Managerial Turnover in a Changing World*

Daniel Garrett  
Northwestern University  
dfgarrett@gmail.com

Alessandro Pavan  
Northwestern University  
alepavan@northwestern.edu

December 2010

Abstract

This paper develops a dynamic theory of managerial turnover in a world where the quality of the match between a firm and its top managers changes stochastically over time. Shocks to managerial productivity are anticipated at the time of contracting but privately observed by the managers. We characterize the joint dynamics of retention, compensation and effort decisions under the profit-maximizing contract and compare them to their efficient counterparts. Our key positive result shows that the firm’s optimal retention decisions become more permissive with time. What in the eyes of an external observer may thus look like "entrenchment" is, in our theory, the result of a fully-optimal contract in a world where incumbent managers possess privileged information about the firm’s prospects under their own control. Our key normative result shows that, compared to what is efficient, the firm’s optimal contract either induces excessive retention (i.e., inefficiently low turnover) at all tenure levels, or excessive firing at the early stages of the relationship followed by excessive retention after sufficiently long tenure. These results are obtained by endogenizing the firm’s separation payoff accounting for the fact that its performance under each new hire is going to be affected by the same information frictions as in the relationship with the incumbent.

*JEL classification: D82

Keywords: managerial turnover, termination clauses, dynamic mechanism design, adverse selection, moral hazard.

*For useful comments and suggestions, we thank Igal Hendel, Jin Li, Alessandro Lizzeri, Robert Miller, Dale Mortensen, Marco Ottaviani, Bill Rogerson, and seminar participants at Berkeley, Harvard, MIT, Northwestern, Princeton, Stanford, Yale, the European University Institute, the 2010 Gerzensee European Symposium in Economic Theory, the 2010 Econometric Society Winter Meetings, and the 2010 Econometric Society World Congress.
1 Introduction

The job security and pay of a firm’s top manager typically rests on the firm’s consistently good performance and future prospects. This makes sense given the substantial impact that top managers are believed to have on firms’ fortunes. At the same time, the environment in which most firms operate has become increasingly dynamic, implying that managers who are able to deliver high profits in the present may not be able to do so in the future.¹ Shocks to managerial productivity may originate from the opening of new markets, the arrival of new technologies, industry consolidation, or the introduction of new legislation.

The contracts that successful firms offer to their top employees are thus designed not only to incentivize their effort but also to guarantee the desired level of turnover. This is not an easy task given that managers typically have superior information than the board about the determinants of the firm’s profits, the quality of their match with the firm, and the evolution of their own productivity. Optimal contracts must therefore provide managers with incentives not only to exert effort but also to report promptly to the board variations in the environment that affect the firm’s prospects under their own control and for leaving the firm when these prospects deteriorate (equivalently, when the quality of their match with the firm is not satisfactory anymore).

In this paper, we develop a dynamic theory of managerial contracting which, in addition to the familiar theme of incentivizing effort, accounts explicitly for the following possibilities: (i) managerial ability to generate profits is bound to change (stochastically) over time; (ii) shocks to managerial productivity are anticipated at the time of contracting, but privately observed by the managers over time; (iii) at each point in time, the board can respond to poor future prospects by replacing an incumbent manager with a new hire; (iv) the firm’s performance under each new hire is going to be affected by the same information frictions as in the relationship with the incumbent. Accounting for these possibilities not only is realistic, it sheds new light on the joint dynamics (and inefficiency) of effort, retention, and compensation decisions.

Model Preview. In each period, the firm’s cash flows are the result of (i) the incumbent manager’s productivity (equivalently, the quality of the match between the firm and the manager—hereafter the manager’s “type”), (ii) managerial effort, and (iii) noise. Each manager’s productivity is positively correlated over time and each manager has private information about his current and past productivity, as well as about his effort choices. The board only observes the stream of cash flows generated by each manager.

Upon separating from the incumbent, the firm goes back to the labor market and is randomly matched with a new manager. All managers are ex-ante identical. In particular, upon joining the firm, each manager’s productivity evolves according to the same stochastic process. This process is

¹See, for example, Fine (1998), who argues that technology is increasing the speed at which business environments evolve across a plethora of industries.
meant to capture how the interaction of the environment with the tasks that the manager is asked
to perform affects the evolution of the manager’s productivity within the firm (equivalently, the
quality of the match). The environment is perfectly stationary in the sense that the firm faces the
same problem with each manager it hires. As a result, the board offers the same contract to each
manager.\textsuperscript{2,3} A contract is described in terms of: (i) the effort policy it induces—this policy specifies
the effort that the firm recommends to the manager as a function of the evolution of his productivity;
(ii) the compensation that the manager receives over time as a function of the observed cash flows;
and (iii) a turnover policy that specifies under which conditions separation will occur.\textsuperscript{4}

The positive and normative properties of the joint dynamics of effort, turnover, and performance
are identified by characterizing the contract that maximizes the firm’s expected profits (net of man-
gerual compensation) and comparing it to the contract that a benevolent planner would offer to
each manager to maximize welfare (defined to be the sum of the firm’s expected cash flows and of
all managers’ expected payoffs (hereafter, the efficient contract)). Both the profit-maximizing and
the efficient contracts are obtained by comparing, after each history, the value of continuing the
relationship with the incumbent (taking into account the dynamics of future effort and retention
decisions) with the expected value from starting a new relationship with a manager of unknown
productivity. Importantly, both these values are evaluated from an ex-ante perspective, i.e., at the
time each manager is hired. Given the stationarity of the environment, the payoff from hiring a
new manager must coincide with the payoff that the firm expected from hiring the incumbent. Both
the profit-maximizing and the efficient contracts are thus obtained through a fixed-point dynamic-
programming problem that internalizes all relevant trade-offs and whose solution endogenizes the
firm’s separation payoff.

**Key positive results.** Our key positive prediction is that the firm’s optimal retention decisions
become *more permissive over time*: the productivity level that the firm requires for each manager
to be retained declines with the number of periods that the manager has been working for the firm.

\textsuperscript{2}While our analysis focuses on a representative firm, both our positive and normative results apply to a competitive
labor market where managerial outside options are endogenous and coincide with the surplus that each manager expects
from going back to the market and being randomly matched with another identical firm.

\textsuperscript{3}For simplicity, we assume that all managers are risk-neutral. What prevents the principal from “selling the firm
to the managers” is the fact that the latter have persistent private information about their ability to generate profits
and hence about their value for the firm. Another reason why the principal may not be able to sell the firm to the
managers is that the latter may be cash-constrained. While we do not model explicitly such a possibility, our results
do not hinge on the absence of limited-liability constraints, as explained in Section 4.

\textsuperscript{4}In general, a turnover policy based solely on observed cash flows cannot induce the optimal sequence of separation
decisions. It may be essential that managers keep communicating with the board, e.g., by explaining the determinants
of past performances and/or by describing the firm’s prospects under their control. A key role of the optimal contract
in our theory is precisely to induce a prompt exchange of information between the managers and the board, in addition
to the more familiar role of incentivizing effort through performance-based compensation.
This result originates from the combination of the following two assumptions: (i) the effect of a manager’s initial productivity on his future productivities declines over time;\textsuperscript{5} and (ii) variations in managerial productivity are anticipated, but privately observed.

The explanation rests on the board’s desire to limit the compensation to the managers who are most productive at the initial contracting stage that is necessary to separate them from the less productive ones. Similar to Laffont and Tirole (1986), such a “rent” originates from the possibility for the most productive managers of generating the same distribution of present and future cash flows as the less productive ones by working less, thus economizing on the disutility of effort.\textsuperscript{6} Contrary to Laffont and Tirole’s static analysis, in our dynamic environment firms have two instruments to limit such rents: first, they can induce less productive managers to work less (e.g., by offering them contracts with low-powered incentives where compensation is relatively insensitive to realized cash flows); in addition, they can commit to a replacement policy that is more severe to a manager whose initial productivity is low in terms of the future productivity and performance levels required for retention. Both instruments play the role of discouraging those managers who are most productive at the contracting stage from mimicking the less productive ones and are thus most effective when targeted at those managers whose initial productivity is low.

The key observation is that, when the effect of a manager’s initial productivity on his subsequent productivity is expected to decline over time, the effectiveness of such instruments is higher when they are used at the early stages of the relationship than in the distant future. The reason is that, from the perspective of a manager who is initially most productive, his ability to “do better” than a manager who is initially less productive is prominent at the early stages, but expected to decline over time due to the imperfect serial dependence of the productivity process.

The firm’s profit-maximizing retention policy is then obtained by trading off two considerations. On the one hand, the desire to respond promptly and efficiently to variations in the environment that affect the firm’s prospects under the incumbent’s control, of course taking into account the dynamics of future effort and retention decisions. This concern calls for retaining managers whose productivity is expected to remain or turn high irrespective of whether or not their initial productivity was low. On the other hand, the value of offering a contract that reduces the compensation that the firm must leave to the managers who are most productive at the hiring stage. This second concern calls for committing to a retention policy that is most severe to those managers whose initial productivity is low. However, because the value of such commitments declines with the length of the employment relationship, the profit-maximizing retention policy becomes gradually more lenient over time.

Our theory thus offers a possible explanation for what in the eyes of an external observer may look like "\textit{entrenchment}". That managers with a longer tenure are retained under the same conditions

\textsuperscript{5}Below, we will provide a formal statement of this assumption in terms of a statistical property of the process governing the evolution of the managers’ productivity.

\textsuperscript{6}Equivalently, by the possibility of generating higher cash flows for the same amount of effort.
that would have called for separation at a shorter tenure is, in our theory, the result of a fully optimal contract, as opposed to the result of a lack of commitment or of good governance. In this respect, our explanation is fundamentally different from the alternative view that managers with longer tenure are "entrenched" because they are able to exert more influence over the board, either because of manager-specific investments, as in Shleifer and Vishny (1989), or because of the appointment of less independent directors, as in Hermalin and Weisbach (1998) – see also, Weisbach (1988), Denis, Denis, and Sarin (1997), Hadlock and Lumer (1997), Rose and Shepard (1997), Almazan and Suarez (2003), Bebchuk and Fried (2004), and Fisman, Kuhrana, and Rhodes-Kropf (2005).

We also show that, under the profit-maximizing contract, effort can be optimally incentivized through linear schemes. These schemes combine a fixed (but time-varying) payment with a bonus which is linear in cash flows. The slope of the linear scheme and hence the induced effort level, increases, on average, with the manager’s tenure in the firm.\(^7\) As with retention, this property originates from the assumption that the effect of each manager’s initial productivity on his subsequent productivity is expected to decline over time. This property implies that the benefit of distorting downward the effort in the contracts of those managers who are less productive at the initial stages, so as to discourage the most productive types from mimicking them, declines over time. As a result, on average, the slope of the linear scheme, and hence the sustained effort level, increases with tenure. This property is consistent with the practice of putting more stocks and options in the package of managers with a longer tenure in the firm (see, e.g., Lippert and Porter, 1997, but also Gibbons and Murphy, 1991).\(^8\)

**Key normative results.** Turning to the normative results, we find that, compared to what is efficient, the firm’s profit-maximizing contract either induces excessive retention at all tenure levels, or excessive firing at the early stages of the relationship, followed by excessive retention in the long run. By excessive retention we mean the following. Any manager who is fired after \(t\) periods of employment under the profit-maximizing contract is either fired in the same period or earlier under the efficient policy. By excessive firing we mean the exact opposite: any manager fired at the end of period \(t\) under the efficient policy is either fired at the end of the same period or earlier under the profit-maximizing contract.

\(^7\)That effort increases with tenure is not what drives the result that retention decisions become more lenient over time. The same result obtains in an environment where effort is either exogenously, or endogenously, fixed over time, as for example in DeMarzo and Fishman (2007) where optimal and efficient effort are constant and coincide with "no stealing." However, note that a model in which effort is \textit{exogenously} fixed may deliver the potentially misleading prediction of a negative correlation between tenure and performance. One should also be careful in not interpreting the property that effort increases with tenure as meaning that "older" managers supply more labor than "younger" ones. In fact, effort in our theory simply stands for the impact of a manager’s activity on the distribution of the firm’s cash flows.

\(^8\)While in Gibbons and Murphy (1991) the optimality of seniority-based schemes originates from learning in a career concerns model, in our theory it originates from time-varying persistent private information.
The result that retention decisions become less efficient over time may appear in contrast to findings in the dynamic mechanism design literature that "distortions" in optimal contracts typically decrease over time and vanish in the long-run. (This property has been documented by various authors, going back at least to Besanko’s (1985) seminal work; see Pavan, Segal, and Toikka, 2009, for a recent unifying explanation based on the statistical property of declining impulse responses, and Garrett and Pavan, 2010, for a discussion in the context of managerial compensation.)

The reason why we do not find convergence to efficiency in the long run in the current setting is that the principal’s outside option is endogenous and affected by the same information and incentive frictions as in the relationship with the incumbent. To understand this result, recall that, from an ex-ante viewpoint (i.e., at the time of hiring), the firm expects to extract more and more surplus from each incumbent as time goes on. As explained above, the reason is that the effect of each manager’s initial private information on his future productivity is expected to decline over time. The fact that the firm can appropriate more and more surplus as time passes in turn implies that the flow (i.e., per period) payoff that the firm expects from retaining the incumbent increases over time and eventually converges to the flow payoff that a benevolent planner would expect from retaining the same manager. On the other hand, the payoff that the firm expects from going back to the labor market and starting a relationship with a new manager of unknown productivity is necessarily lower than the surplus that the planner would expect from starting a new relationship. The reason is precisely that the new relationship is expected to be affected by the same informational and incentive problems as the one governing the interaction with the incumbent. This implies that, to reduce the compensation paid to the new hire, the firm expects to introduce distortions in future effort and retention decisions which reduce the surplus of the new relationship relative to the one created under the efficient contract. This in turn implies that, from an ex-ante viewpoint, after tenure grows long enough, replacement becomes less attractive for a firm maximizing profits than for a planner maximizing efficiency. As a result, firms eventually become excessively lenient when it comes to their retention decisions.\(^9\)

This last result suggests that policy interventions aimed at inducing firms to sustain a higher turnover, e.g., by offering them temporary tax incentives after a change in management, or through the introduction of a mandatory retirement age for top employees, can, in principle, increase welfare.\(^10\) Of course, such policies might be expected to encounter opposition on other grounds whose discussion is beyond the scope of this analysis.

**Layout.** The rest of the paper is organized as follows. In the remainder of this section we briefly review the pertinent literature. Section 2 introduces the model. Section 3 characterizes the efficient contract. Section 4 characterizes the firm’s profit-maximizing contract and uses it to establish the key positive results. Section 5 compares the dynamics of retention decisions under the efficient contract

---

\(^9\)This result also applies to a setting in which effort is required to be constant over time.

\(^10\)See Lazear (1979) for alternative explanations for why mandatory retirement can be beneficial.
with those under the profit-maximizing contract and establishes the key normative results. All proofs are in the Appendix.

1.1 Related literature

The paper is related to various lines of research in the managerial compensation and turnover literature. A vast body of work documents how the threat of replacement plays an important role in incentivizing effort.\textsuperscript{11,12} Recent contributions in this area include Tchistyi (2005), Clementi and Hopenhayn (2006), DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007), Biais et al. (2007), and He (2009).\textsuperscript{13} Related to this line of research is also the work by Spear and Wang (2005), Wang (2008) and Sannikov (2008) where it is shown how a risk-averse agent may be optimally induced to cease to exert effort and then retire, once his promised continuation utility becomes either too high or too low, making it too costly for the firm to incentivize further effort.

While not all of these works focus explicitly on turnover, they do offer implications for the dynamics of retention decisions. For example, Wang (2008) shows how a worker with a shorter tenure faces a higher probability of an involuntary layoff and a lower probability of voluntary retirement than a worker with a longer tenure. In a financial contracting setting, Clementi and Hopenhayn (2006) show how, on average, a borrower’s promised continuation utility increases over time and how this requires an increase in the likelihood that the loan is rolled over. Similarly, Fong and Li (2010) find that the turnover rate eventually decreases in the duration of the employment relationship, but, because contracts are relational, they also find that the turnover rate may initially increase.

The above literature does not account for the possibility of changes in the quality of the match between the firm and its managers. It therefore misses the possibility that turnover is driven by variations in match quality as opposed to concerns for incentivizing effort. Such a possibility has long been recognized as important by another body of the literature that dates back at least to Jovanovic\textsuperscript{11} Despite the vast attention that this property has received in the theoretical literature, the empirical evidence of the effect of turnover on incentives is mixed. See Jenter and Lewellen (2010) for a recent discussion and Gayle, Golan, and Miller (2008) for a recent empirical study of the relationship between promotion, turnover, and compensation in the market for executives.

