial position.

This reasoning yields a somewhat surprising outcome. It shows that *any evaluation factor* other than λ is preferable to the DM, given an initial position of $Q_0 = Q^*_\lambda$. That is, the evaluation factor that generates the lowest expected payoff, given some initial position, is the one that imposes the same steady level. As it appears, once production converges to a stable level, the DM can only profit from either an increased or a decreased evaluation factor, though the two generate inverse incentives.

Theorem 4. *Given the interior-solution property, the steady output level Q^*_λ strictly increases as a function of λ. Moreover, for every $\lambda_1 \neq \lambda_2$, and given an initial position of $Q^*_{\lambda_1}$, the* DM*'s payoff is higher under the λ_2-evaluation rather than under the λ_1-evaluation, i.e., $\hat{\pi}_{\lambda_2}\left(Q^*_{\lambda_1}\right) > \hat{\pi}_{\lambda_1}\left(Q^*_{\lambda_1}\right)$. In addition, If $\lambda_1 > \lambda_2$, then*

- $\hat{\pi}_{\lambda_2}(Q) > \hat{\pi}_{\lambda_1}(Q)$, *for every* $Q \geqslant Q^*_{\lambda_1}$*;*
- $\hat{\pi}_{\lambda_2}(Q) < \hat{\pi}_{\lambda_1}(Q)$, *for every* $Q \leqslant Q^*_{\lambda_2}$*;*
- $\hat{\pi}'_{\lambda_2}(Q) > \hat{\pi}'_{\lambda_1}(Q)$, *for every* $Q \leqslant Q^*_{\lambda_1}$.

Theorem 4 is best understood through Figure 1. First, if recent performance is weighted more heavily (i.e., $\lambda_1 > \lambda_2$), then incentives are sharpened such that: (i) output converges to a higher level of $Q^*_{\lambda_1} > Q^*_{\lambda_2}$; and (ii) dependence on the initial condition weakens, and the derivative w.r.t. Q_0 decreases. Next, in case the initial position Q_0 is low, the DM would prefer a lower evaluation of past production (blue line), quickly neglecting past performance, rather than a high evaluation of past production (red line). The opposite statement holds whenever the initial position is high.

The proof of Theorem 4, given in Appendix D, is based on Bellman's principle of optimality and the Contraction Mapping Theorem. We first define the standard contracting operator from the set of bounded functions to itself, according to the Bellman equation from Theorem 3. We then consider a closed set of bounded functions that maintain a certain property, e.g., one of the properties given in Theorem 3, and prove that the new operator-induced function preserves this property. Following the Contraction Mapping Theorem, we establish that the payoff function also maintains the given property, as needed.

5 Concluding remarks

In this paper we present a dynamic decision problem where the dependency of the compensation scheme over past performance produces an oscillatory production pattern. Though our setting is robust to the governing output-generating process, and though it underlines probabilistic environment, there are several important extensions to follow. First, to capture the oscillating performance in a general dynamic market one could accommodate for the interaction between multiple players. In light of the theoretical complexity behind such a model (as the dynamic interaction between DMs induces a stochastic game), the implementation of numerical analysis is imminent. Second, an extension of our model that associates incentives with screening, while taking into account the uncertainty regarding the agent's subjective abilities, is evident. We believe that the main obstacle lays in capturing the same phenomenon in a simple, yet general, static model. Lastly, a comprehensive analysis of our model while taking into account a wider range of evaluation processes, could produce important insights into the oscillating-performance problem.

References

Admati, Anat Ruth and Paul Pfleiderer, "Does it all add up? Benchmarks and the compensation of active portfolio managers," *Journal of Business*, 1997, *70* (3), 323–350. Publisher: University of Chicago Press.

Arrow, Kenneth J., Theodore Harris, and Jacob Marschak, "Optimal Inventory Policy," *Econometrica*, 1951, *19* (3), 250–272.

Arrow, Kenneth Joseph., Samuel. Karlin, and Herbert. Scarf, *Studies in the mathematical theory of inventory and production*, Stanford Univ. Pr, 1958.

Artés, Joaquín, "Long-run versus short-run decisions: R&D and market structure in Spanish firms," *Research Policy*, February 2009, *38* (1), 120–132.

Bailey, Martin J., "The Welfare Cost of Inflationary Finance," *Journal of Political Economy*, April 1956, *64* (2), 93–110.

Barras, Laurent, Olivier Scaillet, and Russ Wermers, "False Discoveries in Mutual Fund Performance: Measuring Luck in Estimated Alphas," *Journal of Finance*, 2010, *65*, 179–216. Publisher: WileyAmerican Finance Association.

Barro, Robert J., "A theory of monopolistic price adjustment," *Review of Economic Studies*, January 1972, *39* (1), 17–26.

Bellman, Richard E., Irving L. Glicksberg, and Oliver A. Gross, "On the Optimal Inventory Equation," *Management Science*, 1955, *2* (1), 83–104. Publisher: INFORMS.

Berk, Jonathan B. and Richard C. Green, "Mutual Fund Flows and Performance in Rational Markets," *Journal of Political Economy*, 2004, *112* (6), 1269–1295. Publisher: University of Chicago Press.

Bessler, Wolfgang, David Blake, Peter Lückoff, and Ian Tonks, "Fund Flows, Manager Changes, and Performance Persistence*," *Review of Finance*, August 2018, *22* (5), 1911–1947.

Blackwell, David, "Discounted Dynamic Programming," *Annals of Mathematical Statistics*, 1965, *36* (1), 226–235. Publisher: Institute of Mathematical Statistics.

Bollen, Nicolas P. B. and Jeffrey A. Busse, "Short-Term Persistence in Mutual Fund Performance," *Review of Financial Studies*, 2005, *18* (2), 569–597.

Brock, William A and JoséA Scheinkman, "Global asymptotic stability of optimal control systems with applications to the theory of economic growth," *Journal of Economic Theory*, February 1976, *12* (1), 164–190.

Brown, Stephen J. and William N. Goetzmann, "Performance Persistence," *Journal of Finance*, June 1995, *50* (2), 679–698. Publisher: WileyAmerican Finance Association.