\textsuperscript{12} Another paper where dismissal helps creating incentives is Sen (1996). In this paper, the manager’s private information is the productivity of the firm, which is assumed to be constant over time and independent of the manager who runs it. As in the current paper, commitments to replace the initial manager help reducing informational rents. However, contrary to the current paper, the manager does not have hidden actions and there is a single replacement decision. The analysis in Sen (1996) thus does not permit one to study how the leniency of retention decisions evolves over time.

\textsuperscript{13} The reason why the threat of termination is essential in these papers is that the agent is protected by limited liability. This implies that incentives provided entirely through performance-based compensation need not be strong enough. The threat of termination is also crucial in the “efficiency wages” theory; in particular, see Shapiro and Stiglitz’s (1984) seminal work. However, contrary to the literature cited above, in this theory, under the optimal contract, no worker shirks and hence replacement does not occur in equilibrium.
This paper considers an environment where productivity (equivalently, the match quality) is constant over time but unknown to both the firm and the worker who jointly learn it over time through the observation of realized output. Because of learning, turnover becomes less likely over time. Our theory differs from Jovanovic (1979) in a few respects. First, and importantly, we allow learning about match quality to be asymmetric between the workers and the firm, with the former possessing superior information than the latter. Second, we explicitly model managerial effort and account for the fact that it must be incentivized. Third, we consider more general processes for the evolution of the match quality. These distinctions lead to important differences in the results. First, while in Jovanovic's model the leniency of turnover decisions originates from the accumulation of information over time, in our model turnover decisions become more lenient over time even when conditioning on the accuracy of available information (formally, even when the kernels, i.e., the transition probabilities, remain constant over time). Second, while in Jovanovic's model turnover decisions are always second-best efficient, in our model, turnover decisions are second-best inefficient and the inefficiency of such decisions typically increases over time.

More recent papers where turnover is also driven by variations in match quality include Acharya (1992), Mortensen and Pissarides (1994), Atkeson and Cole (2005), and McAdams (2010). Acharya (1992) studies how the market value of a firm changes after the announcement to replace a CEO and how the probability of replacement is affected by the CEO’s degree of risk aversion. Mortensen and Pissarides (1994) show how the optimal turnover policy takes the form of a simple threshold policy, with the threshold being constant over time. Along with the assumption that productivity is drawn independently each time it changes and the fact that the revisions follow a Poisson process, this implies that the probability of terminating a relationship does not vary with tenure. In contrast, in a model of stochastic partnerships, McAdams (2010) finds that relationships become more stable over time due to a survivorship bias. Atkeson and Cole (2005) show how managers who delivered high performance in the past have a higher continuation utility and are then optimally rewarded with some form of job stability. Because a longer tenure implies a higher probability of having delivered a high performance in the past, their model also offers a possible explanation for why retention decisions may become more lenient over time.

---

14 Allgood and Farrell (2003) provide empirical support to the importance of variations in match quality for turnover decisions.
15 Related is also Holmstrom’s (1999) career concerns model. While this paper does not characterize the optimal turnover policy, the evolution of career concerns have been recognized as a possible determinant for turnover; see, for example, Mukherjee (2008).
16 Inefficiencies originate in our theory from the combination of asymmetric information with search frictions. Because neither the firms nor the managers can appropriate the entire surplus, contractual decisions are distorted relative to their second-best counterparts.
17 Acharya (1992) also documents the possible optimality of permanently tenuring a CEO, a possibility that we also accommodate but which we show to never be optimal in our model.
An important distinction between our paper and the two bodies of the literature discussed above is that, in our theory, variations in match quality are anticipated but privately observed. As a result, a properly designed contract must not only incentivize effort but also provide managers with incentives for truthfully reporting to the board variations in match quality that call for adjustments in the compensation scheme and possibly for separation decisions. The importance of private information for turnover decisions has been recognized by another body of the literature that includes Levitt and Snyder (1997), Banks and Sundaram (1998), Eisfeldt and Rampini (2008), Yang (2009), Gayle, Golan and Miller (2008), and Inderst and Mueller (2010). Some of these papers show how asymmetric information may lead to a form of entrenchment, i.e., to situations in which the agent remains in place (or the project continues) although the principal would prefer ex-post to replace him (or discontinue the project). What is missing in this literature is an account of the possibility that the managers’ private information may change over time and hence an analysis of how the leniency of optimal turnover decisions evolves with the managers’ tenure in the firm.\(^{18}\)

Another important difference between our work and each of the various papers mentioned above is that it offers an analysis of how the inefficiency of turnover decisions evolves over time. To the best of our knowledge, this analysis has no precedents in the literature. As explained above, this is made possible by endogenizing the firm’s separation payoff and recognizing that the relationship with each new hire is going to be affected by the same frictions as the one with each incumbent. Recognizing this possibility is essential to our normative result about the excessive leniency of retention decisions after a long tenure.

From a technical viewpoint, the paper builds on recent developments in the theory of dynamic mechanism design with persistent shocks to the agents’ private information and in particular on Pavan, Segal and Toikka (2009).\(^{19}\) We refer the reader to that paper for a discussion of these developments and for a survey of the dynamic mechanism design literature. The results in the current paper extend the techniques of Pavan, Segal, and Toikka (2009) to a setting with both adverse selection and moral hazard and where the principal’s outside option is endogenous because of the possibility of replacement.\(^{20}\)

\(^{18}\)An exception is Gayle, Golan and Miller (2008). They use a longitudinal data set to evaluate the importance of moral hazard and job experience in jointly determining promotion, turnover rates, and compensation, and to study how the latter changes across the different layers of an organization. The focus of their analysis is, however, very different from ours.

\(^{19}\)The analysis in the current paper, as well as in Pavan, Segal, and Toikka (2009), is in discrete time. Recent contributions in continuous time include Williams (2009), Zhang (2009), and Strulovici (2010). These works show how the solution to a class of dynamic adverse selection problems with persistent private information (but without replacement) can be obtained in a recursive way with the level and derivative of promised utility as state variables. In contrast, both the optimal and the efficient contracts in our paper are obtained through a fixed-point dynamic programming problem whose solution is not recursive, thus permitting us to show how effort, compensation, and retention decisions depend explicitly on the entire history of productivity shocks.

\(^{20}\)See also Edmans and Gabaix (2009b), and Edmans, Gabaix, Sadzik, and Sannikov (2009) for recent contributions
Finally, related is obviously also the literature on dynamic managerial compensation without replacement. This literature is too vast to be successfully summarized here. We refer the reader to Garrett and Pavan (2010) and to Edmans and Gabaix (2009a) for an overview. In Garrett and Pavan (2010), we abstract from the possibility of replacement and instead investigate how the optimality of seniority-based schemes (that is, schemes that provide managers with longer tenure with more high-powered incentives) is affected by the managers’ degree of risk aversion. For simplicity, the current paper does not account for the possibility that managers are risk averse; however, we expect all our results to remain true for a suitably low degree of risk aversion.

2 Model

Players. A principal (the board of directors, acting on behalf of the shareholders of the firm) is in charge of designing a new employment contract to govern the firm’s interaction with its managers.\textsuperscript{21} The firm is expected to operate for infinitely many periods and each manager is expected to live as long as the firm. There are infinitely many managers. All managers are ex-ante identical and have a productivity (equivalently, an intrinsic ability to generate cash flows for the firm) that is expected to evolve over time according to the same Markov process as described below.

Stochastic process. Let $t \in \mathbb{N}$ denote the number of periods that a given manager has been working for the firm. Irrespective of the date of hiring, the manager’s period-$t$ productivity $\theta_t$ (i.e., his productivity during the $t^{th}$ period of employment) is drawn from a cumulative distribution function $F_t(\cdot|\theta_{t-1})$ defined on the interval $\Theta_t = [\theta_t, \bar{\theta}_t]$ with $\theta_t, \bar{\theta}_t \in \mathbb{R}$, $\bar{\theta}_t > \theta_t$, and with $\theta_0$ known. The set $\Theta_t$ is the support of the marginal distribution of the random period-$t$ productivity. Given $\theta_{t-1}$, the support of the conditional distribution $F_t(\cdot|\theta_{t-1})$ can, however, be a strict subset of $\Theta_t$.\textsuperscript{22}

where, as in Laffont and Tirole (1986) and in the current paper, the moral hazard problem is solved using techniques from the mechanism design literature. These works consider a setting where (i) there is no turnover, (ii) managers possess no private information at the time of contracting, and (iii) it is optimal to induce a constant level of effort over time. Relaxing (i) and (ii) is essential for our results. On the other hand, allowing effort to change over time is realistic and permits us to investigate the joint dynamics of effort and retention decisions, but is not essential to either our positive or normative results.

\textsuperscript{21} As anticipated above, the focus of the analysis here is on the contracts offered by a representative firm for given contracts offered by all other competing firms (equivalently, for given managers’ outside options). However, the profit-maximizing and efficient contracts characterized below are also equilibrium and welfare-maximizing contracts in a setting where unemployed managers are randomly matched with many (ex-ante identical) firms. Indeed, as it will become clear from the results in Propositions 1 and 3 below, as long as the number of potential managers is large compared to the number of competing firms, so that the matching probabilities remain independent of the contracts selected, then the managers’ outside option (i.e., their payoff after separation occurs) has an effect on the level of compensation but not on the profit-maximizing and efficient effort and retention policies.

\textsuperscript{22} Allowing for more than two periods is essential to being able to examine the dynamics of retention decisions. Allowing for more than two productivity levels is also essential. One can easily verify that with two productivity levels, the optimal retention policy takes one of the following three forms: (i) either the manager is never replaced, irrespective
Hereafter, we identify the process governing the evolution of the managers’ productivity with the collection of kernels $F \equiv \langle F_t(\cdot) \rangle_{t=1}^\infty$. Let $\theta^t \equiv (\theta_1, ..., \theta_t) \in \Theta^t \equiv \times_{s=1}^t \Theta_s$.\footnote{Throughout the entire manuscript, we will use superscripts to denote sequences of variables.} For each $t$, then let $R_t^t \equiv \{ \theta^t \in \Theta^t : \theta_1 \in \Theta_1$ and $\theta_t \in \text{Supp}[F_t(\cdot|\theta_{t-1})], \text{ all } l = 2, \ldots, t \}$ denote the set of possible histories of productivities that are compatible with the process $F$. For any $t$, any $\theta_{t-1}$, $F_t(\cdot|\theta_{t-1})$ is absolutely continuous over $\mathbb{R}$ with density $f_t(\cdot|\theta_{t-1}) > 0$ over a connected subset of $\Theta_t$. Moreover, for any $t$, any $\theta_t \in \mathbb{R}$, almost any $\theta_{t-1} \in \Theta_t$, $\partial F_t(\theta_t|\theta_{t-1})/\partial \theta_{t-1}$ exists.\footnote{Note that $\partial F_t(\theta_t|\theta_{t-1})/\partial \theta_{t-1}$ (respectively, $\partial F_t(\theta_t|\theta_{t-1})/\partial \theta_{t-1}$) denotes the right-hand (respectively, left-hand) derivative of $F_t$ with respect to $\theta_{t-1}$.}

These impulse response functions are the nonlinear analogs of the familiar constant linear impulse response functions for autoregressive processes. If productivity evolves according to an AR(1) process $\gamma \theta_{t-1} + \varepsilon_t$, the impulse response of $\theta_t$ on $\theta_{t-1}$, $\gamma > 0$, is simply given by $J_t^\gamma = \gamma^{t-1}$. More generally, the impulse response $J_t^\gamma(\theta_{t-1})$ of $\theta_t$ on $\theta_{t-1}$ captures the total effect of an infinitesimal variation of $\theta_t$ on $\theta_{t-1}$, taking into account all effects on intermediate types $(\theta_{t+1}, ..., \theta_{t-1})$:

$$\frac{\partial \mathbb{E}[\theta_t|\theta_{t-1}]}{\partial \theta_{t-1}} = \mathbb{E}[J^\gamma_{t-1}(\theta_{t-1})].$$

As shown below, these functions play a key role in determining the dynamics of profit-maximizing effort and turnover policies.

We assume throughout that the process $F$ satisfies the property of “first-order stochastic dominance in types”:\footnote{of the evolution of his productivity; or (ii) he is retained if and only if his initial productivity was high; or (iii) he is fired as soon as his productivity turns low. In all cases, the retention policy (i.e., whether the manager is retained as a function of his period-t productivity) is independent of the length of the employment relationship.} for all $t \geq 2$, $\theta_{t-1} > \theta_{t-1}'$ implies $F_t(\theta_t|\theta_{t-1}) \leq F_t(\theta_t|\theta_{t-1}')$ for all $\theta_t$. Note that the assumption of first-order stochastic dominance in types implies that, for all $t$ and $\tau > t$, all $\theta^\tau \in R^\tau$, $J_t^\tau(\theta^\tau_{t-1}) \geq 0$.

We will say that the process is “autonomous” if, for all $t, s \geq 2$ and for any $\theta \in \Theta_{t-1} \cap \Theta_{s-1}$, $F_t(\cdot|\theta) = F_s(\cdot|\theta)$. Throughout, we will maintain the assumption that types evolve independently across managers.

For each $t$, all $\theta^t \in R^t$, let $J_t^\gamma(\theta^t_{t-1}) \equiv \Pi_{k=t+1}^{t} I_{k-1}^\gamma (\theta_k, \theta_{k-1})$,

with each $I_{k-1}^\gamma (\theta_k, \theta_{k-1})$ defined by

$$I_{k-1}^\gamma (\theta_k, \theta_{k-1}) \equiv \frac{-\partial F_k(\theta_k|\theta_{k-1})/\partial \theta_{k-1}}{f_k(\theta_k|\theta_{k-1})}.$$
**Effort, cash flows, and payoffs.** After learning his period-1 productivity $\theta_t$, the manager currently employed by the firm must choose an effort level $e_t \in E \equiv \mathbb{R}$. The firm’s per-period cash flows, gross of the manager’s compensation, are given by

$$\pi_t = \theta_t + e_t + \nu_t,$$

(1)

where $\nu_t$ is a transitory noise shock. The shocks $\nu_t$ are i.i.d. over time, independent across managers, and drawn from the distribution $\Phi$, with expectation $E[\nu_t] = 0$. The sequences of productivities $\theta^t$ and efforts $e^t \equiv (e_1, ..., e_t) \in E^t$ are each manager’s private information. In contrast, the history of cash flows $\pi^t \equiv (\pi_1, ..., \pi_t) \in \Pi^t \equiv \mathbb{R}^t$ generated by each manager is verifiable and can be used as a basis for compensation.

By choosing effort $e_t$ in period $t$, the manager suffers a disutility $\psi(e_t)$. Denoting by $c_t$ the compensation that the manager receives in period $t$ (equivalently, his period-$t$ consumption), the manager’s preferences over (lotteries over) streams of consumption levels $c \equiv (c_1, c_2, ...) $ and streams of effort choices $e \equiv (e_1, e_2, ...) $ are described by an expected utility function with (Bernoulli) utility given by

$$U^A(c, e) = \sum_{t=1}^{\infty} \delta^{t-1} \left[ c_t - \psi(e_t) \right],$$

(2)

where $\delta < 1$ is the (common) discount factor.

The principal’s objective is to maximize the discounted sum of the firm’s expected profits, defined to be cash flows net of the managers’ compensation. Formally, let $\pi_{it}$ and $c_{it}$ denote, respectively, the cash flow generated and the compensation received by the $i^{th}$ manager employed by the firm in his $t^{th}$ period of employment. Then, let $T_i$ denote the number of periods for which manager $i$ works for the firm. The contribution of manager $i$ to the firm’s payoff, evaluated at the time manager $i$ is hired, is given by

$$X_i(\pi_{i}^{T_i}, C_{i}^T) = \sum_{t=1}^{T_i} \delta^{t-1} [\pi_{it} - c_{it}],$$

Next, denote by $I \in \mathbb{N} \cup \{+\infty\}$ the total number of managers hired by the firm over its infinite life. The firm’s payoff, given the cash flows and payments $(\pi_{i}^{T_i}, c_{i}^{T_i})_{i=1}^{I}$, is then given by

$$U^P = \sum_{i=1}^{I} \delta^{\sum_{j=1}^{i-1} T_j} X_i(\pi_{i}^{T_i}, c_{i}^{T_i}).$$

(3)

Given the stationarity of the environment, with an abuse of notation, throughout the entire analysis, we will omit all indices $i$ referring to the identities of the managers.

**Timing and labor market.** The firm’s interaction with the labor market unfolds as follows. Immediately after being matched with the firm, the manager privately learns his period-1 productivity $\theta_1$ (drawn from the distribution $F_1$) and then decides whether or not to sign the employment contract described in full detail below. If the manager refuses to sign, he then leaves the firm and
receives an outside option equal to $U^o \geq 0$, irrespective of $\theta_1$. In this case, the principal must wait until the next period before going back to the labor market. This assumption is meant to capture (admittedly in a crude way) the idea that it takes time (in the model, one period) to find a new manager so that replacement decisions must be planned in advance. Without such a friction, the board would continue sampling until it finds a manager of the highest possible productivity, which is unrealistic and makes the analysis uninteresting. We also assume that it is never optimal for the principal to let the firm operate without a manager.

After signing the contract, the manager sends a message $m_1$ to the firm (think of $m_1$ as the choice of an element of the contract, such as a clause pertaining to either the compensation or the retention policy). After sending the message $m_1$, the manager privately chooses effort $e_1$. Nature then draws $\nu_1$ from the distribution $\Phi$ and finally the firm’s (gross) cash flows $\pi_1$ are determined according to (1). After observing the cash flows $\pi_1$, the firm pays the manager a compensation $c_1$ (which may depend on both the clause $m_1$ and the cash flows $\pi_1$) and the contract then stipulates whether or not the manager is to be retained into the second period (in general, we allow retention to depend on $m_1$ and $\pi_1$).

The manager’s second-period productivity is drawn from the distribution $F_2(\cdot|\theta_1)$. After privately learning $\theta_2$, the manager decides whether or not to leave the firm. If he leaves, the manager obtains a continuation payoff equal to $U^o \geq 0$. If he stays, he then sends a new message $m_2$ (again, think of this message as the choice of some new contractual term); he then privately chooses effort $e_2$; cash flows $\pi_2$ are realized; the manager is paid compensation $c_2$ (which may depend on aspects of the contract determined in the previous and current period, i.e., on $m^2 = (m_1, m_2)$, as well as the entire history of observed cash flows $\pi^2 = (\pi_1, \pi_2)$); finally, given $m^2$ and $\pi^2$, the contract again stipulates whether the manager will be retained into the next period.

The entire sequence of events described above repeats itself over time until the firm decides to

---

25 While our model is meant to apply to markets where productivity is match-specific—in which case the outside option $U^o$ is a constant—all our results extend qualitatively to a setting where productivity is manager-specific, and hence correlated across jobs. This is true as long as the derivative of the outside option $U^o (\theta_1)$ with respect to current productivity is sufficiently small that the single-crossing conditions of Section 4 are preserved. This is the case, for example, when (i) the discount factor is not very high, and/or (ii) it takes a long time for a manager to find a new job.

26 From the perspective of the firm under examination here, this outside option is exogenous. However, in equilibrium, $U^o$ will coincide with the equilibrium continuation payoff that each manager expects from going back to the labor market and being randomly matched (possibly after an unemployment phase) with another firm offering the same contract characterized below.

27 That retention decisions are specified explicitly in the contract simplifies the exposition but is not essential. For example, by committing to pay a sufficiently low compensation after all histories that are supposed to lead to separation, the firm can always implement the desired retention policy by delegating to the managers the choice of whether or not to stay in the relationship. It will become clear from the analysis below that, while both the optimal and the efficient retention policies are unique, there are many ways these policies can be implemented (see, e.g., Yermack, 2006 for a description of the most popular termination clauses and "golden handshakes" practices).
separate from its incumbent manager or the latter decides to leave the firm. After separation has been decided, at the beginning of the subsequent period, the firm goes back to the labor market and is randomly matched with a new manager. The new manager’s productivity $\theta_1$ is then drawn from the same (time-invariant) distribution $F_1$. The relationship between the new manager and the firm then unfolds as described above until separation occurs (as mentioned above, this will be determined at some period $t$, with effect from period $t + 1$, where $t$ will typically be random). In the period immediately following the one in which separation has been decided, the firm goes back to the labor market and is randomly matched with a new manager. The same sequence of events described above then applies to the new relationship as well as to any future one.\footnote{The assumption of random matching is quite standard in the labor/matching literature (see, e.g., Jovanovic, 1979). In our setting, it implies that there is no direct competition among managers for employment contracts. This distinguishes the environment from an auction-like setting where, in each period, the principal consults simultaneously with multiple managers and then chooses which one to hire/retain.}

**Technical assumptions.** To validate a certain dynamic envelope theorem (see Pavan, Segal, and Toikka 2009 for details), guarantee interior solutions, and be able to apply the Contraction Mapping Theorem, we will make the following technical assumptions. We will assume that the sets $\Theta_t$ are uniformly bounded, i.e., that there exists $K < +\infty$ such that $|\theta| < K$ for all $\theta \in \Theta_t$. We will also assume that the functions $J_t^\tau(\cdot)$ are uniformly bounded, in the sense that there exists $J < +\infty$ such that $J_t^\tau(\theta^\tau_{t-1}) < J$ for all $t$ and $\tau > t$, all $\theta^\tau \in R^\tau$. Lastly, we will impose the following conditions on the disutility function $\psi$. Firstly, $\psi(e) = 0$ for all $e \leq 0$. Secondly, $\psi$ is continuously differentiable over $\mathbb{R}$. Thirdly, there exists a scalar $\bar{e} > 0$ such that (i) $\psi'(\bar{e}) > 1$, and (ii) $\psi$ is thrice continuously differentiable over $(0, \bar{e})$ with $\psi''(e), \psi'''(e) > 0$ and $\psi''''(e) \geq 0$ for all $e \in (0, \bar{e})$. Finally, to allow a direct application of the dynamic envelope theorem, we will assume that there exist scalars $C > 0$ and $L > 1$ such that $\psi(e) = Le - C$ for all $e > \bar{e}$. These conditions are satisfied, for example, when $\bar{e} > 1$, $\psi(e) = (1/2)e^2$ for all $e \in (0, \bar{e})$, and $\psi(e) = \bar{e}e - \bar{e}^2/2$ for all $e > \bar{e}$.

### 2.1 The employment relationship as a dynamic mechanism

Because all managers are ex-ante identical, time is infinite, and types evolve independently across managers, the firm offers the same contract to each manager. Under such a contract, the compensation that the firm pays to each manager is a function of the messages that the manager sends over time and of the cash flows that he generates, but is independent of both the calendar time at which the manager is hired and of the history of messages and cash flows generated by other managers. Hereafter, we will thus maintain the notation that $t$ denotes the number of periods that a representative manager has been working for the firm and not the calendar time.

Furthermore, because the firm can commit, one can conveniently describe the firm’s contract
as a direct revelation mechanism which specifies, for each period \(t\), a recommended effort choice, a compensation, and a retention decision. In principle, the recommended effort choice and the retention decision may depend both on the history of reported productivities \(\theta^t\) and on the history of past cash flow realizations \(\pi^{t-1}\). However, because the latter are a noisy transformation of effort and productivity, one can easily verify that, under both the efficient and the profit-maximizing contracts, the optimal effort and retention decisions will depend only on reported productivities \(\theta^t\). On the other hand, because the effort decisions are “hidden actions” (i.e., because of moral hazard) it is essential that the compensation be allowed to depend both on the reported productivities \(\theta^t\) as well as on past and current cash flows \(\pi^t\).

Hereafter, we will thus model the employment relationship induced by the profit-maximizing and the efficient contracts as a direct revelation mechanism \(\Omega \equiv (\xi, s, \kappa)\), where the latter consists of sequences of functions \(\xi \equiv (\xi_t : \Theta^t \rightarrow E)_{t=1}^{\infty}\), \(s \equiv (s_t : \Theta^t \times \Pi^t \rightarrow \mathbb{R})_{t=1}^{\infty}\) and \(\kappa \equiv (\kappa_t : \Theta^t \rightarrow \{0, 1\})_{t=1}^{\infty}\) such that:

- \(\xi_t(\theta^t)\) is the recommended period-\(t\) effort;
- \(s_t(\theta^t, \pi^t)\) is the compensation paid at the end of period \(t\);
- \(\kappa_t(\theta^t)\) is the retention decision for period \(t\), with \(\kappa_t(\theta^t) = 1\) if the manager is to be retained, which means he is granted the possibility of working for the firm also in period \(t + 1\), regardless of his period-(\(t + 1\)) productivity \(\theta_{t+1}\),\(^{29}\) and \(\kappa_t(\theta^t) = 0\) if (i) either he is dismissed at the end of period \(t\), or (ii) he was dismissed in previous periods; i.e., \(\kappa_t(\theta^t) = 0\) implies \(\kappa_s(\theta^s) = 0\) for all \(s > t\), all \(\theta^s\).\(^{30}\) Given any sequence \(\theta^\infty\), we then denote by \(\tau(\theta^\infty) \equiv \min \{t : \kappa_t(\theta^t) = 0\}\) the corresponding length of the employment relationship.

In each period \(t\), given the previous reports \(\hat{\theta}^{t-1}\) and cash flows realizations \(\pi^{t-1}\), the employment relationship unfolds as follows:

- After learning his period-\(t\) productivity \(\theta_t \in \Theta_t\), and upon deciding to stay in the relationship, the manager sends a report \(\hat{\theta}_t \in \Theta_t\);
- The mechanism then prescribes effort \(\xi_t(\hat{\theta}^{t-1}, \hat{\theta}_t)\) and specifies a reward scheme \(s_t(\hat{\theta}^{t-1}, \hat{\theta}_t, \pi^{t-1}, \cdot) : \Pi_t \rightarrow \mathbb{R}\) along with a retention decision \(\kappa_t(\hat{\theta}^{t-1}, \hat{\theta}_t)\);
- The manager then chooses effort \(e_t\);
- After observing the realized cash flows \(\pi_t = e_t + \theta_t + \nu_t\), the manager is paid \(s_t(\hat{\theta}^{t-1}, \hat{\theta}_t, \pi^{t-1}, \pi_t)\) and is then either retained or replaced according to the decision \(\kappa_t(\hat{\theta}^{t-1}, \hat{\theta}_t)\).

\(^{29}\)Recall that separation decisions must be planned one period in advance.

\(^{30}\)For expositional convenience, we allow the policies \(\xi_t\), \(s_t\), and \(\kappa_t\) to be defined over all possible histories, including those histories that lead to separation at some \(s < t\). This, of course, is inconsequential for the analysis.
By the revelation principle, we restrict attention to direct mechanisms for which (i) a truthful and obedient strategy is optimal for the manager, and (ii) after any truthful and obedient history, the manager finds it optimal to stay in the relationship whenever offered the possibility of doing so (i.e., the manager never finds it optimal to leave the firm when he has the option to stay).

3 The efficient contract

We begin by describing the effort and turnover policies, $\xi^E$ and $\kappa^E$, that maximize ex-ante welfare, defined to be the sum of a representative manager’s expected payoff and of the firm’s expected profits (the “efficient” policies).\(^{31}\) Although we are clearly interested in characterizing these policies for the same environment as described above, it turns out that these policies coincide with the ones that maximize ex-ante welfare in an environment with symmetric information, in which the managers’ productivities and effort choices are observable and verifiable. In turn, because all players’ payoffs are linear in payments, these policies also coincide with the equilibrium ones that each firm would choose under symmetric information to maximize expected profits. For simplicity, in this section, we thus assume information is symmetric and then show in Section 5 – Corollary 2 – that the efficient policies under symmetric information remain implementable also under asymmetric information.

The efficient effort policy is very simple: Because all players are risk neutral and because each manager’s productivity has no effect on the marginal cost or the marginal benefit of effort, the efficient effort level is independent of the history of realized productivities.

The efficient turnover policy, on the other hand, is the solution to a dynamic programming problem. Because the firm does not know the future productivity of its current manager, nor the productivities of its future hires, this problem involves a trade-off in each period between experimenting with a new manager and continuing experimenting with the incumbent. Define $A^E \equiv \bigcup_{t=1}^{\infty} (\Theta_t \times \{t\})$ and denote by $B^E$ the set of all bounded functions from $A^E$ to $\mathbb{R}$. The solution to the aforementioned trade-off can be represented as a value function $W^E \in B^E$ that, for any $t$, any $\theta_t$, gives the firm’s expected continuation payoff when the incumbent manager’s productivity is $\theta_t$ and he has been employed for $t$ periods. Clearly, the value $W^E(\theta_t, t)$ takes into account the possibility of replacing the manager at the end of period $t$, as well as at the end of any subsequent period. The function $W^E$ can be conveniently described as the fixed point of the mapping defined in the following proposition.

**Proposition 1** The efficient effort and turnover policies satisfy the following properties.\(^{32}\) (i) For all $t$, all $\theta^t \in R^t$, $\xi^E(\theta^t) = e^E$, with $e^E$ implicitly defined by $\psi'(e^E) = 1$. (ii) There exists a sequence of thresholds $(\theta^E_t)_{t=1}^{\infty}$ such that, conditional on being employed in period $t$, the manager is

\(^{31}\)Because all managers and all firms are identical and because payoffs are transferable, these policies also maximize the ex-ante sum of all managers’ expected payoffs and of all firms’ expected profits.

\(^{32}\)The efficient policies are "essentially unique", i.e., unique up to a zero-measure set of histories.
retained at the end of period \( t \) if and only if \( \theta_t \geq \theta^*_t \). The thresholds \( \{\theta^*_t\}_{t=1}^\infty \) are defined as follows. Let \( W^E \) be the unique fixed point to the mapping \( T_E : B^E \to B^E \) defined, for all \( W \in B^E \), all \( (\theta_t, t) \in A^E \), by \( T_E W(\theta_t, t) = \theta_t + e^E - \psi(e^E) + \delta \max \{ \mathbb{E}_{\theta_{t+1} | \theta_t}[W(\tilde{\theta}_{t+1}, t + 1) | \mathbb{E}_{\tilde{\theta}_1}[W(\tilde{\theta}_1, 1)] \}. \) For any \( t, \theta^*_t = \inf \{ \theta_t \in \Theta : \mathbb{E}_{\theta_{t+1} | \theta_t}[W^E(\tilde{\theta}_{t+1}, t + 1)] \geq \mathbb{E}_{\tilde{\theta}_1}[W^E(\tilde{\theta}_1, 1)] \}, \) unless the set is empty, in which case \( \theta^*_t = \tilde{\theta}_t \).

The proof uses the Contraction Mapping Theorem to establish existence and uniqueness of a function \( W^E \) that is a fixed point to the mapping \( T_E : B^E \to B^E \) defined in the proposition. It then shows that this function is indeed the value function for the problem described above. Finally, it establishes that, for any \( t \), the function \( W^E(\cdot, t) \) is nondecreasing. These properties, together with the assumptions that the process is Markov and satisfies the property of first-order stochastic dominance in types in turn imply that turnover decisions must be taken according to the cut-off rule given in the proposition.

4 The profit-maximizing contract

We now turn to the contract that maximizes each representative firm’s expected profits in a setting where neither the managers’ productivities nor their effort choices are observable. We begin by providing some sufficient conditions for given effort and turnover policies to be implementable.\(^{33}\)

Note that these policies are defined over the supports of the marginal distributions and thus specify effort and retention decisions also for zero-measure histories \( \theta^t \not\in R^t \). The reason for extending the policies from \( R^t \) to \( \Theta^t \) is that this permits us to specify a course of action also off-equilibrium, i.e., for sequences of reports that reveal a departure from a truthful and obedient strategy in previous periods. This in turn facilitates the verification of incentive compatibility on the equilibrium path.

**Proposition 2** Let \( \xi \) and \( \kappa \) be effort and turnover policies that depend on \( \theta \) only. Suppose that \( \xi \) and \( \kappa \) are such that the following single-crossing conditions hold for all \( t \geq 1 \), all \( \theta_t, \tilde{\theta}_t \in \Theta_t \), all \( \tilde{\theta}^{t-1} \in \Theta^{t-1} \) such that \( \kappa_{t-1}(\tilde{\theta}^{t-1}) = 1 \):

\[
\mathbb{E}_{\theta_{t}^{\infty} | \theta_t} \left[ \sum_{k=t}^{\tau(\theta^{t-1}, \tilde{\theta}_{t}^{\infty})} \delta^{k-t} \psi\left( \xi_{k}(\tilde{\theta}^{t-1}, \theta_t, \tilde{\theta}^{k}_{>t}) \right) - \sum_{k=t}^{\tau(\tilde{\theta}^{t-1}, \theta_t, \tilde{\theta}^{\infty}_{>t})} \delta^{k-t} \psi\left( \xi_{k}(\tilde{\theta}^{t-1}, \tilde{\theta}_t, \tilde{\theta}^{k}_{>t}) \right) \right] \geq [0 - \tilde{\theta}_t] \geq 0. \tag{4}
\]

Then there exists a linear reward scheme of the form

\[
s_t(\theta^t, \pi_t) = S_t(\theta^t) + \alpha_t(\theta^t)\pi_t \text{ all } t, \text{ all } \theta^t \in \Theta^t, \tag{5}
\]

such that, irrespective of the distribution \( \Phi \) of the transitory noise, the following are true: (i) after any history \( h_t = (\theta^t, \tilde{\theta}^{t-1}, e^{t-1}, \pi^{t-1}) \) such that \( \kappa_{t-1}(\tilde{\theta}^{t-1}) = 1 \), each manager prefers to follow a

\(^{33}\)By “implementable” we mean such that there exists a compensation scheme \( s \) that, given \( \xi \) and \( \kappa \), induces the manager to follow a truthful and obedient strategy.
truthful and obedient strategy in the entire continuation game that starts in period $t$ with history $h_t$ than following any other strategy; (ii) the lowest period-1 type’s expected payoff from following a truthful and obedient strategy in the entire game is exactly equal to his outside option $U_o$; and (iii) after any history $h_t = (\theta_t, \hat{\Theta}^{t-1}, e_t^{t-1}, \pi_t^{t-1})$ such that $\kappa_t = 1$, each manager’s continuation payoff under a truthful and obedient strategy remains at least as high as his outside option $U_o$.