Busse, Jeffrey A., Amit Goyal, and Sunil Wahal, "Performance and Persistence in Institutional Investment Management," *Journal of Finance*, 2010, *65*, 765–790. Publisher: WileyAmerican Finance Association.

Carhart, Mark M., "On Persistence in Mutual Fund Performance," *Journal of Finance*, March 1997, *52* (1), 57–82.

Cass, David and Karl Shell, "The structure and stability of competitive dynamical systems," *Journal of Economic Theory*, February 1976, *12* (1), 31–70.

Chen, Daniel L., Tobias J. Moskowitz, and Kelly Shue, "Decision Making Under the Gambler's Fallacy: Evidence from Asylum Judges, Loan Officers, and Baseball Umpires *," *Quarterly Journal of Economics*, August 2016, *131* (3), 1181–1242.

Cornell, Bradford, Jason Hsu, and David Nanigian, "Does Past Performance Matter in Investment Manager Selection?," *Journal of Portfolio Management*, July 2017, *43* (4), 33–43.

Dangl, Thomas, Youchang Wu, and Josef Zechner, "Market Discipline and Internal Governance in the Mutual Fund Industry," *Review of Financial Studies*, September 2008, *21* (5), 2307–2343.

Dvoretzky, Aryeh, Jack Kiefer, and Jacob Wolfowitz, "The Inventory Problem: I. Case of Known Distributions of Demand," *Econometrica*, 1952, *20* (2), 187–222.

_ , _ , and _ , "On the Optimal Character of the (s, S) Policy in Inventory Theory," *Econometrica*, 1953, *21* (4), 586–596.

Elton, Edwin J., Martin J. Gruber, Sanjiv Das, and Matthew Hlavka, "Efficiency with Costly Information: A Reinterpretation of Evidence from Managed Portfolios," *Review of Financial Studies*, January 1993, *6* (1), 1–22. Publisher: Arthur Wiesenberger and Co., New York.

Goetzmann, William N and Roger G Ibbotson, "Do Winners Repeat?," *Journal of Portfolio Management*, January 1994, *20* (2), 9–18. Publisher: Institutional Investor Journals.

Goyal, Amit and Sunil Wahal, "The Selection and Termination of Investment Management Firms by Plan Sponsors," *Journal of Finance*, 2008, *63*, 1805–1847. Publisher: WileyAmerican Finance Association.

Grinblatt, Mark and Sheridan Titman, "The Persistence of Mutual Fund Performance," *Journal of Finance*, December 1992, *47* (5), 1977–1984. Publisher: WileyAmerican Finance Association.

Hendricks, Darryll, Jayendu Patel, and Richard Zeckhauser, "Hot Hands in Mutual Funds: Short-Run Persistence of Relative Performance, 1974-1988," *Journal of Finance*, March 1993, *48* (1), 93–130. Publisher: WileyAmerican Finance Association.

Holmstrom, Bengt, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies*, January 1999, *66* (1), 169–182. Publisher: Oxford University Press.

Ishiguro, Shingo, "Management cycles," *Economic Theory*, February 2022, *73* (1), 257–300.

Lagziel, David and Ehud Lehrer, "Reward schemes," *Games and Economic Behavior*, 2018, *107*, 21–40. Publisher: Academic Press.

Lynch, Anthony W. and David K. Musto, "How Investors Interpret past Fund Returns," *Journal of Finance*, 2003, *58* (5), 2033–2058.

Rockafellar, Tyrrell R., "Saddle points of Hamiltonian systems in convex Lagrange problems having a nonzero discount rate," *Journal of Economic Theory*, February 1976, *12* (1), 71–113.

Scarf, Herbert, "The Optimality of (S, s) Policies in the Dynamic Inventory Problem," in Kenneth J. Arrow, Samuel Karlin, and Patrick Suppes, eds., *Mathematical Models in Social Sciences*, Stanford University Press, 1959.

Scheinkman, Alexandre J., "On optimal steady states of n-sector growth models when utility is discounted," *Journal of Economic Theory*, February 1976, *12* (1), 11–30.

Sheshinski, Eytan and Yoram Weiss, "Inflation and Costs of Price Adjustment," *Review of Economic Studies*, June 1977, *44* (2), 287–303.

Sheshinski, Eytan. and Yoram. Weiss, *Optimal pricing, inflation, and the cost of price adjustment*, MIT Press, 1993.

Stokey, Nancy L., Robert E. Lucas, and Edward C. Prescott, *Recursive Methods in Economic Dynamics*, Harvard University Press, 1989.

Stoughton, Neal M., "Moral Hazard and the Portfolio Management Problem," *Journal of Finance*, 1993, *48* (5), 2009.

Appendices

A The deterministic case - analysis and proofs

Eq. (1) produces the following sequential optimization problem

$$\hat{\pi}(e_0) = \sup_{\underline{e} \in E^\infty} \sum_{t=1}^{\infty} \beta^{t-1}[-e_t + R(Q_{t-1} + Q_t)], \tag{3}$$

where $\hat{\pi}$ is an optimal payoff given initial effort level $e_0 \in E$. We follow the standard analysis of the corresponding Bellman equation,

$$\hat{\pi}(e_0) = \sup_{e \in E} \left[-e + R(Q_0 + Q(e)) + \beta\hat{\pi}(e)\right]. \tag{4}$$

where the existence of the value function $\hat{\pi}$ follows from the properties of E (non-empty, compact, and convex), and from the continuity of Q and R, which suggests that the single-period payoffs are bounded.

At this point we follow the results given in Chapter 4 of Stokey et al. (1989). To facilitate readability and to not repeat similar arguments in subsequent proofs of Theorems 2 and 3, we elaborate in the current proof regarding our use of the aforementioned results. First, in terms of notations, the set of endogenous state variables X and the action set Y and $\Gamma(X)$, in Stokey et al. (1989), translate to our action set E. Second, the single-stage action (effort) set E is non-empty, compact, and convex, and the corresponding payoff (output) functions are continuous, therefore the single-stage payoffs are bounded. This ensures the existence of the value function $\hat{\pi}$ in Eq. (4). To formally support this claim, we refer to Chapter 4 of Stokey et al. (1989), specifically to Assumptions 4.1 − 4.4 and Theorems[12] SLP-4.3, SLP-4.4, and SLP-4.5. Their analysis is based on considering the (more basic) n-stage discounted problem, for which the solution evidently exists, and then taking the limit of the value function as n tends to infinity.