The single-crossing conditions in the proposition say that higher reports about current productivity lead, on average, to higher chances of retention and to higher effort choices both in the present as well as in subsequent periods, where the average is weighted by the impulse responses.

A special case of interest is when both the turnover and the effort policies are strongly monotone, i.e., when each $\xi_t(\cdot)$ and $\kappa_t(\cdot)$ is nondecreasing in $\theta_t$. This property is satisfied, for example, by the efficient policies of Proposition 1, appropriately extended from $R^t$ to $\Theta^t$ (see Corollary 2). Below, we will identify conditions on the primitives of the environment (namely on the stochastic process $F$ governing the evolution of the managers’ productivities) that guarantee that this property is satisfied also by the policies that maximize each firm’s expected profits. However, as the result in the proposition makes clear, these conditions are stronger than needed. In fact, linear schemes permit the firms to also implement policies that are not strongly monotone. Furthermore, policies which are not implementable by linear schemes may be implementable by other schemes.

One of the reasons for focusing on linear schemes, in addition to their simplicity and the fact that they are often used in practice, is the following: These schemes do not require any knowledge, either by the firms, or by the managers, or by both, about the details of the distribution $\Phi$ of the transitory noise terms (see also Caillaud, Guesnerie, and Rey, 1992, for a related result in a static setting). Another advantage of the proposed schemes is that they guarantee that, at any period $t \geq 1$, if the manager finds it optimal to stay in the relationship and follow a truthful and obedient strategy from period $t$ onwards after a history in which he has been truthful and obedient in all past periods, he then also finds it optimal to do the same after any other history. In both respects, the proposed schemes thus offer a form of robust implementation.

Turning to the two components $\alpha$ and $S$ of the proposed linear scheme, the coefficients

$$\alpha_t(\theta_t) \equiv \psi_t(\xi_t(\theta_t))$$

where chosen so as to provide the manager with the right incentives to choose effort obediently. Because neither future cash flows nor future retention decisions depend on current cash flows (and, as a result, on current effort), it is easy to see that, when the sensitivity of the manager’s compensation to the current cash flows is given by (6), by choosing effort $e_t = \xi_t(\theta_t)$, the manager equates the marginal disutility of effort to its marginal benefit and hence maximizes his continuation payoff. This is irrespective of whether or not the manager has reported his productivity truthfully. Under
the proposed scheme, the moral-hazard part of the problem is thus controlled entirely through the variable components \( \alpha \equiv (\alpha_t(\cdot))_{t=1}^{\infty} \).

Given \( \alpha \), the fixed components \( S \equiv (S_t(\cdot))_{t=1}^{\infty} \) are then chosen to control for the adverse-selection part of the problem, i.e., to induce the managers to reveal their productivity. As we show in the Appendix, this can be accomplished by setting each fixed component equal to

\[
S_t(\theta^t) = \psi(\xi_t(\theta^t)) - \alpha_t(\theta^t)(\xi_t(\theta^t) + \theta_t) + (1 - \delta) U^o
\]

where

\[
\psi(\xi_t(\theta^t)) = \beta_t(\theta^t) \xi_t(\theta^t) \quad \text{and} \quad \beta_t(\theta^t) = \beta_{t-1}(\theta_{t-1}^t) \psi(\xi_{t-1}(\theta^t))
\]

for \( t = 1, 2, \ldots \), and

\[
\alpha_t(\theta^t) = \beta_{t-1}(\theta_{t-1}^t) \psi(\xi_{t-1}(\theta^t))
\]

for \( t = 2, 3, \ldots \), with \( \beta_1(\theta_1^1) = 1 \). The function \( \psi(\cdot) \) denotes the manager’s period-(t+1) continuation payoff (over and above his outside option) under the truthful and obedient strategy. Combining the two components \( \alpha \) and \( S \), one can then verify that, when the policies \( \xi_t \) and \( \kappa_t \) satisfy the single-crossing conditions in the proposition, then after any history \( (\theta^t, \theta^t_{t-1}, \psi_t, \psi_{t-1}, \ldots, \psi_{t-1}) \), the manager finds one-stage deviations from the truthful and obedient strategy unprofitable. Together with a form of continuity-at-infinity discussed in the Appendix, this then implies that no other deviations are profitable either.

By inspecting the formula for the linear scheme in (5) and using the formula for the fixed component in (7), one can also verify that when (i) the transitory noise shocks \( \nu \) are not too negative, (ii) the level of the outside option \( U^0 \) is not too small, and (iii) the discount factor \( \delta \) is not too high, then the above linear schemes entail a nonnegative payment to the manager in every period and for any history. In this case, the proposed schemes offer a valid implementation also in settings in which the managers are protected by limited liability. Because both the efficient policies identified in Proposition 1 as well as the profit-maximizing policies identified below will be implemented with linear schemes, this result shows that neither our positive nor our normative results depend critically on our simplifying assumption of disregarding limited liability (or cash) constraints.

So far, we have shown that linear compensation schemes permit the firms to sustain a fairly rich set of effort and turnover policies. We now show that, under certain regularity conditions, the policies that maximize the firms’ expected profits indeed belong to this set. From arguments similar to those in Garrett and Pavan (2010) one can verify that, in any incentive-compatible mechanism \( \Omega \equiv (\xi, s, \kappa) \), each manager’s period-1 expected payoff under a truthful and obedient strategy \( V^\Omega(\theta_1) \) must satisfy

\[
V^\Omega(\theta_1) = V^\Omega(\theta_1) + \int_{\theta_1}^{\theta_1} \beta_{s_{t>1}}^\infty \delta^{t-1} J^t_1(s, \theta^t_{s_{t>1}}) \psi'(\xi_t(s, \theta^t_{s_{t>1}})) \, ds.
\]
Importantly, note that the formula in (9) must hold irrespective of whether or not the compensation is linear. That (9) is necessary for incentive-compatibility in fact follows by applying the dynamic envelope theorem of Pavan, Segal, and Toikka (2009) to the problem under examination here.

The formula in (9) confirms the intuition that the expected surplus that the principal must leave to each period-1 type is determined by the dynamics of effort and retention decisions under the contracts offered to the less productive types. As anticipated in the Introduction, this is because those managers who are most productive at the contracting stage expect to be able to obtain a "rent" when mimicking the less productive types. This rent originates from the possibility of generating the same cash flows as the less productive types by working less, thus economizing on the disutility of effort. The amount of effort they expect to save must, however, take into account the fact that their own productivity, as well as that of the types they are mimicking, will change over time. This is done by weighting the amount of effort saved in all subsequent periods by the impulse response functions \( J_t \), which, as explained above, control for how the effect of the initial productivity on future productivity evolves over time.

Using (9), the firm’s expected profits from each manager it hires can be conveniently expressed as

\[
\mathbb{E}_{\tilde{\theta}^\infty, \theta^\infty} \left[ \sum_{t=1}^{\tau(\tilde{\theta}^\infty)} \delta^{t-1} \left\{ \tilde{\theta}_t + \xi_t(\tilde{\theta}_t^*) + \nu_t - \psi(\xi_t(\tilde{\theta}_t^*)) - \eta(\tilde{\theta}_1, J_t^*(\tilde{\theta}_t^*)) \right\} - (1 - \delta) U^o \right] + U^o - V^\Omega(\tilde{\theta}_1), \tag{10}
\]

where \( \eta(\theta_1) \equiv \frac{1 - F_1(\theta_1)}{f_1(\theta_1)} \) denotes the inverse hazard rate of the first-period distribution. The formula in (10) is the dynamic analog of the familiar virtual surplus formula for static adverse selection settings. It expresses the firm’s expected profits as the discounted expected total surplus generated by the relationship, net of terms that control for the surplus that the firm must leave to the manager (over and above his outside option) to induce him to truthfully reveal his private information.

Equipped with the aforementioned representation, now consider the “relaxed program” that consists of choosing policies \((\xi_1(\cdot), \kappa_1(\cdot))_{t=1}^\infty\) so as to maximize the firm’s total expected profits, taking into account that the expected net contribution of each manager is given by (10), and subject to the participation constraints of the lowest period-1 types \( V^\Omega(\tilde{\theta}_1) \geq U^o \). Hereafter, we denote by \((\xi_1^*(\cdot), \kappa_1^*(\cdot))_{t=1}^\infty\) the policies that solve such a relaxed program. Using the result in Proposition 2, it is then easy to see that these policies indeed maximize the firm’s total expected profits whenever there exists an extension of such policies from \( R \) to \( \Theta \) such that the extended policies satisfy all the single-crossing conditions in Proposition 2. In the following proposition, we thus first characterize the policies \((\xi_1^*(\cdot), \kappa_1^*(\cdot))_{t=1}^\infty\) that solve the relaxed program. We then provide a simple sufficient condition for the existence of such an extension. (Recall that the role of these extensions is to permit the managers to truthfully reveal their types also off-equilibrium, i.e., after histories that involve a departure from truthful and obedient behavior in past periods.)

Let \( A \equiv \cup_{t=1}^\infty \left( R^t \times \{ t \} \right) \) and denote by \( B \) the space of bounded functions from \( A \) to \( \mathbb{R} \).
Proposition 3 Consider the policies $\xi^\ast$ and $\kappa^\ast$ defined by (i) and (ii) below. (i) For all $t$, all $\theta^t \in \mathbb{R}^t$, the effort policy $\xi^\ast_t (\theta^t)$ is implicitly given by

$$
\psi'(\xi^\ast_t (\theta^t)) = 1 - \eta(\theta_1)J_1^1(\theta^t)\psi''(\xi^\ast_t (\theta^t)).
$$

(ii) Let $W^\ast$ be the unique fixed point to the mapping $T : \mathcal{B} \to \mathcal{B}$ defined, for all $W \in \mathcal{B}$, all $(\theta^t, t) \in A$, by

$$
TW(\theta^t, t) \equiv \xi^\ast_t (\theta^t) + \theta_t - \psi(\xi^\ast_t (\theta^t)) - \eta(\theta_1)J_1^1(\theta^t)\psi'(\xi^\ast_t (\theta^t)) + \delta \max\{\mathbb{E}_{\theta^{t+1}}[W(\tilde{\theta}^{t+1}, t+1)], \mathbb{E}_{\theta_1}[W(\tilde{\theta}, 1)]\}.
$$

For all $t$, all $\theta^t \in \mathbb{R}^t$, the retention policy $\kappa^\ast$ is such that, conditional on being employed in period $t$, the manager is retained at the end of period $t$ if and only if $\mathbb{E}_{\theta^{t+1}}[W^\ast(\tilde{\theta}^{t+1}, t+1)] \geq \mathbb{E}_{\theta_1}[W^\ast(\tilde{\theta}, 1)]$.

Suppose that there exists an extension of the policies $(\xi^\ast, \kappa^\ast)$ from $R$ to $\Theta$ such that the extended policies satisfy the single-crossing conditions of Proposition 2. Then any contract that maximizes the firm’s profits implements the policies $(\xi^\ast, \kappa^\ast)$ given above. A sufficient condition for such an extension to exist is that each function $\eta(\cdot)J_1^1(\cdot)$ is nonincreasing on $\mathbb{R}^t$, all $t$. Furthermore, when this is the case, the optimal retention policy takes the form of a cut-off rule: There exists a sequence of threshold functions $(\theta^\ast_t (\cdot))_{t=1}^\infty$, $\theta^\ast_t : \Theta^{t-1} \to \mathbb{R}$, all $t \geq 1$, such that, conditional on being employed in period $t$, the manager is retained at the end of period $t$ if and only if $\theta_t \geq \theta^\ast_t (\theta^{t-1})$, with $\theta^\ast_t (\cdot)$ nonincreasing.

Under the assumptions in the proposition, the effort and turnover policies that maximize the firm’s expected profits are thus the “virtual analogs” of the policies $\xi^E$ and $\kappa^E$ that maximize efficiency, as given in Proposition 1. Note that, in each period $t$, and for each history $\theta^t \in \mathbb{R}^t$, the optimal effort $\xi^\ast_t (\theta^t)$ is chosen so as to trade off the effect of a marginal variation in effort on total surplus $\epsilon_t + \theta_t - \psiT(\epsilon_t)$ with its effect on the managers’ informational rents, as computed from period one’s perspective (i.e., at the time the managers are hired). The fact that both the firm’s and the managers’ preferences are additively separable over time implies that this trade-off is unaffected by the possibility that the firm replaces the managers. Furthermore, because the rent $V^{\Omega}(\theta_1)$ that each type $\theta_1$ expects at the time he is hired is increasing in the effort $\xi_t(\hat{\theta}_1, \theta^t_{>1})$ that the firm asks each less productive type $\hat{\theta}_1 < \theta_1$ to exert in each period $t \geq 1$, the effort policy that maximizes the firm’s profits is systematically downward distorted with respect to its efficient counterpart $\xi^E$. Such a downward distortion should be expected, for it essentially comes from the same considerations as in familiar static models like Laffont and Tirole (1986).

More interestingly, note that, fixing the initial type $\theta_1$, the dynamics of effort in subsequent periods is entirely driven by the dynamics of the impulse response functions $J_1^1$. These functions,

---

34 For simplicity, we assume throughout that the profit-maximizing policy specifies positive effort choices in each period $t$ and for each history $\theta^t \in \mathbb{R}^t$. This amounts to assuming that, for all $t$ all $\theta^t \in \mathbb{R}^t$, $\psi''(0) < 1/\eta(\theta_1)J_1^1(\theta^t)$. When this condition does not hold, optimal effort is simply given by $\xi^\ast_t (\theta^t) = 0$.

35 Because $\theta_0$ is given, in period one, the cut-off function $\theta^\ast_1 : \mathbb{R}^0 \to \mathbb{R}$ reduces to a constant cut-off $\theta^\ast_1 \in \mathbb{R}$.
by describing the effect of period-one productivity on subsequent productivity, capture how the persistence of the managers’ initial private information evolves over time. Because such persistence is what makes more productive (period-one) types expect larger surplus in subsequent periods than initially less productive types, the dynamics of the impulse responses $J^1_t$ are what determine the dynamics of effort decisions $\xi^*_t$.

Next, consider the turnover policy, which is the focus of the analysis. The characterization of the profit-maximizing policy $\kappa^*$ parallels the one for the efficient policy $\kappa^E$ in Proposition 1. The proof in the Appendix first establishes existence of the value function $W^*$ associated with the problem that consists of choosing the turnover policy so as to maximize the firm’s virtual surplus (given for each manager by (10)) taking as given the profit-maximizing effort policy $\xi^*$. For any $(\theta^t, t) \in A$, $W^*(\theta^t, t)$ gives the firm’s expected value from continuing the relationship with a manager who has worked already for $t - 1$ periods and who will continue working for at least one more period (period $t$). As with the efficient policy, this value is computed taking into account future retention and effort decisions. However, contrary to the case of efficiency, the value $W^*(\theta^t, t)$ in general depends on the entire history of productivities $\theta^t$, as opposed to only the current productivity $\theta_t$. The reason is twofold. First, as shown above, the profit-maximizing effort policy typically depends on the entire history $\theta^t$. Second, even if effort were exogenously fixed at a constant level, the “virtual value” of continuing the relationship after $t$ periods would typically depend on the entire history $\theta^t$. The reason is that this value is computed from an ex-ante perspective, and takes into account managers’ informational rents, as expected at the time of hiring. As shown above, these rents are determined by the dynamics of the impulse response functions $J^1_t$. Because the rents typically depend on the entire history of productivity realizations $\theta^t$, so does the profit-maximizing turnover policy $\kappa^*$.\(^{36}\) The profit-maximizing turnover policy can then be determined straightforwardly from $W^*$: the incumbent manager is replaced whenever the expected value $E_{\theta_1} \left[ W^*(\theta_1, 1) \right]$ of starting a relationship with a new manager of unknown productivity exceeds the expected value $E_{\theta^{t+1}, \theta_t} \left[ W^*(\theta^{t+1}, t + 1) \right]$ of continuing the relationship with the incumbent manager who has experienced a history of productivities $\theta^t$. Once again, these values are calculated from the perspective of the time at which the incumbent is hired.

Having characterized the policies that maximize the firm’s virtual surplus, the second part of the proposition then identifies a simple sufficient condition for these policies to satisfy the single-crossing conditions of Proposition 2. Under the assumption that each function $\eta(\cdot)J^1_t(\cdot)$ is nonincreasing.\(^{37}\)

\(^{36}\)A notable exception is when $\theta_t$ evolves according to an autoregressive processes, as in Example 1 below. In this case, the impulse responses $J^1_t$ are scalars and the expected value from continuing a relationship after $t$ periods depends only on the current productivity $\theta_t$, the initial productivity $\theta_1$, and the length $t$ of the employment relationship.

\(^{37}\)Note that this assumption is the dynamic analog of the regularity condition for static mechanism design; it combines the familiar condition of monotone hazard rate of the first-period distribution $F_1$ with the assumption of nonincreasing impulse responses $J^1_t$.  

21
that the firm expects from an incumbent manager (net of information rents) who has been working for \( t \) periods are nondecreasing in \( t \). Together with the property of “first-order stochastic dominance in types”, this property in turn implies that the value \( W^*(\cdot, t) \) of continuing the relationship after \( t \) periods is nondecreasing. In this case, the turnover policy \( \kappa^* \) that maximizes the firm’s virtual surplus is nondecreasing and takes the form of a cut-off rule, with cut-off functions \( (\theta^*_t(\cdot))_{t=1}^{\infty} \) satisfying the properties in the proposition. Together with the fact that the effort policy \( \xi^* \) is also monotone, this assumption in turn guarantees that, starting from the policies \( (\xi^*, \kappa^*) \) that maximize the firm’s virtual surplus, one can construct an extension of these policies from \( R \) to \( \Theta \) so that the corresponding extended policies satisfy all the single-crossing conditions of Proposition 2. The result in that proposition then implies existence of a linear compensation scheme \( s^* \) such that the mechanism \( \Omega^* = (\xi^*, \kappa^*, s^*) \) is incentive compatible and individually rational and gives a manager with initial type \( \theta^*_{t_0} \) an expected payoff equal to his outside option. That the mechanism \( \Omega^* \) is optimal then follows directly from the fact that the firm’s expected profits from each manager it hires are given by (10) under any mechanism that is incentive compatible and individually rational for the manager.\(^{38}\)

Combining together the various conditions, the result in Proposition 3 identifies the policies that maximize the firm’s expected profits when the process that governs the evolution of the managers’ productivity satisfies the following properties: (i) the supports of the marginal distributions \( \Theta_t \) are uniformly bounded over \( t \); (ii) all kernels \( F_t(\cdot|\theta^{t-1}) \) are first-order Markov and satisfy the condition of first-order stochastic dominance in types; (iii) the impulse response functions \( J^*_t(\cdot) \) are uniformly bounded over \( t, \tau, \tau > t, \) and \( R^\tau \); (iv) each function \( \eta(\cdot)J^1_t(\cdot) \) is nonincreasing. Although these conditions are restrictive, they are stronger than necessary. As discussed above, condition (iv) is introduced only to guarantee that the effort and turnover policies that maximize the firm’s virtual surplus are monotone in each period \( t \), which is more than what is required to guarantee that the single-crossing conditions of Proposition 2 are satisfied. The policies of Proposition 3 must therefore remain optimal also for a larger class of processes.

For the purposes of establishing our key positive and normative results below, we will restrict attention to stochastic processes satisfying conditions (i)-(iv) above. Two examples of processes satisfying these conditions are given below.\(^{39}\)

---

\(^{38}\)By incentive-compatible we mean such that the truthful and obedient strategy is sequentially optimal at all histories. By individually rational we mean such that participation is sequentially optimal at all histories.

\(^{39}\)The proof that these processes satisfy all the conditions in Proposition 3 is quite straightforward and available upon request.
Example 1 Let $\Theta_1 \subset \mathbb{R}$ be a bounded interval and $F_1$ be an absolutely continuous c.d.f. with support $\Theta_1$, strictly increasing over $\Theta_1$, with nonincreasing inverse hazard rate $\eta(\cdot)$. Let $(\gamma_t)_{t=2}^\infty$ be a sequence of non-negative scalars such that $\sum_{s=1}^t (x_{t-s+1}^{t} \gamma_{s+1})$ is bounded uniformly across $t$.\(^{40}\) For each $t$, let $G_t$ be an absolutely continuous c.d.f. with support $[\xi_t, \xi_t]$, where $|\xi_t|$ and $|\xi_t|$ are bounded uniformly across $t$. Consider the following (possibly non-autonomous) first-order autoregressive process: (i) $\theta_1$ is drawn from $F_1$; (ii) for all $t \geq 2$, $\theta_t = \gamma_t \theta_{t-1} + \epsilon_t$, with $\epsilon_t$ drawn from $G_t$, independently from $\theta_1$ and $\epsilon_{-t}$. This process satisfies all the conditions of Proposition 3.

Example 2 Let $\Theta_1 \subset \mathbb{R}$ be a bounded interval and $F_1$ be an absolutely continuous c.d.f. with support $\Theta_1$, strictly increasing over $\Theta_1$, with nonincreasing inverse hazard rate $\eta(\cdot)$. Let $\Theta' = [\hat{\theta}, \bar{\theta}] \supset \Theta_1$ and take an arbitrary continuously differentiable function $z : \Theta' \to \mathbb{R}_{++}$ satisfying (i) $\bar{\theta} \leq \hat{\theta} - z(\theta)$, (ii) $z'(\cdot) / z(\cdot)$ nondecreasing, and (iii) $z'(\theta) \in [-1, 0]$ for all $\theta \in \Theta'$.\(^{41}\) For each $t \geq 2$, let $G_t$ be an absolutely continuous c.d.f. with support $[0, 1]$. Consider the following process: (i) $\theta_1$ is drawn from $F_1$; (ii) for all $t \geq 2$, $\theta_t = \hat{\theta} - z(\theta_{t-1}) \epsilon_t$, with $\epsilon_t$ drawn from $G_t$, independently from $\theta_1$ and $\epsilon_{-t}$. This process satisfies all the conditions of Proposition 3.

We are now ready to establish our key positive result. We start with the following definition.

Definition 1 The process $F$ satisfies the property of “declining impulse responses” if, for any $s > t \geq 1$, any $(\theta^t, \theta^s_{>t})$, $\theta_s \geq \theta_t$ implies that $J_1(\theta^t, \theta^s_{>t}) \leq J_1(\theta^t)$.

As anticipated in the Introduction, this property captures the idea that the effect of a manager’s initial productivity on his future productivities declines with the length of the employment relationship, a property that seems reasonable for most cases of interest. This property is satisfied, for example, by an autonomous AR(1) process with coefficient $\gamma$ of linear dependence smaller than one (a special case of the class of processes in Example 1) and by the class of processes in Example 2.

The following result about the dynamics of profit-maximizing turnover applies to processes that satisfy declining impulse responses and that are autonomous.

Proposition 4 Suppose that the process $F$ is autonomous and satisfies the property of declining impulse responses. Take an arbitrary pair of periods $s, t$, with $s > t$, and an arbitrary history of productivities $\theta^s = (\theta^t, \theta^s_{>t}) \in R^s$. Then $\theta_s \geq \theta_t$ implies that $\mathbb{E}_{\theta^{t+1}}[W^*(\bar{\theta}^{s+1}, s + 1)] \geq \mathbb{E}_{\theta^{t+1}}[W^*(\bar{\theta}^{t+1}, t + 1)]$.

Proposition 4 thus establishes that the value of continuing the relationship in period $s > t$ with a manager whose period-$s$ productivity is no lower than it was in period $t$ is at least as high as it was

\(^{40}\)An example is an autonomous AR(1) process with $\gamma_t = \gamma \in (0, 1)$ for all $t$.

\(^{41}\)An example is $z(\theta) = e^{-\theta}$ for all $\theta \in [\hat{\theta}, \bar{\theta}]$, with $\bar{\theta} > 0$ and $\bar{\theta} - e^{-2} \bar{\theta} \geq \hat{\theta}$.
in period \( t \). This result follows from the fact that, when the process is autonomous and satisfies the property of declining impulse responses then, for any given productivity, the net flow payoff that the firm expects (ex-ante) from retaining the incumbent, as captured by (12), increases with the length of the employment relationship, a property which is then inherited by the value function \( W^* \).

The following corollary is then an immediate implication of Proposition 4.

**Corollary 1** Suppose that the conditions in Proposition 4 hold. Take an arbitrary period \( s > 1 \) and let \( \theta^s \in \mathbb{R}^s \) be such that \( \kappa_{s-1}^s(\theta^{s-1}) = 1 \). Furthermore, suppose \( \theta^s \) is such that \( \theta_s \geq \theta_t \) for some \( t < s \). Then \( \kappa_s(\theta^s) = 1 \).

The result in Corollary 1 says that the productivity level which the firm requires for retention declines with the length of the employment relationship. That is, the manager is retained in any period \( s \) whenever his period-\( s \) productivity is no lower than in all previous periods. In other words, when separation occurs, it must necessarily be the case that the manager’s productivity is at its historical lowest.

The reason why the retention policy becomes gradually more permissive over time is the one anticipated in the Introduction. When the effect of the initial productivity on future productivity declines over time, a commitment to replace a manager in the distant future is less effective in reducing the informational rent that the manager is able to obtain thanks to his initial private information than a commitment to replace him in the near future (for given productivity at the time of dismissal).

The result that the optimal turnover policy becomes more permissive over time together with the result that the productivity level \( \theta^1(\theta^0) \) required for retention decreases with the productivity experienced in past periods may help explain the practice of rewarding managers that are highly productive at the early stages (and hence, on average, generate higher profits) by offering them job stability once their tenure in the firm becomes long enough. Thus what in the eyes of an external observer may look like "entrenchment" can actually be the result of a profit-maximizing contract in a world where managerial productivity is expected to change over time and be the managers’ private information. Importantly, note that this property holds independently of the level of the managers’ outside option \( U^0 \). We thus expect such a property to hold irrespective of whether one looks at a given firm or at the entire market equilibrium.\(^{42}\)

\(^{42}\)Of course, as explained above, this is true only as long as search and matching frictions prevent managers from appropriating the entire surplus.

It is, however, important to recognize that, while the property that retention decisions become more permissive over time holds when conditioning on productivity (equivalently, on match quality), it need not hold when averaging across the entire pool of productivities of retained managers. Indeed, while the probability of retention for a given productivity level necessarily increases with tenure,
the unconditional probability of retention need not be monotonic in the length of the employment relationship because of composition effects that can push in the opposite direction. It is thus essential for the econometrician testing for our positive prediction to collect data that either directly, or indirectly, permit him to condition on managerial productivity.

5 On the (in)efficiency of optimal retention decisions

Having established that profit-maximizing firms offer, in equilibrium, contracts that lead to more permissive retention decisions over time, we now turn to the normative implications of this result. We start by verifying that the first-best effort and turnover policies identified in Section 3 remain implementable also when productivity and effort choices are the managers’ private information.

**Corollary 2** Assume that both productivity and effort choices are the managers’ private information. There exists a linear compensation scheme of the type described in Proposition 2 that implements the first-best effort and turnover policies of Proposition 1.\(^\text{43}\)

We can now compare the firms’ profit-maximizing policies with their efficient counterparts. As shown in the previous section, when impulse responses decline over time and eventually vanish in the long run, effort under the firms’ optimal contracts gradually converges to its efficient level as the length of the employment relationship grows sufficiently large. One might expect a similar result to apply also to retention decisions. This conjecture, however, fails to take into account that firms’ separation payoffs are endogenous and affected by the same frictions as in the relationship with each incumbent. Taking this endogeneity into account is indeed what leads to the fixed-point result in Proposition 3. Having endogenized the firms’ separation payoff in turn permits us to establish our key normative result that, once the length of the employment relationship has grown sufficiently large, profit-maximizing retention decisions eventually become excessively permissive as compared to what efficiency requires. We formalize this result in Proposition 5 below. Before doing that, as a preliminary step to understanding the result, it is useful to first consider the following simplified example.

**Example 3** Consider a firm operating for only two periods and assume that this is commonly known. In addition, suppose that both \(\theta_1\) and \(\varepsilon_2\) are uniformly distributed over \([-0.5, +0.5]\) and that \(\theta_2 = \gamma \theta_1 + \varepsilon_2\). Finally, suppose that \(\psi(\varepsilon) = \varepsilon^2/2\) for all \(\varepsilon \in [0,1]\), and that \(U^o = 0\). In this example, the profit-maximizing contract induces too much (respectively, too little) turnover if \(\gamma > 0.845\) (respectively, if \(\gamma < 0.845\)).

\(^{43}\)The result follows by extending the efficient policies \(\xi^E\) and \(\kappa^E\) from \(R\) to \(\Theta\) in a way that preserves strong monotonicity (as in the proof of Proposition 3), and then applying Proposition 2 to the extended policies.
The relation between the profit-maximizing thresholds $\theta_1^*$ and the impulse responses $J_1^2 = \gamma$ are depicted in Figure 1 below (the efficient threshold is $\theta_1^E = 0$).

![Graph showing the relation between $\gamma$ and $\theta_1^*$](image)

The example indicates that whether the profit-maximizing threshold for retention is higher or lower than its efficient counterpart depends crucially on the magnitude of the impulse response with respect to the first-period productivity. When $\gamma$ is small, the effect of $\theta_1$ on $\theta_2$ is small, in which case the firm can appropriate a large fraction of the surplus generated by the incumbent in the second period. As a result, the firm optimally commits in period one to retain the incumbent for a large set of his period-one productivities. In particular, when $\gamma$ is very small (i.e., when $\theta_1$ and $\theta_2$ are almost independent) the firm optimally commits to retain the incumbent irrespective of his period-one productivity. Such a low turnover is clearly inefficient, for efficiency requires that the incumbent be retained only when his expected period-2 productivity is higher than that of a newly hired manager, which is the case only when $\theta_1 \geq \theta_1^E = 0$.

On the other hand, when $\gamma$ is close to 1, the threshold productivity for retention under the profit-maximizing policy is higher than the efficient one. To see why, suppose that productivity is fully persistent, i.e. that $\gamma = 1$. Then, as is readily checked, $VS_1(\theta_1) = \mathbb{E}_{\theta_2}[VS_2(\theta_1, \theta_2)]$, where the functions $VS_1$ and $VS_2$ are given by (12). In this example, $VS_1$ is strictly convex. Noting that $\theta_1^E = \mathbb{E}[\tilde{\theta}_1]$, we then have that $\mathbb{E}[VS_1(\tilde{\theta}_1)] > VS_1(\theta_1^E) = \mathbb{E}_{\theta_2}[VS_2(\theta_1^E, \tilde{\theta}_2)]$, i.e., the expected value of replacing the manager is greater than the value from keeping the manager when his first-period productivity equals the efficient threshold. The same result holds for $\gamma$ close to 1. When productivity is highly persistent, the firm’s optimal contract may thus induce excessive firing (equivalently, too high a level of turnover) as compared to what is efficient.

As shown below, the above comparative statics have a natural analog in a dynamic setting by replacing the degree of serial correlation $\gamma$ in the example with length of the employment relationship. We start with the following definition.

**Definition 2** The process $F$ satisfies the property of “vanishing impulse responses” if, for any $\epsilon > 0$, there exists $t_\epsilon$ such that, for all $t > t_\epsilon$, $\eta(\theta_1)J_1^2(\theta_1) < \epsilon$ for all $\theta_1 \in \mathbb{R}$.

This condition simply says that the effect of the managers’ initial productivity on their subsequent productivity eventually vanishes after sufficiently long tenure, and that this occurs uniformly over
all histories. Under this assumption, which seems plausible for most cases of interest, we then have the following result.