Define the correspondence $\sigma : E \to 2^E$ such that

$$\sigma(e_0) = \{e \in E \mid \hat{\pi}(e_0) = F(e_0, e) + \beta\hat{\pi}(e)\},$$

where $F(e_0, e) = -e + R(Q(e_0) + Q(e))$ for every $e_0 \in E$. The function F follows the notation of Chapter 4, Section 4.1, in Stokey et al. (1989), where F is the single-stage payoff given an initial state of e_0 and an action e. Moreover, the correspondence $\sigma(e_0)$ is defined according to the equivalent (optimal) *policy correspondence* $G(x)$ of Stokey et al. (1989) (see Eq.(2) in p.78). Since all functions are bounded and continuous, σ is well defined. Moreover, Theorem SLP-4.6 proves that σ is a compact-valued, upper hemi-continuous correspondence, that generates the DM's optimal strategy (the proof of Theorem 4.6

[12]Hereafter, we refer to Theorems 4.2 − 4.11 in p.71 − 85 of Stokey et al. (1989) as SLP-4.XX.

is based on Blackwell's sufficient condition for a contraction mapping, and the Contraction Mapping Theorem). The functions Q and R are monotonic and concave, so $F(e_0, e)$ is concave w.r.t. e_0 and e, and strictly concave w.r.t. e_0, thus the conditions (Assumptions $4.3 - 4.4$ and $4.7 - 4.8$) to SLP-4.8 are met, and σ is a single-valued continuous function according to this theorem. In addition, the fact that F is strictly increasing in e_0 implies that $\hat{\pi}$ is strictly increasing as well (see SLP-4.7).

Proving that $\hat{\pi}$ is differentiable (using SLP-4.11) requires $\sigma(E)$ to be the interior points of E, which holds by the interior-solution property. Hence, $\hat{\pi}$ is continuously differentiable and we can use the envelope theorem to differentiate Eq. (4) w.r.t. e and evaluate the derivative in $\sigma(e_0)$ to get

$$R'(Q(e_0) + Q(\sigma(e_0)))Q'(\sigma(e_0)) + \beta\hat{\pi}'(\sigma(e_0)) = 1. \tag{5}$$

Eq. (5) enables us to study the properties of σ. We start with monotonicity. Consider a small increase of e_0 to $e_0 + \varepsilon > e_0$. If $\sigma(e_0 + \varepsilon) \geqslant \sigma(e_0)$, then the LHS of Eq. (5) decreases, violating the equality, since R', Q', and $\hat{\pi}'$ are (non-negative) decreasing functions due to the strict concavity of R, Q, and $\hat{\pi}$. Thus, σ is a strictly decreasing continuous function from E to itself, and it has a unique fixed point $e^* \in E$ (also an interior point of E) such that $R'(Q(e^*) + Q(e^*))Q'(e^*) + \beta\hat{\pi}'(e^*) = 1$.

The next step of our analysis shows that, for every $e_0 \in E$, the sequences $\left(\sigma^{2n}(e_0)\right)_{n\in\mathbb{N}}$ and $\left(\sigma^{2n+1}(e_0)\right)_{n\in N}$ monotonically converge to e^*, as n tends to infinity. In addition, the uniqueness of e^*, along with the continuity of σ, imply that e^* is bounded between $\sigma^n(e_0)$ and $\sigma^{n+1}(e_0)$ for every n. That is, we show that the repeated use of σ tends monotonically to the fixed point e^*, and $\sigma^n(e_0)$ oscillates around e_0 as a function of n, with an amplitude converging to 0.

Now define $H(x, y) = R'(Q(x) + Q(y))Q'(y) + \beta\hat{\pi}'(y)$ and note that the concavity of Q and $\hat{\pi}$ imply that $H(x, y) > H(y, x)$ for every $y < x$ where $x, y \in E$. In addition, H is continuous and strictly decreasing in both coordinates. Therefore, H satisfies the conditions of the following Lemma 1.

Lemma 1. *Let $H : E^2 \to \mathbb{R}$ be a continuous function, strictly decreasing in both coordinates, such that*

$$H(x, y) > H(y, x), \tag{6}$$

for every $y < x$ where $x, y \in E$. Then,

1. *there exists $c \in \mathbb{R}$ such that for all $x \in E$, the eq. $H(x, y) = c$ has a unique solution, $y_c(x)$;*

2. *the function y_c is a continuous, strictly decreasing function with a unique fixed point x_c;*

3. *for every $x \in E$, the sequences $\left(y_c^{2n}(x)\right)_{n\in\mathbb{N}}$ and $\left(y_c^{2n+1}(x)\right)_{n\in\mathbb{N}}$ monotonically tend to x_c, where $y_c^k(x) = y_c(y_c^{k-1}(x))$ for every $k > 1$;*

4. the fixed point x_c is bounded between $y_c^n(x)$ and $y_c^{n+1}(x)$ for every $n \in \mathbb{N}$ and every $x \in E$.

Proof. Ineq. (6) implies that $H(e_{\max}, e_{\min}) > H(e_{\min}, e_{\max})$ and we can fix c between these two values of H. It follows from the strict monotonicity of H that $H(x, e_{\min}) > c > H(x, e_{\max})$ for every $x \in E$. By continuity, for every $x \in E$ there exists a solution $y_c(x)$ for $H(x, y) = c$, and strict monotonicity suggests that $y_c(x)$ is unique. In addition, the same two properties of H imply that y_c is continuous and strictly decreasing. Moreover, y_c is defined from E to E, and therefore has a unique fixed point, denoted x_c. We conclude that $H(x_c, y_c(x_c)) = H(x_c, x_c) = c$.

Fix $x \in E$ such that $x > x_c$. Since $H(x_c, y_c(x_c)) = c$ where H is strictly decreasing, we deduce that $y_c(x) < y_c(x_c) = x_c$. Assume, contrary to the stated lemma, that $y_c^2(x) = x$. Then,

$$c = H(y_c(x), y_c^2(x)) = H(y_c(x), x) < H(x, y_c(x)) = c,$$

where the inequality follows from Ineq. (6). A contradiction. Since $H(y_c(x), x) < c$, we conclude that $x_c < y_c^2(x) < x$, as required for the result to hold. A similar proof holds for $x > x_c$. Hence, we can now consider the sequences $\left(y_c^{2n}(x)\right)_{n\in\mathbb{N}}$ and $\left(y_c^{2n+1}(x)\right)_{n\in\mathbb{N}}$ bounding x_c. Each of the two sequences tends closer to x_c as n grows. Since both are monotonic and bounded, they converge. Assume, by contradiction, that a sequence, e.g., $\left(y_c^{2n+1}(x)\right)_{n\in\mathbb{N}}$, converges to $x'_c \neq x_c$. Then,

$$x'_c = \lim_{n\to\infty} y_c^{2n+3}(x) = \lim_{n\to\infty} y_c^2\left(y_c^{2n+1}(x)\right) = y_c^2\left(\lim_{n\to\infty} y_c^{2n+1}(x)\right) = y_c^2(x'_c),$$

contradicting the strict monotonicity of the sequences when $x \neq x_c$, and concluding the proof. ■

To conclude, consider the previously defined function $H(x, y) = R'(Q(x) + Q(y))Q'(y) + \beta\hat{\pi}'(y)$, and note that it sustains all the conditions of Lemma 1. Fix $c = 1$ and substitute x and y by e_0 and σ, respectively. The previous analysis, prior to Lemma 1, shows that σ has a unique fixed point e^* given $c = 1$, and Lemma 1 ensures that an iteration of σ cyclically converges to the fixed point e^*. Thus, we conclude the proof of Theorem 1.

B The stochastic case - analysis and proof

Similarly to the previous analysis, we transform the optimization problem derived from Eq. (2) to the following Bellman equation

$$\pi(\omega_0, Q_0) = \sup_{e\in E} \mathbf{E}_{\omega_0}\left[-e + R(Q_0 + Q(\tilde{\omega}, e)) + \beta\pi(\tilde{\omega}, Q(\tilde{\omega}, e))\right], \tag{7}$$

where the expectation relates to the random variable $\tilde{\omega}$, drawn according to P and ω_0.

Proof of Theorem 2.

To prove Theorem 2, we follow Chapter 9 of Stokey et al. (1989), specifically using Assumptions $9.4 - 9.12$ and $9.16 - 9.17$. To facilitate this for the reader, note that the DM's actions set E is fixed throughout the stages in our model, therefore it is equivalent to $\Gamma(X, Z)$, as well as Y, in the notations of Stokey et al. (1989). In addition, in their notation, the set of endogenous state variables X is equivalent to our set S, which is the convex hull of all possible outputs $Q(\Omega, E)$, and their set Z of exogenous state variables translates to our finite set Ω of states. Let us go through the relevant basic assumptions in Chapter 9 of Stokey et al. (1989), that are needed for the current proof:

- Assumption 9.4 requires the set S to be a convex Borel set.
- Assumption 9.5 relies on the assumption that Ω is countable.
- Assumptions $9.6, 9.9, 9.11$ and 9.16 all refer to the DM's actions set E which is a fixed convex interval in our set-up, thus independent of Ω. As a result, all these requirements are met.
- Assumption 9.7 relates to the expected single-stage payoff function

  $$F(\omega_0, Q_0, e) = \mathbf{E}_{\omega_0, Q_0}\left[-e + R(Q_0 + Q(\tilde{\omega}, e))\right],$$

  stating that F is bounded and continuous, for every $\omega \in \Omega$. This assumption holds according to the properties of R and Q (similarly to Assumption 4.4 used in the proof of Theorem 1);
- Assumptions 9.8 and 9.10 require F to be concave w.r.t. Q_0 and e and increasing in Q_0. These assumptions hold similarly to Assumptions 4.5 and 4.7 used in Appendix A.
- Assumption 9.12 requires F to be differentiable w.r.t. Q_0 and e, given $\omega \in \Omega$, equivalently to Assumption 4.9 used in the proof of Theorem 1. Evidently, this assumption holds for similar reasons.
- Assumption 9.17 requires that that the function $Q(\omega, e)$ is continuous in e, for every given $\omega \in \Omega$.

It is important to note that we repeatedly use the linearity of the expectation operator, mainly in the context of the function F, to extend the aforementioned assumptions from Appendix A to the current stochastic setting. To sum up, Assumptions $9.4 - 9.12$ and $9.16 - 9.17$ stated in Chapter 9 of Stokey et al. (1989) hold in our setting, so we can apply their results in this section. Specifically, we follow SLP-9.6 through SLP-9.10, and use Exercise 9.7 which extends these theorems to our setting:

- SLP-9.6 proves that the optimal policy is a non-empty, compact-valued, upper hemi-continuous correspondence.

- SLP-9.7 states that the value function is increasing in the initial condition Q_0, given ω_0.

- SLP-9.8 establishes that the value function is strictly concave in Q_0, given ω_0, and that the optimal policy is a continuous, single-valued, function.

- SLP-9.10 ensures that the value function is differentiable in Q_0, given ω_0.

In addition, the output in stage t depends solely on the realized action and the realized state in stage $t-1$, thus the required conditions in Exercise 9.7 in Stokey et al. (1989) are met, (with the exception of $9.7-g$ which is irrelevant for the current theorem) and the result follows by the interior-solution property. To sum-up, based on previous results, there exists a continuously increasing, strictly concave, differentiable payoff function $\hat{\pi}(\omega_0, Q_0)$ (all w.r.t. Q_0), and there exists a unique, stationary, and continuous optimal-strategy $\sigma : \Omega \times S \to E$.