**Proposition 5** Suppose that the process $F$ satisfies the property of vanishing impulse responses.

(i) There exists $\bar{t}$ such that, for any $t > \bar{t}$, any $\theta^t \in R^t$ such that $\theta_t \geq \theta^E_t$, $\mathbb{E}_{\theta^{t+1}}[W^*(\theta^{t+1}, t+1)] > \mathbb{E}_{\theta_{1}}[W^*(\theta_{1}, 1)]$. (ii) Suppose that, in addition to the above assumptions, $F$ satisfies the following properties: (a) there exists a constant $\beta \in \mathbb{R}_{++}$ such that, for each $t \geq 2$, each $\theta_t \in \Theta_t$, the function $\eta(\theta_1)J_1^t((\theta_1, \cdot))$ is Lipschitz continuous over $\Theta^t(\theta_1) \equiv \{\theta_{t-1} \in \Theta_{t-1} : (\theta_1, \theta_{t-1}) \in R^t\}$ with Lipschitz constant $\beta$; and (b) there exists a constant $\rho \in \mathbb{R}_{++}$ such that, for each $t \geq 2$, each $\theta_t \in \Theta_t$, the function $f_t(\theta_t|\cdot)$ is Lipschitz continuous with Lipschitz constant $\rho$. Then there exists $\bar{t}$ such that, for any $t > \bar{t}$, any $\theta^{t-1} \in R^{t-1}$ such that $\kappa^{t-1}_t(\theta^{t-1}) = 1$, if $\theta^E_{t-1} \in \text{int}(\text{Supp}[F_t(\cdot|\theta_{t-1})])$, then $\theta^E_{t-1}(\theta^{t-1}) < \theta^E_t$.

Part (i) of Proposition 5 establishes existence of a critical length $\bar{t}$ for the employment relationship such that each manager who is retained after $t > \bar{t}$ periods under the efficient contract when his period-$t$ productivity is $\theta_t$, he is also retained under the profit-maximizing contract when his period-$t$ productivity is the same. The additional conditions of Part (ii) imply continuity in $\theta_t$ of the expected continuation payoffs $\mathbb{E}_{\theta^{t+1}}[W^*(\theta^{t+1}, t+1)]$ and $\mathbb{E}_{\theta_{t+1}}[W^E(\theta_{t+1}, t+1)]$ for any period $t \geq 2$ and history of productivities $\theta^{t-1} \in R^{t-1}$. This in turn establishes that the profit-maximizing retention thresholds will eventually become strictly smaller than their efficient counterparts (provided the efficient thresholds are interior to the support of the conditional distributions $F_t(\cdot|\theta_{t-1})$).

The proof for Proposition 5 can be understood heuristically by considering the “fictitious problem” that consists of maximizing the firm’s expected profits in a setting where the firm can observe its incumbent manager’s types and effort choices, but not those of its future hires. In this environment, the firm optimally asks the incumbent to follow the efficient effort policy in each period, it extracts all surplus from the incumbent (i.e., the incumbent receives his outside option), and offers the contract identified in Proposition 3 to each new hire.

Now, consider the actual problem. After a sufficiently long tenure, the cutoffs for retaining the incumbent in this problem must converge to those in the fictitious problem. The reason is that effort and retention decisions after a sufficiently long tenure have almost no effect on the incumbent’s ex-ante information rent. Together with the fact that the firm’s “outside option” (i.e., its expected payoff from hiring a new manager) is the same in the two problems, this implies that the firm’s decision on whether or not to retain the incumbent must eventually coincide in the two problems.

Next, note that the firm’s outside option in the fictitious problem is strictly lower than the firm’s outside option in a setting where the firm can observe all managers’ types and effort choices. The reason is that, with asymmetric information, it is impossible for the firm to implement the efficient policies while extracting all surplus from the managers, whilst this is possible with symmetric
information. It follows that, after a sufficiently long tenure, the value the firm assigns to retaining the incumbent relative to hiring a new manager is necessarily higher in the fictitious problem (and therefore in the actual one) than in a setting with symmetric information; the profit the firm obtains in each period the incumbent is in power is the same, while the payoff from hiring a new manager is lower. Furthermore, because the value the firm assigns to retaining the incumbent (relative to hiring a new manager) in a setting with symmetric information coincides with the one assigned by the planner when maximizing welfare,\textsuperscript{44} we have that the firm’s retention policy necessarily becomes more permissive than the efficient one after sufficiently long tenure.

The findings of Propositions 4 and 5 can be combined together to establish the following proposition, which summarizes our key normative results.

**Proposition 6** Suppose that the process $F$ governing the evolution of managerial productivity is autonomous and satisfies both the properties of declining and vanishing impulse responses. Suppose further that the profit-maximizing policy retains each manager after the first period with positive probability. Then either (i) the profit-maximizing contract induces excessive retention as compared to the efficient contract in each period and after almost any history $\theta^t \in R^t$; or (ii) there exist dates $t, \bar{t} \in \mathbb{N}$, with $2 \leq t \leq \bar{t}$, such that the following are true: (a) for any $t < t$, and almost any $\theta^t \in R^t$, if $\kappa_{t-1}(\theta^{t-1}) = 1$ and $\kappa_t^E(\theta^t) = 0$, then $\kappa_t^*(\theta^t) = 0$; and (b) for any $t > \bar{t}$, and almost any $\theta^t \in R^t$, if $\kappa_{\bar{t}-1}(\theta^{\bar{t}-1}) = 1$ and $\kappa_{\bar{t}}^*(\theta^{\bar{t}}) = 0$, then $\kappa_{\bar{t}}^E(\theta^{\bar{t}}) = 0$.

Profit-maximizing contracts thus either induce excessive retention (i.e., too little turnover) throughout the entire relationship, or they induce excessive firing at the early stages followed by excessive retention in the long run. Formally, there exist dates $t, \bar{t}$ such that any manager who is fired at the end of period $t < \bar{t}$ under the efficient policy is either fired at the end of the same period or earlier under the profit-maximizing contract, whereas any manager fired at the end of period $t > \bar{t}$ under the profit-maximizing contract is either fired at the end of the same period or earlier under the efficient policy.\textsuperscript{45}

### 6 Conclusions

We developed a tractable, yet rich, model of dynamic managerial contracting that explicitly accounts for the following possibilities: (i) turnover is driven by variations in the managers’ ability to generate profits for the firm (equivalently, in the match quality); (ii) variations in managerial productivity are

\textsuperscript{44}Recall that welfare under the efficient contract with asymmetric information coincides with the sum of the firm’s expected profits and of all the managers’ outside options under the contract that the firm would offer if information about all managers’ effort and productivities were symmetric.

\textsuperscript{45}One can easily verify that both cases (i) and (ii) of the proposition are possible. For instance, case (ii) obtains by extending Example 3 to an infinite horizon setting with $\delta > 0$ sufficiently small.
anticipated at the time of contracting but privately observed by the managers; (iii) at each point in time, the firm can go back to the labor market and replace an incumbent manager with a new hire; (iv) the firm’s prospects under the new hire are affected by the same information frictions as in the relationship with each incumbent.

Allowing for the aforementioned possibilities permitted us to identify important properties of the employment relationship. On the positive side, we showed that profit-maximizing contracts require job instability early in the relationship followed by job security later on. These dynamics balance the firm’s concern for responding promptly to variations in the environment that call for a change in management with its concern for limiting the level of managerial compensation that is necessary to induce a truthful exchange of information between the management and the board. What in the eyes of an external observer may thus look like "entrenchment" driven by poor governance or lack of commitment, can actually be the result of a fully optimal contract in a world where the board’s objectives are perfectly aligned with those of the shareholders. This result, however, does not mean that firms’ retention decisions are efficient. We showed that the contracts that firms offer to their top managers either induce excessive retention (i.e., insufficiently low turnover) at all tenure levels, or excessive firing at the early stages followed by excessive retention after long tenure.

Throughout the analysis, we maintained the assumption that the process that matches managers to firms is exogenous. Endogenizing the matching process is an important, yet challenging, direction for future research which is likely to shed further light on the joint dynamics of compensation, performance, and retention decisions.

References


Appendix

Proof of Proposition 1. That the efficient effort policy is given by $E_t (\theta_t) = e^E$ for all $t$, all $\theta_t$, follows directly from inspection of the firm’s payoff (3), the managers’ payoff (2), and the definition of cash flows (1).

Consider the retention policy. Because all managers are ex-ante identical, and because the process governing the evolution of the managers’ productivities is Markov, it is immediate that, in each period, the decision of whether or not to retain a manager must depend only on the length of the employment relationship $t$ and on the manager’s current productivity $\theta_t$. We will denote by $W^E : A^E \to \mathbb{R}$ the value function associated with the problem that involves choosing the efficient Markovian retention policy, given the constant effort policy described above. For any $(\theta_t, t) \in A^E$, $W^E (\theta_t, t)$ specifies the maximal continuation expected welfare that can be achieved when the incumbent manager’s productivity is $\theta_t$, he has been working for the firm for $t - 1$ periods already and will be working for the firm for at least one more period (recall that separation decisions must be planned one period in advance). It is immediate that $W^E$ is the value function of the problem described above only if it is a fixed point to the mapping $T^E$ given in the proposition.

Now let $\mathcal{N}^E \subset \mathcal{B}^E$ denote the space of bounded functions from $A^E$ to $\mathbb{R}$ that, for all $t$, are nondecreasing in $\theta_t$. Below, we first establish existence and uniqueness of a function $\hat{W}^E \in \mathcal{N}^E$ such that $T^E \hat{W}^E = \hat{W}^E$. Next, we verify that $W^E = \hat{W}^E$.

Note that the set $\mathcal{N}^E$, together with the uniform metric, is a complete metric space. Because the sequence of sets $(\Theta_t)_{t=1}^\infty$ is uniformly bounded, and because the process satisfies the assumption...
of first-order-stochastic dominance in types, \( N^E \) is closed under \( T_E \). Moreover, “Blackwell’s sufficient conditions” (namely, “monotonicity” and “discounting”, where the latter is guaranteed by the assumption that \( \delta < 1 \)) imply that \( T_E \) is a contraction. Therefore, by the Contraction Mapping Theorem (see, e.g., Theorem 3.2 of Stokey and Lucas, 1989), for any \( W \in N^E \), \( \hat{W}^E = \lim_{n \to \infty} T^E_n W \) exists, is unique, and belongs to \( N^E \).

Now, we claim that the following retention policy is efficient: for any \( t \), any \( \theta^t \in R^t \), \( \kappa_{t-1}(\theta^{t-1}) = 1 \) implies \( \kappa_t(\theta^t) = 1 \) if \( E_{\theta_{t+1}|\theta_t} [\hat{W}^E(\hat{\theta}_{t+1}, t + 1)] \geq E_{\theta_{1}|\theta_t} [\hat{W}^E(\hat{\theta}_1, 1)] \) and \( \kappa_t(\theta^t) = 0 \) otherwise. Note that, because the process satisfies the assumption of first-order-stochastic dominance in types, and because \( \hat{W}^E (., t) \) is nondecreasing for each \( t \), this retention policy is a cut-off policy. This property, together with the fact that the “flow payoffs” \( \theta_t + e^E - \psi(e^E) \) and \( \hat{W}^E \) are uniformly bounded on \( A^E \), then permit one to verify, via standard verification arguments, that the constructed policy is indeed efficient and that \( W^E = \hat{W}^E \).\(^{46} \)

**Proof of Proposition 2.** Consider the linear reward scheme \( s = (s_t : \Theta^t \times \mathbb{R} \to \mathbb{R})_{t=1}^\infty \) where \( s_t(\theta^t, \pi_t) = S_t(\theta^t) + \alpha_t(\theta^t) \pi_t \) for all \( t \), with \( \alpha_t(\theta^t) \) as in (6) and \( S_t(\theta^t) \) as in (7). Note that, because retention does not depend on cash flows, it does not affect the manager’s incentives for effort. From the law of iterated expectations, it then follows that, for any given history of reports \( \hat{\theta}^{t-1} \) such that the manager is still employed in period \( t \geq 1 \) (i.e., \( \kappa_{t-1}(\theta^{t-1}) = 1 \)) and for any period-\( t \) productivity \( \theta_t \), the manager’s continuation payoff at the beginning of period \( t \) when the manager plans to follow a truthful and obedient strategy from period \( t \) onwards is given by \( U^o + u_t(\theta_t; \hat{\theta}^{t-1}) \) where\(^{47} \)

\[
u_t(\theta_t; \hat{\theta}^{t-1}) = \int_{\hat{\theta}^t}^{\theta_t} E_{\theta_{t+1}|\theta_t} \left[ \sum_{k=1}^t \tau(\theta^{t-1}, s, \hat{\theta}_{>t}^k) \delta^{k-t} j_t^k(s, \hat{\theta}_{>t}^k) \psi(\xi_t(\hat{\theta}^{t-1}, s, \hat{\theta}_{>t}^k)) \right] ds.
\]

Because \( u_t(\theta_t; \hat{\theta}^{t-1}) \geq 0 \), the above scheme guarantees that, after any truthful and obedient history, the manager finds it optimal to stay in the relationship whenever the firm’s retention policy permits him to do so, which proves Part (iii) in the proposition. That \( u_1(\hat{\theta}_1) = 0 \) then implies Part (ii).

Consider then Part (i). Take an arbitrary history of past reports \( \hat{\theta}^{t-1} \). Suppose that, in period \( t \), the manager’s true type is \( \theta_t \) and that he reports \( \hat{\theta}_t \), then optimally chooses effort \( \xi_t(\hat{\theta}^{t-1}, \hat{\theta}_t) \) in period \( t \), and then, starting from period \( t+1 \) onwards, he follows a truthful and obedient strategy. One can easily verify that, under the proposed linear scheme, the manager’s continuation payoff is then given by

\[
\hat{u}_t(\theta_t; \hat{\theta}_t; \hat{\theta}^{t-1}) = u_t(\theta_t; \hat{\theta}^{t-1}) + \psi(\xi_t(\hat{\theta}^{t-1}, \hat{\theta}_t))(\theta_t - \hat{\theta}_t) + \delta \kappa_t(\theta_t; \hat{\theta}^{t-1}) \left\{ E_{\theta_{t+1}|\theta_t}[u_{t+1}(\hat{\theta}_{t+1}; \hat{\theta}^{t-1}, \hat{\theta}_t)] - E_{\hat{\theta}_{t+1}|\theta_t}[u_{t+1}(\hat{\theta}_{t+1}; \hat{\theta}^{t-1}, \hat{\theta}_t)] \right\}.
\]

\(^{46} \)This verification is standard in dynamic programming and hence omitted for brevity; the proof is available upon request.

\(^{47} \)Note that, under the proposed scheme, a manager’s continuation payoff depends on past announcements \( \hat{\theta}^{t-1} \), but not on past productivities \( \theta^{t-1} \), effort choices \( e^{t-1} \), or cash flows \( \pi^{t-1} \).
The single-crossing conditions in the proposition then imply that, for all \( t \), all \( \hat{\theta}^{t-1} \in \Theta^{t-1} \), all \( \theta_t, \hat{\theta}_t \in \Theta_t \),
\[
\left[ \frac{d u_t(\theta_t; \hat{\theta}^{t-1})}{d \theta_t} - \frac{d u_t(\theta_t, \hat{\theta}_t; \hat{\theta}^{t-1})}{d \theta_t} \right] \left[ \theta_t - \hat{\theta}_t \right] \geq 0.
\]

One can easily verify that this condition in turn implies that following a truthful and obedient strategy from period \( t \) onwards gives type \( \theta_t \) a higher continuation payoff than lying in period \( t \) by reporting \( \hat{\theta}_t \), then optimally choosing effort \( \xi_t(\hat{\theta}^{t-1}, \hat{\theta}_t) \) in period \( t \), and then going back to a truthful and obedient strategy from period \( t + 1 \) onwards.

Now, to establish the result in the proposition, it suffices to compare the manager’s continuation payoff at any period \( t \), given any possible type \( \theta_t \) and any possible history of past reports \( \hat{\theta}^{t-1} \in \Theta^{t-1} \) under a truthful and obedient strategy from period \( t \) onwards, with the manager’s expected payoff under any continuation strategy that satisfies the following property. In each period \( s \geq t \), and after any possible history of reports \( \hat{\theta}^s \in \Theta^s \), the effort specified by the strategy for period \( s \) coincides with the one prescribed by the recommendation policy \( \xi_s \); that is, after any sequence of reports \( \hat{\theta}^s \), effort is given by \( \xi_s(\hat{\theta}^s) \), where \( \xi_s(\hat{\theta}^s) \) is implicitly defined by
\[
\psi'(\xi_s(\hat{\theta}^s)) = \alpha_s(\hat{\theta}^s). \tag{13}
\]

Restricting attention to continuation strategies in which, at any period \( s \geq t \), the manager follows the recommended effort policy \( \xi_s(\hat{\theta}^s) \) is justified by: (i) the fact that the compensation paid in each period \( s \geq t \) is independent of past cash flows \( \pi^{s-1} \); (ii) under the proposed scheme, the manager’s period-\( s \) compensation, net of his disutility of effort, is maximized at \( e_s = \xi_s(\hat{\theta}^s) \); (iii) cash flows have no effect on retention. Together, these properties imply that, given any continuation strategy that prescribes effort choices different from those implied by (13), there exists another continuation strategy whose effort choices comply with (13) for all \( s \geq t \), all \( \hat{\theta}^s \), which gives the manager a (weakly) higher expected continuation payoff.