By the differentiability of the RHS of Eq. (7) w.r.t. e, we use the envelope theorem and plug-in $\sigma(\omega_0, Q_0)$ after taking the FOC, to get

$$0 = \mathbf{E}_{\omega_0}\left[-1 + \left(R'(Q_0 + Q(\tilde{\omega}, e)) + \beta \frac{\partial \pi(\tilde{\omega}, Q(\tilde{\omega}, e))}{\partial Q(\tilde{\omega}, e)}\right) \frac{\partial Q(\tilde{\omega}, e)}{\partial e}\right]_{e=\sigma(\omega_0, Q_0)}. \tag{8}$$

The monotonicity and concavity of the output function, the reward function, and the payoff function imply that the derivatives on the RHS decrease when either Q_0 or $\sigma(\omega_0, Q_0)$ increase. Thus, an increase in Q_0 must follow a decrease in $\sigma(\omega_0, Q_0)$ to maintain Eq. (8). Hence, $\mathbf{E}_{\sigma,\omega_{t-1}}[Q_t|Q_{t-1}] = \mathbf{E}_{\omega_{t-1}}[Q(\tilde{\omega}, \sigma(\omega_{t-1}, Q_{t-1})]$, and $\sigma(\omega_{t-1}, Q_{t-1})$ decreases w.r.t Q_{t-1}.

To prove a cyclic performance, we start with the simple case where the state variable is absorbed, with probability (w.p.) 1, to some fixed state $\omega \in \Omega$. In such a case, the proof of Theorem 1 holds and a cyclic performance follows. Otherwise, assume w.l.o.g. that the chain is irreducible. For every $\omega \in \Omega$, consider the function $\psi_\omega(q) = \mathbf{E}_\omega\left[Q(\tilde{\omega}, \sigma(\omega, q))\right]$. By the continuity and monotonicity of σ along with the compactness assumption over $Q(E)$, there exists a unique fixed point q_ω such that $\psi_\omega(q_\omega) = q_\omega$. Since the Markov chain is finite and irreducible, we can take the stationary distribution μ and define $Q^* = \mathbf{E}_\mu[q_{\tilde{\omega}}]$, where the expectation is taken w.r.t. μ.

Fix $\overline{\omega}, \underline{\omega} \in \Omega$ such that $q_{\overline{\omega}} > q_\omega > q_{\underline{\omega}}$, for every $\omega \in \Omega \backslash \{\overline{\omega}, \underline{\omega}\}$. We will show that for every $\varepsilon > 0$, w.p. 1, every trajectory visits the two intervals $(-\infty, q_{\underline{\omega}} + \varepsilon]$ and $[q_{\overline{\omega}} - \varepsilon, \infty)$ infinitely many times, thus oscillating around Q^* as needed. The idea behind this statement is that both $\overline{\omega}$ and $\underline{\omega}$ are visited infinitely many times, and whenever the realized output is within $[q_{\underline{\omega}}, q_{\overline{\omega}}]$, the expected production in the subsequent period is outside $[q_{\underline{\omega}}, q_{\overline{\omega}}]$. Namely, the monotonicity of σ implies that for every state ω, the inequality $\psi_\omega(q) > q_\omega$ holds if and only if $q < q_\omega$. Meaning, a realized production below (or above) the fixed point q_ω guarantees that the next-stage's expected output is above (or below, resp.) the fixed point. In other words, the output *oscillates in expectation.*

Fix a small $\varepsilon > 0$ such that $Q^* \in (q_{\underline{\omega}} + \varepsilon, q_{\overline{\omega}} - \varepsilon)$. The compactness of $Q(E)$ along with the oscillation-in-expectation property guarantees that there exists $\delta > 0$ such that $\Pr(Q(\tilde{\omega}, \sigma(\overline{\omega}, q)) > q_{\overline{\omega}}) > \delta$ for every $q < q_{\overline{\omega}}$ and, equivalently, $\Pr(Q(\tilde{\omega}, \sigma(\underline{\omega}, q)) < q_{\underline{\omega}}) > \delta$ for every $q > q_{\underline{\omega}}$. We will now turn to a proof by contradiction.

Denote $I = [q_{\underline{\omega}} + \varepsilon, q_{\overline{\omega}} - \varepsilon]$ and assume there is a positive probability event $D = \bigcup_{t \in \mathbb{N}} D_t$ where D_t includes all histories such that the realized output from stage t onwards is in I. Since D has positive probability, there exists $D_T \subseteq D$ with positive probability, and a positive-probability finite history h, of length greater than T stages, such that $\Pr(D_T|h) > 1 - \delta$. Now consider all continuations of h. Each continuation h' settles in $\overline{\omega}$ infinitely often. Let $\tau[h']$ be the first stage, after h, where $\overline{\omega}$ is the state variable according to a continuation h'. The construction implies that $Q_{\tau[h']} \in I$, and specifically $Q_{\tau[h']} < q_{\overline{\omega}}$. By the previous statement, we know that for every $Q_{\tau[h']} < q_{\overline{\omega}}$,

$$\Pr\left(Q\left(\tilde{\omega}, \sigma\left(\omega_{\tau[h']}, Q_{\tau[h']}\right)\right) > q_{\overline{\omega}} \middle| \, h, \omega_{\tau[h']} = \overline{\omega}\right) > \delta.$$

Summing up over all stages $\tau(h')$, we get that $\Pr(\overline{D_T}|h) > \delta$, contradicting the initial assumption that $\Pr(D_T|h) > 1 - \delta$ and concluding the proof. ■

C Proof of Theorem 3

Proof. In this proof we follow the same analysis presented in Appendix A. However, to simplify the notation, we revert to the auxiliary problem where the DM's action set is $Q(E)$ instead of E, and the DM's t-stage payoff is given by $-Q^{-1}(Q_t) + R((1-\lambda)\widehat{Q_{t-1}} + \lambda Q_t)$. To clarify, as stated in Section 2, the stage payoff $-e_t + R((1-\lambda)\widehat{Q_{t-1}} + \lambda Q(e_t))$ is equivalent to $-Q^{-1}(Q_t) + R((1-\lambda)\widehat{Q_{t-1}} + \lambda Q_t)$, where in the former expression the DM chooses the t-stage effort level e_t, whereas in the latter the DM's chooses the t-stage output level Q_t. The equivalence follows from the one-to-one mapping between E and $Q(E)$.

Denote the initial position by $Q_0 \in Q(E)$. Therefore, the equivalent functional equation to Eq. (4) becomes

$$\hat{\pi}(Q_0) = \sup_{q \in Q(E)} \left[-Q^{-1}(q) + R((1-\lambda)Q_0 + \lambda q) + \beta \hat{\pi}((1-\lambda)Q_0 + \lambda r)\right], \tag{9}$$

where the DM chooses a production level q, receives a payoff of $-Q^{-1}(q) + R((1-\lambda)Q_0 + \lambda q)$, and moves on to the next stage with evaluation $(1-\lambda)Q_0 + \lambda q$. Note that Q^{-1} is the inverse function of Q, and therefore strictly increasing, strictly convex and continuously differentiable.

By the properties of E, Q, and R we can use SLP-4.2, SLP-4.3, and SLP-4.6 (similarly to Theorem 1) to prove the existence, uniqueness, and continuity of $\hat{\pi}$. Re-define the correspondence $\sigma : Q(E) \rightarrow$

$2^{Q(E)}$ such that

$$\sigma(Q_0) = \{q \in Q(E) \mid \hat{\pi}(q) = F(Q_0, q) + \beta\hat{\pi}((1-\lambda)Q_0 + \lambda q)\},$$

where $F(Q_0, q) = -Q^{-1}(q) + R((1-\lambda)Q_0 + \lambda q)$. Theorems SLP-4.4, SLP-4.5, and SLP-4.6 prove that σ is a compact-valued, upper hemi-continuous correspondence, that generates the DM's optimal strategy.

To show that σ is a single-valued continuous function, we need to prove that $F(Q_0, q)$ is concave w.r.t. Q_0 and q, and strictly concave w.r.t. Q_0 (see SLP-4.8). By the strict convexity of Q^{-1} and by the same analysis as in Appendix A, the concavity condition of F holds and σ is a single-valued continuous function, while the value function $\hat{\pi}$ is strictly concave. In addition, the fact that F is strictly increasing in Q_0 implies that the value function $\hat{\pi}$ is also strictly increasing (see SLP-4.7).

The interior-solution property and SLP-4.11 prove that the value function is continuously differentiable, and by the envelope theorem we can follow the analysis of Chapter 4 in Stokey et al. (1989), to write down the following FOC of the Bellman equation,

$$0 = -\frac{1}{\lambda Q'(Q^{-1}(\sigma(Q_0)))} + R'((1-\lambda)Q_0 + \lambda\sigma(Q_0)) + \beta\hat{\pi}'((1-\lambda)Q_0 + \lambda\sigma(Q_0)),$$

or equivalently,

$$\lambda Q'(Q^{-1}(\sigma(Q_0)))\left[R'((1-\lambda)Q_0 + \lambda\sigma(Q_0)) + \beta\hat{\pi}'((1-\lambda)Q_0 + \lambda\sigma(Q_0))\right] = 1 \tag{10}$$

Next, consider a small increase of Q_0 to $Q_0 + \varepsilon > Q_0$. If $\sigma(Q_0 + \varepsilon) \geqslant \sigma(Q_0)$, then the LHS of the last equation decreases (since Q, R, and $\hat{\pi}$ are concave), violating the equality. Hence, we proved that σ is a strictly decreasing continuous function from $Q(E)$ to $Q(E)$, thus it has a unique, interior, fixed point $Q^* \in Q(E)$ such that

$$\lambda Q'(Q^{-1}(Q^*))\left[R'(Q^*) + \beta\hat{\pi}'(Q^*)\right] = 1. \tag{11}$$

Combining the last two equations (the FOC equality and the last fixed-point equality) yields

$$\lambda Q'(Q^{-1}(Q^*))\left[R'(Q^*) + \beta\hat{\pi}'(Q^*)\right] = \lambda Q'(Q^{-1}(\sigma(Q_0)))\left[R'(\hat{Q}) + \beta\hat{\pi}'(\hat{Q})\right],$$

where $\hat{Q} = (1-\lambda)Q_0 + \lambda\sigma(Q_0)$.

Assume $Q_0 < Q^*$. The monotonicity of σ implies that $\sigma(Q_0) > Q^*$, and so $Q'(Q^{-1}(Q^*))) > Q'(Q^{-1}(\sigma(Q_0)))$. Therefore, it follows from the Equation (11) that $R'(\hat{Q}) + \beta\hat{\pi}'(\hat{Q}) > R'(Q^*) + \beta\hat{\pi}'(Q^*)$, or equivalently $Q^* > (1-\lambda)Q_0 + \lambda\sigma(Q_0) > Q_0$. In words, we showed that an initial position of $Q_0 < Q^*$ imposes a production above Q^* in the subsequent stage, while maintaining the subsequent position below Q^*. By induction, the same result applies in every stage to follow. Symmetrically, one reaches a similar conclusion given $Q_0 > Q^*$, and we derive that the sequence $(\widehat{Q_t})_{n\in\mathbb{N}}$ generated by σ and Q_0, monotonically converges to Q^*. ■

D Proof of Theorem 4

Proof. To simplify the proof, we use the same notation as in the proof of Theorem 3 where the relevant Bellman equation is given by Eq. (9),

$$\hat{\pi}_\lambda(Q_0) = \sup_{q \in Q(E)} \left[-Q^{-1}(q) + R((1-\lambda)Q_0 + \lambda q) + \beta \hat{\pi}((1-\lambda)Q_0 + \lambda q) \right],$$

such that Q^{-1} is the (strictly increasing and convex) inverse function of Q. To use Bellman's principle of optimality and Blackwell's Contraction Mapping Theorem, we need to find a contracting operator from the set of bounded functions to itself. Let B be the set of bounded real-valued functions over $Q(E)$. For every λ, define the operator $T_\lambda : B \to B$ such that

$$(T_\lambda f)(Q_0) = \max_{q \in Q(E)} \left[-Q^{-1}(q) + R((1-\lambda)Q_0 + \lambda q) + \beta f((1-\lambda)Q_0 + \lambda q) \right],$$

for every $Q_0 \in Q(E)$. This operator, along with the results of Chapter 4 of Stokey et al. (1989), was used implicitly to prove Theorem 3, and will be similarly used in the current proof.