Next, it is easy to see that, under any continuation strategy that satisfies the aforementioned effort property, the manager’s expected payoff in each period \( s \geq t \) is bounded uniformly over \( \Theta_s \times \Theta^{s-1} \). In turn, this implies that a continuity-at-infinity condition similar to that in Fudenberg and Levine (1983) holds in this environment. Precisely, for any \( \epsilon > 0 \), there exists \( t \) large enough such that, for all \( \theta_t \in \Theta_t \), and all \( \hat{\theta}^{t-1}, \hat{\theta}^{t-1} \in \Theta^{t-1} \),
\[
\delta t \left| \hat{u}_t(\theta_t; \hat{\theta}^{t-1}) - \hat{u}_t(\theta_t; \hat{\theta}^{t-1}) \right| < \epsilon,
\]
where \( \hat{u}_t \) and \( \hat{u}_t \) are continuation payoffs under arbitrary continuation strategies satisfying the above effort restriction, given arbitrary histories of reports \( \hat{\theta}^{t-1} \) and \( \hat{\theta}^{t-1} \). This continuity-at-infinity property, together with the aforementioned property about one-stage deviations from a truthful and obedient strategy, imply that, after any history, the manager’s continuation payoff under a truthful and obedient strategy from that period onwards is weakly higher than the expected payoff under any other continuation strategy. We thus conclude that, whenever the pair of policies \((\xi, \kappa)\) satisfies all the single-crossing conditions in the proposition, it can be implemented by the proposed linear reward scheme. ■
Proof of Proposition 3. The proof proceeds in three steps. Step 1 characterizes the effort and retention policies $\xi^*$ and $\kappa^*$ (defined over the set of feasible productivity histories $R$) that maximize the firm’s ex-ante expected payoff, where the profits that the firm expects from each manager it hires are given by (10). Step 2 then shows that, when these policies can be extended from $R$ to $\Theta$ so that the extended policies satisfy all the single-crossing conditions of Proposition 2, then the firm implements the policies $\xi^*$ and $\kappa^*$ under any optimal contract. Lastly, Step 3 shows that a sufficient condition for such an extension to exist is that each function $h_t(\cdot) \equiv -\eta(\cdot)J_1^t(\cdot)$ is nondecreasing. This condition guarantees that both policies $\xi^*$ and $\kappa^*$ are nondecreasing, in which case it is always possible to construct an extension of these policies from $R$ to $\Theta$ so that the extended policies are also nondecreasing and hence trivially satisfy all the single-crossing conditions of Proposition 2. Furthermore, Step 3 shows that, in this case, the optimal retention policy takes the form of the cut-off rule in the proposition.

Step 1. First, consider the effort policy. It is easy to see that the policy $\xi^*$ that maximizes the firm’s total profits is independent of the retention policy $\kappa$ and is such that $\xi_t^*(\theta^t)$ is given by (11) for all $t$, all $\theta^t \in R^t$. Next, consider the retention policy. We first prove existence of a unique fixed point $W^* \in B$ to the mapping $T$ defined in the proposition. To this end, endow $B$ with the uniform metric. That $B$ is closed under $T$ is ensured by the restrictions on $\psi$ and by the definition of $\xi^*$, which together imply that each function $VS_t : R^t \to \mathbb{R}$ defined by

$$VS_t(\theta^t) \equiv \xi_t^*(\theta^t) + \theta_t - \psi(\xi_t^*(\theta^t)) - \eta(\theta_1)J_1^t(\theta^t) \psi(\xi_t^*(\theta^t)) - (1 - \delta)U_o$$

is uniformly bounded over $A$. Blackwell’s theorem implies that $T$ is a contraction mapping and the Contraction Mapping Theorem (see Stokey and Lucas, 1989) then implies the result. Standard arguments then permit one to verify that $W^*(\theta^t, t)$ is indeed the value function associated with the problem that consists in choosing a retention policy from period $t$ onwards that, given the history of productivities $\theta^t \in R^t$ and given the profit-maximizing effort policy $\xi^*$, maximizes the firm’s total expected discounted profits.\(^{48}\) Having established this result, it is then easy to see that any retention policy $\kappa^*$ that, given the effort policy $\xi^*$, maximizes the firm’s total profits must satisfy the conditions in the proposition.

Step 2. Suppose now that there exists an extension of the policies $(\xi^*, \kappa^*)$ from $R$ to $\Theta$ such that the extended policies satisfy all the single-crossing conditions of Proposition 2. The result in that proposition then implies existence of a linear compensation scheme $s^*$ such that the mechanism $\Omega^* = (\xi^*, s^*, \kappa^*)$ is incentive-compatible and satisfies the following properties: (i) it implements the policies $(\xi^*, \kappa^*)$; (ii) type $\theta_1$ obtains an expected payoff under $\Omega^*$ equal to his outside option, i.e., $V^{\Omega^*}(\theta_1) = U_o$; and (iii) the manager never finds it optimal to leave the firm when offered the possibility of staying. That the mechanism $\Omega^*$ is optimal then follows from the fact that the firm’s

---

\(^{48}\)The reason why the term $-(1 - \delta)U_o$ disappears from the mapping $T$ is that this term is constant across $t$ and across all managers.
expected profits from each manager that it hires are given by (10) under any mechanism that is incentive compatible and individually rational for the managers.

**Step 3.** Now assume that each function \( h_t(\cdot) \equiv -\eta(\cdot)J_t^i(\cdot) \) is nondecreasing. Because the function \( g(e, h, \theta) \equiv e + \theta - \psi(e) + hw'(e) \) has the strict increasing differences property with respect to \( e \) and \( h \), each function \( \xi^*_t(\cdot) \) is nondecreasing. This property follows from standard monotone comparative statics results by noting that, for each \( (\theta^t, t) \in A \), \( \xi^*_t(\theta^t) = \arg\max_{e \in E} g(e, h_t(\theta^t), \theta_t) \).

Next, we show that, for all \( t \), the function \( W^*(\cdot, t) \) is nondecreasing. To this aim, let \( \mathcal{N} \subset \mathcal{B} \) denote the set of all bounded functions from \( A \) to \( \mathbb{R} \) that, for each \( t \), are nondecreasing in \( \theta^t \). Note that, since \( -\eta(\cdot)J_t^i(\cdot) \) is nondecreasing, so is the function \( V_S_t(\cdot) \) — this is an immediate implication of the envelope theorem. This property, together with the fact that the process describing the evolution of the managers’ productivities satisfies the assumption of “first-order stochastic dominance in types” implies that \( \mathcal{N} \) is closed under the operator \( T \). It follows that \( \lim_{n \to \infty} T^n W \) is in \( \mathcal{N} \). The fact that \( T : \mathcal{B} \to \mathcal{B} \) admits a unique fixed point then implies that \( \lim_{n \to \infty} T^n W = W^* \).

The last result, together with “first-order stochastic dominance in types” implies that, for each \( t \), each \( \theta^{t-1} \in R^{t-1} \), \( \mathbb{E}_{\theta^{t+1} | \theta^{t-1}} [W^*(\theta^{t+1}, t+1)] \) is nondecreasing. Given the monotonicity of each function \( \mathbb{E}_{\theta^{t+1} | \theta^{t-1}} [W^*(\theta^{t+1}, t+1)] \), it is then immediate that the retention policy \( \kappa^* \) that maximizes the firm’s profits must be a cut-off rule with cut-off functions \( (\theta^*_t(\cdot))_{t=1}^{\infty} \) satisfying the conditions in the proposition. A sequence of cut-off functions \( (\theta^*_t(\cdot))_{t=1}^{\infty} \) satisfying these conditions is, for example, the following: for any \( t \), any \( \theta^{t-1} \in R^{t-1} \),

\[
\theta^*_t(\theta^{t-1}) = \begin{cases} 
-K & \text{if } \mathbb{E}_{\theta^{t+1} | \theta^t} [W^*(\theta^{t+1}, t+1)] > \mathbb{E}_{\theta^t} [W^*(\theta^t, 1)] \text{ for all } \theta_t \text{ s.t. } (\theta^{t-1}, \theta_t) \in R^t \\
K & \text{if } \mathbb{E}_{\theta^{t+1} | \theta^t} [W^*(\theta^{t+1}, t+1)] < \mathbb{E}_{\theta^t} [W^*(\theta^t, 1)] \text{ for all } \theta_t \text{ s.t. } (\theta^{t-1}, \theta_t) \in R^t \\
\min \{ \theta_t \in \Theta_t : (\theta^{t-1}, \theta_t) \in R^t \text{ and } \mathbb{E}_{\theta^{t+1} | \theta^t} [W^*(\theta^{t+1}, t+1)] \geq \mathbb{E}_{\theta^t} [W^*(\theta^t, 1)] \} & \text{if } \left\{ \theta_t \in \Theta_t : (\theta^{t-1}, \theta_t) \in R^t \text{ and } \mathbb{E}_{\theta^{t+1} | \theta^t} [W^*(\theta^{t+1}, t+1)] = \mathbb{E}_{\theta^t} [W^*(\theta^t, 1)] \right\} \neq \emptyset
\end{cases}
\]

where \( K \in \mathbb{R}_{++} \) is the uniform bound on each \( \Theta_t \), as defined in the model set-up.

Given the policies \( (\xi^*, \kappa^*) \), then consider the extension of these policies from \( R \) to \( \Theta \) constructed as follows. For any \( t \geq 2 \), any \( 2 \leq s \leq t \), let \( \varphi_s : \Theta^t \to \Theta^t \) be the function defined, for all \( \theta^t \in \Theta^t \), by

\[
\varphi_s(\theta^t) \equiv \begin{cases} 
\theta^t & \text{if } \theta_s \in \{ \text{Supp}[F_s(\cdot | \theta_{s-1})] \} \\
(\theta^{s-1}, \min \{ \text{Supp}[F_s(\cdot | \theta_{s-1})] \} , \theta_{s+1}, ..., \theta_t) & \text{if } \theta_s < \min \{ \text{Supp}[F_s(\cdot | \theta_{s-1})] \} \\
(\theta^{s-1}, \max \{ \text{Supp}[F_s(\cdot | \theta_{s-1})] \} , \theta_{s+1}, ..., \theta_t) & \text{if } \theta_s > \max \{ \text{Supp}[F_s(\cdot | \theta_{s-1})] \}
\end{cases}
\]

For all \( \theta_1 \in \Theta_1 \), let \( \lambda_1 (\theta_1) \equiv \theta_1 \), while for any \( t \geq 2 \), let \( \lambda_t : \Theta^t \to \Theta^t \) be the function defined, for all \( \theta^t \in \Theta^t \), by \( \lambda_t(\theta^t) \equiv \varphi_{t-1} \circ \varphi_{t-2} \circ \cdots \circ \varphi_2(\theta^t) \). Note that the function \( \lambda_t \) maps each vector of reports \( \theta^t \in \Theta^t \setminus R^t \) which reveals that the manager has been untruthful in previous periods into a vector of reports \( \hat{\theta}^t = \lambda_t(\theta^t) \) that is consistent with truth-telling. This is obtained by replacing recursively any report \( \theta_s \) that, given \( \theta_{s-1} \), is smaller than any feasible type with \( \hat{\theta}_s = \min \{ \text{Supp}[F_s(\cdot | \theta_{s-1})] \} \), and, likewise,
by replacing any report $\theta_s$ that is higher than any feasible type with $\hat{\theta}_s = \max\{\text{Supp}[F_s(\cdot|\theta_{s-1})]\}$. Now, let $\tilde{\xi}^*$ and $\tilde{\kappa}^*$ be the policies defined by $\tilde{\xi}^*(\theta^t) = \xi^*_t(\lambda_t(\theta^t))$ and $\tilde{\kappa}^*_t(\theta^t) = \kappa^*_t(\lambda_t(\theta^t))$, all $t$, all $\theta^t \in \Theta^t$. The property that each $\xi^*_t(\cdot)$ and $\kappa^*_t(\cdot)$ are nondecreasing implies that so are the policies $\tilde{\xi}^*_t(\cdot)$ and $\tilde{\kappa}^*_t(\cdot)$. This in turn guarantees that the extended policies $\tilde{\xi}^* = (\tilde{\xi}^*_t)_{t=1}^\infty$ and $\tilde{\kappa}^* = (\tilde{\kappa}^*_t)_{t=1}^\infty$ satisfy all the single-crossing conditions of Proposition 2.

**Proof of Proposition 4.** Let $\tilde{\mathcal{N}}$ denote the subclass of all functions $W \in \mathcal{B}$ satisfying the following properties: (a) for each $t$, $W(\cdot, t)$ is non-decreasing over $\Theta^t$; and (b) for any $s > t$, any $\theta^t \in \Theta^t$ and any $\theta^s_{>t}$ such that $\theta^s = (\theta^t, \theta^s_{>t}) \in R^s$, if $\theta_s \geq \theta_t$, then $W(\theta^s, s) \geq W(\theta^t, t)$.

We established already in the proof of Proposition 3 that the operator $T$ preserves property (a). The property of declining impulse responses, together with the property of first-order stochastic dominance in types, implies that $T$ also preserves (b). The unique fixed point $W^*$ to the mapping $T : \mathcal{B} \to \mathcal{B}$ thus satisfies properties (a) and (b) above. First-order stochastic dominance in types then implies that $\mathbb{E}_{\tilde{\theta}^t_0|\theta_0}[W^*(\tilde{\theta}^{t+1}_0, s + 1)] \geq \mathbb{E}_{\tilde{\theta}^t_0|\theta_0}[W^*(\tilde{\theta}^{t+1}_0, t + 1)]$.

**Proof of Example 3.** Note that $\eta(\theta_1) = \frac{1}{2} - \theta_1$. Thus, $\xi^*_1(\theta_1) = \frac{1}{2} + \theta_1$ and the payoff from hiring a new manager in period 2 is $\mathbb{E}[\xi^*_1(\theta_1) + \hat{\theta}_1 - \psi(\xi^*_1(\theta_1)) - \eta(\theta_1)\psi'(\xi^*_1(\theta_1))] = 1/6$. The manager is thus retained if and only if $\mathbb{E}_{\tilde{\theta}^t_2|\theta_1}[\xi^*_2(\theta_1) + \hat{\theta}_2 - \psi(\xi^*_2(\theta_1)) - \eta(\theta_1)\psi'(\xi^*_2(\theta_1))] \geq 1/6$, where $\xi^*_2(\theta_1) = 1 - \frac{1}{2} + \gamma \theta_1$ and $\mathbb{E}_{\tilde{\theta}^t_2|\theta_1}[\hat{\theta}_2] = \gamma \theta_1$. The inequality holds for all $\theta_1 \in [-\frac{1}{2}, +\frac{1}{2}]$ if $\gamma \leq 0.242$. Otherwise it holds if and only if $\theta_1 \geq \theta^*_1$ for some $\theta^*_1 \in (-\frac{1}{2}, +\frac{1}{2})$ such that $\theta^*_1 < 0$ if $\gamma \in (0.242, 0.845)$ and $\theta^*_1 > 0$ if $\gamma > 0.845$.

**Proof of Proposition 5.** The proof follows from five lemmas. Lemmas A1-A3 establish Part (i) of the proposition. Lemmas A4 and A5, together with Part (i), establish Part (ii).

Part (i). We start with the following lemma which does not require any specific assumption on the stochastic process and provides a useful property for a class of stopping problems with an exogenous separation payoff.