The proof is divided into five parts with respect to the different parts of the theorem:

Part I proves $\hat{\pi}_{\lambda_2}\left(Q^*_{\lambda_1}\right) \geqslant \hat{\pi}_{\lambda_1}\left(Q^*_{\lambda_1}\right)$, for every two discount factors $\lambda_1 \neq \lambda_2$ such that $Q_0 = Q^*_{\lambda_1}$.

Part II proves $\hat{\pi}_{\lambda_2}(Q_0) > \hat{\pi}_{\lambda_1}(Q_0)$, for every two discount factors $\lambda_1 > \lambda_2$ such that $Q_0 > Q^*_{\lambda_1}$.

Part III proves $\hat{\pi}_{\lambda_2}(Q_0) < \hat{\pi}_{\lambda_1}(Q_0)$, for every two discount factors $\lambda_1 > \lambda_2$ such that $Q_0 < Q^*_{\lambda_2}$.

Part IV proves Q^*_λ is strictly increasing in λ.

Part V proves Part II and Part III for the cases where $Q_0 = Q^*_{\lambda_1}$ and $Q_0 = Q^*_{\lambda_2}$, respectively. Applying Parts II, III, and V to any λ and with respect to higher and lower discount factors produces the desired result.

Part I. Since $Q_0 = Q^*_{\lambda_1}$ is a fixed point of the λ_1-evaluation problem, the DM will repeatedly generate an output of $Q^*_{\lambda_1}$ and a payoff of $\hat{\pi}_{\lambda_1}(Q^*_{\lambda_1})$. Thus, for any function $f \in B$ such that $f(Q^*_{\lambda_1}) \geqslant \pi_{\lambda_1}(Q^*_{\lambda_1})$, it follows that

$$\begin{aligned}
(T_{\lambda_2} f)(Q^*_{\lambda_1}) &\geqslant -Q^{-1}(Q^*_{\lambda_1}) + R(Q^*_{\lambda_1}) + \beta f(Q^*_{\lambda_1}) \\
&\geqslant -Q^{-1}(Q^*_{\lambda_1}) + R(Q^*_{\lambda_1}) + \beta \pi_{\lambda_1}(Q^*_{\lambda_1}) \\
&= \hat{\pi}_{\lambda_1}(Q^*_{\lambda_1}),
\end{aligned}$$

where the first inequality follows from substituting the optimal q with $Q^*_{\lambda_1}$, and the second inequality follows from the assumption over f. By Bellman's principle of optimality and Blackwell's Contraction

Mapping Theorem, along with the fact that the set of bounded functions that sustain the condition $f(Q^*_{\lambda_1}) \geqslant \pi_{\lambda_1}(Q^*_{\lambda_1})$ is closed, it follows that $\hat{\pi}_{\lambda_2}(Q^*_{\lambda_1}) \geqslant \hat{\pi}_{\lambda_1}(Q^*_{\lambda_1})$, as needed.

Part II. Fix a function $f \in B$ such that $f(q) \geqslant \hat{\pi}_{\lambda_1}(q)$ for every $q > Q^*_{\lambda_1}$. By Theorem 3, we know that $Q_0 > Q^*_{\lambda_1}$ implies $\sigma_{\lambda_1}(Q_0) < Q^*_{\lambda_1}$, where σ_{λ_1} is the optimal stationary strategy in the λ_1-evaluation problem, such that the position in the next stage tends towards $Q^*_{\lambda_1}$ from above. Hence,

$$
\begin{aligned}
(T_{\lambda_2} f)(Q_0) \;&\geqslant\; -Q^{-1}(\sigma_{\lambda_1}(Q_0)) + R((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_1}(Q_0)) + \beta f((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_1}(Q_0)) \\
&\geqslant\; -Q^{-1}(\sigma_{\lambda_1}(Q_0)) + R((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_1}(Q_0)) + \beta \hat{\pi}_{\lambda_1}((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_1}(Q_0)) \\
&>\; -Q^{-1}(\sigma_{\lambda_1}(Q_0)) + R((1-\lambda_1)Q_0 + \lambda_1\sigma_{\lambda_1}(Q_0)) + \beta \hat{\pi}_{\lambda_1}((1-\lambda_1)Q_0 + \lambda_1\sigma_{\lambda_1}(Q_0)) \\
&=\; \hat{\pi}_{\lambda_1}(Q_0),
\end{aligned}
$$

where the first inequality follows from substituting the optimal q with $\sigma_{\lambda_1}(Q_0)$, the second inequality follows from the assumption over f, and the third inequality follows from the monotonicity of $\hat{\pi}_\lambda$ (w.r.t. λ) and of R. Since the set of functions f that sustain the required condition is closed, and following the Contraction Mapping Theorem, the statement of Part II holds.

Part III. Similarly to Part II, fix a function $f \in B$ such that $f(q) \geqslant \hat{\pi}_{\lambda_2}(q)$ for every $q < Q^*_{\lambda_2}$. By Theorem 3, we know that $Q_0 < Q^*_{\lambda_2}$ implies $\sigma_{\lambda_2}(Q_0) > Q^*_{\lambda_2}$, where σ_{λ_2} is the optimal stationary strategy in the λ_2-evaluation problem, such that the position in the next stage tends towards $Q^*_{\lambda_2}$ from below. Hence,

$$
\begin{aligned}
(T_{\lambda_1} f)(Q_0) \;&\geqslant\; -Q^{-1}\left(\frac{\lambda_1-\lambda_2}{\lambda_1}Q_0 + \frac{\lambda_2}{\lambda_1}\sigma_{\lambda_2}(Q_0)\right) + R((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_2}(Q_0)) \\
&+\; \beta f((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_2}(Q_0)) \\
&\geqslant\; -Q^{-1}\left(\frac{\lambda_1-\lambda_2}{\lambda_1}Q_0 + \frac{\lambda_2}{\lambda_1}\sigma_{\lambda_2}(Q_0)\right) + R((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_2}(Q_0)) \\
&+\; \beta \hat{\pi}_{\lambda_2}((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_2}(Q_0)) \\
&>\; -Q^{-1}(\sigma_{\lambda_2}(Q_0)) + R((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_2}(Q_0)) + \beta \hat{\pi}_{\lambda_2}((1-\lambda_2)Q_0 + \lambda_2\sigma_{\lambda_2}(Q_0)) \\
&=\; \hat{\pi}_{\lambda_2}(Q_0),
\end{aligned}
$$

where the first inequality follows from substituting the optimal q with $\frac{\lambda_1-\lambda_2}{\lambda_1}Q_0 + \frac{\lambda_2}{\lambda_1}\sigma_{\lambda_2}(Q_0)$, the second inequality follows from the assumption over f, and the third inequality follows from the monotonicity of Q^{-1}. Since the set of functions f sustaining the required condition is closed, and following the Contraction Mapping Theorem, the statement of Part III holds.

Part IV. Assume that $Q^*_{\lambda_1} < Q^*_{\lambda_2}$ for $0 < \lambda_2 < \lambda_1 < 1$ (where the result for the end points is trivial). Take $q \in (Q^*_{\lambda_1}, Q^*_{\lambda_2})$. According to Parts II and III, we get $\hat{\pi}_{\lambda_2}(q) > \hat{\pi}_{\lambda_1}(q) > \hat{\pi}_{\lambda_2}(q)$. A contradiction. Thus, $Q^*_{\lambda_1} \geqslant Q^*_{\lambda_2}$, for $\lambda_1 > \lambda_2$.

Now assume that $Q^*_{\lambda_1} = Q^*_{\lambda_2}$ for $\lambda_2 < \lambda_1$. We can take the FOC of the RHS of the stated Bellman equation (similarly to Theorem 3), along with the derivative of $\hat{\pi}_{\lambda_1}(Q_0)$, to get the two equations,

$$\lambda \left[R'((1-\lambda)Q_0 + \lambda\sigma(Q_0)) + \beta\hat{\pi}'((1-\lambda)Q_0 + \lambda\sigma(Q_0))\right] = \left(Q^{-1}\right)'(\sigma(Q_0))$$

and

$$\hat{\pi}'(Q_0) = (1-\lambda)R'((1-\lambda)Q_0 + \lambda\sigma(Q_0)),$$

where the second equality follows from the envelope theorem. Taking $\lambda = \lambda_1$, $Q_0 = Q^*_{\lambda_1}$, and plugging the second equation into the first yields

$$\lambda_1\left[1 + \beta(1-\lambda_1)\right] = \frac{\left(Q^{-1}\right)'\left(Q^*_{\lambda_1}\right)}{R'\left(Q^*_{\lambda_1}\right)}.$$

Since $\beta \in (0,1)$, the LHS is an increasing function of λ_1, subject to $0 \leqslant \lambda_1 \leqslant 1$. Thus, $Q^*_{\lambda_1} = Q^*_{\lambda_2}$ contradicts the last equality, implying $Q^*_{\lambda_1} > Q^*_{\lambda_2}$, as needed.

Part V. We only prove the relevant case of Part II where $Q_0 = Q^*_{\lambda_1}$, while a similar proof holds for $Q_0 = Q^*_{\lambda_2}$ of Part III. Consider $\lambda_1 > \lambda_2$ and fix $\lambda_3 \in (\lambda_2, \lambda_1)$. According to Part IV, $Q^*_{\lambda_1} > Q^*_{\lambda_3} > Q^*_{\lambda_2}$. Hence by Part II, $\hat{\pi}_{\lambda_2}(Q^*_{\lambda_1}) > \hat{\pi}_{\lambda_3}(Q^*_{\lambda_1})$, whereas by Part I, $\hat{\pi}_{\lambda_3}(Q^*_{\lambda_1}) \geqslant \hat{\pi}_{\lambda_1}(Q^*_{\lambda_1})$, which concludes Part V.

Next, we prove the last statement of the theorem regarding the derivatives. If $Q^*_{\lambda_2} \leqslant Q \leqslant Q^*_{\lambda_1}$, then $\sigma_{\lambda_2}(Q) < Q < \sigma_{\lambda_1}(Q)$, and $(1-\lambda)Q + \lambda\sigma_\lambda(Q)$ increase with λ. If $Q < Q^*_{\lambda_2}$, we consider two cases where either $\sigma_{\lambda_1}(Q) \leqslant \sigma_{\lambda_2}(Q)$ or $\sigma_{\lambda_1}(Q) > \sigma_{\lambda_2}(Q)$. Assume that $\sigma_{\lambda_1}(Q) \leqslant \sigma_{\lambda_2}(Q)$. Thus, $\lambda Q'(Q^{-1}(\sigma_\lambda(Q)))$ increases w.r.t. λ, and by Eq. 10 along with the concavity of R and $\hat{\pi}$, it follows that $(1-\lambda)Q + \lambda\sigma_\lambda(Q)$ increases in λ. Otherwise, $\sigma_{\lambda_1}(Q) > \sigma_{\lambda_2}(Q) > Q$ and, again, we get the same monotonicity of $(1-\lambda)(Q) + \lambda\sigma_\lambda(Q)$ w.r.t. λ. By the previously stated equation $\hat{\pi}'_\lambda(Q) = (1-\lambda)R'((1-\lambda)Q + \lambda\sigma(Q))$, and along with the concavity of R, it follows that $\hat{\pi}'_{\lambda_2}(Q) > \hat{\pi}'_{\lambda_1}(Q)$, as stated.

■