**Lemma A1.** For any $c \in \mathbb{R}$, there exists a unique function $W^{E,c} \in \mathcal{B}^E$ that is a fixed point to the mapping $T_{E,c} : \mathcal{B}^E \to \mathcal{B}^E$ defined, for all $W \in \mathcal{B}^E$, all $(\theta_t, t) \in A^E$, by

$$T_{E,c}W(\theta_t, t) = \theta_t + e^E - \psi(e^E) + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}^t_0|\theta_1}[W(\tilde{\theta}^t_{t+1}, t + 1)] ; c \right\}.$$ 

Fix $c', c'' \in \mathbb{R}$ with $c'' > c'$. There exists $\iota > 0$ such that, for all $t$, all $\theta_t \in \Theta_t$,

$$\mathbb{E}_{\tilde{\theta}^t_0|\theta_1}[W^{E,c''}(\tilde{\theta}^t_{t+1}, t + 1)] \geq c'' \Rightarrow \mathbb{E}_{\tilde{\theta}^t_0|\theta_1}[W^{E,c'}(\tilde{\theta}^t_{t+1}, t + 1)] > c' + \iota.$$

**Proof of Lemma A1.** Take any $c \in \mathbb{R}$. Because $\mathcal{B}^E$, together with the uniform metric, is a complete metric space, and because $T_{E,c}$ is a contraction, $T_{E,c}$ has a unique fixed point $W^{E,c} \in \mathcal{B}^E$. Now take a pair $(c'', c')$, with $c'' > c'$, and let $\mathcal{C}(c'', c') \subset \mathcal{B}^E$ be the space of bounded functions from
\( \mathcal{A} \) to \( \mathbb{R} \) such that, for all \( (\theta, t) \in \mathcal{A} \), \( W(\theta, t) \geq W^{E,c'}(\theta, t) - \delta(c'' - c') \). First note that \( \mathcal{C}(c'', c') \) is closed under \( T_{E,c'} \). To see this, take any \( W \in \mathcal{C}(c'', c') \). Then, for any \( (\theta, t) \),

\[
T_{E,c'} W(\theta, t) - W^{E,c''}(\theta, t) = T_{E,c'} W(\theta, t) - T_{E,c'} W^{E,c'}(\theta, t) \\
= \delta \left( \max \{ \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W(\tilde{\theta}_{t+1}, t+1); c'] \} - \max \{ \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W^{E,c''}(\tilde{\theta}_{t+1}, t+1); c''] \} \right) \\
\geq -\delta(c'' - c').
\]

Also, once endowed with the uniform metric, \( \mathcal{C}(c'', c') \) is a complete metric space. Hence, from the same arguments as in the proofs of the previous propositions, the unique fixed point \( W^{E,c'} \in \mathcal{B} \) to the operator \( T_{E,c'} \) must be an element of \( \mathcal{C}(c'', c') \). That is, for all \( (\theta, t) \), \( W^{E,c'}(\theta, t) - W^{E,c''}(\theta, t) \geq -\delta(c'' - c') \).

Finally, for any \( t \), any \( \theta_t \in \Theta_t \), if \( \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W^{E,c''}(\tilde{\theta}_{t+1}, t+1)] \geq c'' \), then

\[
\mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W^{E,c'}(\tilde{\theta}_{t+1}, t+1)] \geq \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W^{E,c''}(\tilde{\theta}_{t+1}, t+1)] - \delta(c'' - c') \\
\geq c'' - \delta(c'' - c') > c' + \delta
\]

for some \( \delta > 0 \). □

The next lemma establishes a strict ranking between the separation payoffs under the efficient and the profit-maximizing contracts.

**Lemma A2.** \( \mathbb{E}_{\tilde{\theta}_1} \left[ W^E(\tilde{\theta}, 1) \right] > \mathbb{E}_{\tilde{\theta}_1} \left[ W^*(\tilde{\theta}_1, 1) \right] \).

**Proof of Lemma A2.** Let \( \mathcal{D}(W^E) \subset \mathcal{B} \) be the space of bounded functions \( W \) from \( A \) to \( \mathbb{R} \) such that \( W(\theta^t, t) \leq W^E(\theta, t) \) for all \( (\theta^t, t) \in A \) (where \( A \equiv \cup_{t=1}^{\infty} (R^t \times \{t\}) \)). The set \( \mathcal{D}(W^E) \) is closed under \( T \), as defined in Proposition 3. To see this, let \( W \in \mathcal{D}(W^E) \). Then, for all \( (\theta^t, t) \in A \),

\[
TW(\theta^t, t) = \xi_t^*(\theta^t) + \eta_t(\theta^t) - \eta(\theta_t) J_1^t(\theta^t) - \psi(\xi_t^*(\theta^t)) + \delta \max \{ \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W(\tilde{\theta}_{t+1}, t+1)], \mathbb{E}_{\tilde{\theta}_1}[W(\tilde{\theta}_1, 1)] \} \\
\leq c^E + \theta_t - \psi(c^E) \\
+ \delta \max \{ \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W^E(\tilde{\theta}_{t+1}, t+1)], \mathbb{E}_{\tilde{\theta}_1}[W^E(\tilde{\theta}_1, 1)] \} \\
= T_E W(\theta_t, t) = W^E(\theta_t, t).
\]

Since \( \mathcal{D}(W^E) \), together with the uniform metric, is a complete metric space, and since \( T \) is a contraction, given any \( W \in \mathcal{D}(W^E) \), \( \lim_{n \to \infty} T^n W \) exists and belongs to \( \mathcal{D}(W^E) \). Since \( W^* \) is the unique fixed point to the mapping \( T : \mathcal{B} \to \mathcal{B} \), it must be that \( W^* = \lim_{n \to \infty} T^n W \).

Hence, \( W^* \in \mathcal{D}(W^E) \). That is, for any \( (\theta^t, t) \in A \), \( W^*(\theta^t, t) \leq W^E(\theta_t, t) \). The result then follows.
by noting that, for any $\theta_1 \in \Theta_1 \setminus \{\tilde{\theta}_1\}$,

$$W^*(\theta_1, 1) = TW^*(\theta_1, 1)$$

$$= \xi_1^*(\theta_1) + \theta_1 - \psi(\xi_1^*(\theta_1)) - \eta(\theta_1)\psi'(\xi_1^*(\theta_1)) + \delta \max\{\mathbb{E}_{\tilde{\theta}_1}(W^*(\tilde{\theta}_2^2, 2), \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1, 1)]\}$$

$$< \theta_1 + e^E - \psi(e^E)$$

$$+ \delta \max\{\mathbb{E}_{\tilde{\theta}_2|\theta_1}[W^*(\tilde{\theta}_2, 2), \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1, 1)]\}$$

$$= W^E(\theta_1, 1),$$

where the inequality is strict because $\eta(\theta_1) > 0$ on $\Theta_1 \setminus \{\tilde{\theta}_1\}$. ■

The next lemma combines the results in the previous two lemmas to establish Part (i) in the proposition.

**Lemma A3.** There exists $\tilde{t} \geq 1$ such that, for any $t > \tilde{t}$, any $\theta^t \in R^t$,

$$\mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W^E(\tilde{\theta}_{t+1}, t + 1)] \geq \mathbb{E}_{\tilde{\theta}_1}[W^E(\tilde{\theta}_1, 1)] \implies \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W^*(\tilde{\theta}_{t+1}, t + 1)] > \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1, 1)].$$

**Proof of Lemma A3.** Recall that $W^{E,c'}$, as defined in Lemma A1, is the value function for the stopping problem with efficient flow payoffs $\theta_t + e^E - \psi(e^E)$ and exogenous separation payoff $c'$. Now let $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1, 1)].$ Below, we will compare the function $W^{E,c'}$ with the value function $W^*$ associated with the profit-maximizing stopping problem. Recall that the latter is a stopping problem with flow payoffs, for each $t$, and each $\theta^t$, given by

$$VS_t(\theta^t) \equiv \xi_1^*(\theta^t) + \theta_t - \psi(\xi_1^*(\theta^t)) - \eta(\theta_t)J_1^t(\theta^t)\psi'(\xi_1^*(\theta^t))$$

and separation payoff $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1, 1)].$ By the property of “vanishing impulse responses”, for any $\omega > 0$, there exists $\tilde{t}$ such that, for any $t > \tilde{t}$, any $\theta^t \in R^t, \ VS_t(\theta^t) > \theta_t + e^E - \psi(e^E) - \omega.$ That is, for $t > \tilde{t}$, the flow payoff in the stopping problem that leads to the firm’s optimal contract is never less by more than $\omega$ than the corresponding flow payoff in the stopping problem with efficient flow payoffs and exogenous separation payoff $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1, 1)].$ In terms of value functions, this implies that, for all $t > \tilde{t}$, all $\theta^t \in R^t$,

$$W^*(\theta^t, t) \geq W^{E,c'}(\theta^t, t) - \frac{\omega}{1 - \delta}. \quad (14)$$

To see this, consider the set $W \subset B$ of all bounded functions $W$ from $A$ to $\mathbb{R}$ such that, for all $t > \tilde{t}$, all $\theta^t \in R^t, W(\theta^t, t) \geq W^{E,c'}(\theta^t, t) - \frac{\omega}{1 - \delta}$ and consider the operator $T^e : B \to B$ defined, for all $(\theta^t, t)$, by

$$T^eW(\theta^t, t) = VS_t(\theta^t) + \delta \max\{\mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[W(\tilde{\theta}_{t+1}, t + 1)], c'\}.\quad (14)$$
The set $W$ is closed under $T_{c'}$. Indeed, if $W \in W$, then, for any $t > \tilde{t}$, any $\theta^t \in R^t$,

$$T_{c'} W(\theta^t, t) - W^{E,c'}(\theta_t, t) = VS_t(\theta^t) + \delta \max\{E_{\tilde{\theta}^{t+1}}[W(\theta^{t+1}, t+1), c'] - \theta_t + e^E - \psi(e^E), + \delta \max\{E_{\tilde{\theta}^{t+1}}[W^{E,c'}(\tilde{\theta}_{t+1}, t+1), c']\} \geq -\omega - \delta \omega = -\frac{\omega}{1 - \delta}.$$

Since $W$, together with the uniform metric, is a complete metric space, and since $T_{c'}$ is a contraction, given any $W \in W$, $\lim_{n \to \infty} T_{c^n}^n W$ exists and belongs to $W$. Furthermore, because $c' = E_{\tilde{\theta}_1} W^*(\tilde{\theta}_1, 1)$, it must be that $W^* = \lim_{n \to \infty} T_{c^n}^n W$. Hence, $W^* \in W$, which proves (14).

Now, let $c'' = E_{\tilde{\theta}_1} W^E(\tilde{\theta}_1, 1)$. By Lemma A2, $c'' > c'$. Now observe that $W^E = W^{E,c''}$. It follows that, for all $t > \tilde{t}$ and all $\theta^t \in R^t$, if $E_{\tilde{\theta}_t+1;\theta_t} W^E(\tilde{\theta}_{t+1}, t+1) \geq E_{\tilde{\theta}_1} W^E(\tilde{\theta}_1, 1)$, then

$$E_{\tilde{\theta}^{t+1};\theta^t} W^*(\tilde{\theta}^{t+1}, t+1) \geq E_{\tilde{\theta}^{t+1};\theta^t} W^{E,c'}(\tilde{\theta}_{t+1}, t+1) - \frac{\omega}{1 - \delta} \geq E_{\tilde{\theta}_1} W^*(\tilde{\theta}_1, 1) + \omega - \frac{\omega}{1 - \delta}.$$

The first inequality follows from (14), while the second inequality follows from Lemma A1 using $c' = E_{\tilde{\theta}_1} W^*(\tilde{\theta}_1, 1)$ and choosing $\nu$ as in that lemma. The result then follows by choosing $\omega$ sufficiently small that $\nu - \frac{\omega}{1 - \delta} > 0$. ■

**Part (ii).** The proof follows from two lemmas. Lemma A4 establishes Lipschitz continuity in $\theta_t$ of the expected value of continuing the relationship in period $t + 1$, respectively under the firm’s profit-maximizing contract and the efficient contract. This result is then used in Lemma A5 to prove Part (ii) of the proposition.

**Lemma A4.** Suppose that (a) there exists $\beta \in \mathbb{R}_{++}$ such that, for each $t \geq 2$ and each $\theta_1 \in \Theta_1$, the function $\eta(\theta_1) J'_1(\theta_1, \cdot)$ is Lipschitz continuous over $\Theta'_{t+1}(\theta_1) \equiv \{\theta'_{t+1} \in \Theta'_{t+1} : (\theta_1, \theta'_t) \in R_t\}$ with Lipschitz constant $\beta$; and (b) there exists $\rho \in \mathbb{R}_{++}$ such that, for each $t \geq 2$ and each $\theta_t \in \Theta_t$, the function $f_i(\theta_t, \cdot)$ is Lipschitz continuous over $\Theta_{t-1}$ with Lipschitz constant $\rho$. Then, for each $t \geq 2$ and each $\theta^{t-1} \in R^{t-1}$, $E_{\tilde{\theta}^{t+1}(\theta^{t-1}, \cdot)} [W^*(\theta^{t-1}, \cdot, \tilde{\theta}_{t+1}, t+1)]$ is Lipschitz continuous over $\Theta_t(\theta^{t-1}) \equiv \{\theta_t \in \Theta_t : (\theta^{t-1}, \theta_t) \in R^t\}$. Moreover, for each $t \geq 2$, $E_{\tilde{\theta}_{t+1}} [W^E(\tilde{\theta}_{t+1}, t+1)]$ is Lipschitz continuous over $\Theta_t$.\footnote{Note that this result applies also to processes that are not autonomous.}

**Proof of Lemma A4.** We show that, for any $t \geq 2$ any $\theta^{t-1} \in R^{t-1}$, $E_{\tilde{\theta}^{t+1}(\theta^{t-1}, \cdot)} [W^*(\theta^{t+1}, t+1)]$ is Lipschitz continuous in $\Theta_t(\theta^{t-1})$. The proof that $E_{\tilde{\theta}_{t+1}} [W^E(\tilde{\theta}_{t+1}, t+1)]$ is Lipschitz continuous on $\Theta_t$ is similar and omitted. Let

$$M = \frac{e^E + K - \psi(e^E)}{1 - \delta} \quad \text{and} \quad m = \frac{1 + \beta L + 2\delta \rho M K}{1 - \delta}.$$
where recall that $K$ is the bound on $|\Theta_t|$ (uniform over $t$) and $L > 0$ is the bound on $\psi'$.

We will show that, for any $\theta_1 \in \Theta_1$, any $t \geq 2$, the function $W^*(\theta_1, \cdot, t)$ is Lipschitz continuous over $\Theta^t_{>1}(\theta_1)$ with constant $m$. For this purpose, let $\mathcal{L}(M, m) \subset \mathcal{B}$ denote the space of functions $W : A \rightarrow \mathbb{R}$ that satisfy the following properties: (i) for any $(\theta^t, t) \in A$, $|W(\theta^t, t)| \leq M$; (ii) for any $\theta_1 \in \Theta_1$, any $t \geq 2$, $W(\theta_1, \cdot, t)$ is Lipschitz continuous over $\Theta_1(\theta_1)$ with constant $m$; (iii) for any $\theta_1 \in \Theta_1$, any $t \geq 2$, $W(\theta_1, \cdot, t)$ is nondecreasing over $\Theta^t(\theta_1)$.

We first show that $\mathcal{L}(M, m)$ is closed under the operator $T$ defined in Proposition 3. To see this, take an arbitrary $W \in \mathcal{L}(M, m)$. First note that, for any $(\theta^t, t) \in A$,

$$TW(\theta^t, t) = VS_T(\theta^t) + \delta \max\{E_{\theta^t}^{\theta^t+1}(W(\tilde{\theta}^t, t + 1)|\tilde{\Theta}_1), E_{\tilde{\theta}_t}[W(\tilde{\Theta}_1, 1)]\} \leq e^E + K - \psi(e^E) + \delta M = M.$$

Next note that, for any $(\theta^t, t) \in A$, $TW(\theta^t, t) \geq -K - \delta M > -M$. The function $TW$ thus satisfies property (i). To see that the function $TW$ satisfies property (ii), let $t \geq 2$ and consider an arbitrary period $\tau$, $2 \leq \tau \leq t$. Then take two arbitrary sequences $(\theta^{t-1}, \theta^t_{>, \tau}), (\theta^{t-1}, \theta^t_{>, \tau}) \in R^t$. Suppose, without loss of generality, that $\theta^t_{\tau} > \theta^t_{\tau'}$. Then,

$$TW((\theta^{t-1}, \theta^t_{>, \tau}), t) - TW((\theta^{t-1}, \theta^t_{>, \tau}), t)$$

$$= \left[\xi^*_t(\theta^t) + \theta_t - \psi(\xi^*_t(\theta^t)) - \eta(\theta_1)J^t_1(\theta^t)\psi'(\xi^*_t(\theta^t))\right]_{\theta^t=(\theta^{t-1}, \theta^t_{>, \tau}, \theta^t_{>, \tau})}$$

$$- \left[\xi^*_t(\theta^t) + \theta_t - \psi(\xi^*_t(\theta^t)) - \eta(\theta_1)J^t_1(\theta^t)\psi'(\xi^*_t(\theta^t))\right]_{\theta^t=(\theta^{t-1}, \theta^t_{>, \tau}, \theta^t_{>, \tau})}$$

$$+ \delta \left(\max\{E_{\theta^t}^{\theta^t+1}(W(\tilde{\theta}^t, t + 1)|\tilde{\Theta}_1), E_{\tilde{\theta}_t}[W(\tilde{\Theta}_1, 1)]\} - \max\{E_{\theta^t}^{\theta^t+1}(W(\tilde{\theta}^t, t + 1)|\tilde{\Theta}_1), E_{\tilde{\theta}_t}[W(\tilde{\Theta}_1, 1)]\}\right)$$

(15)

The first two terms on the right-hand side of (15) are no greater than $(1 + \beta L)(\theta^t_{\tau} - \theta^t_{\tau'})$. This can be derived as follows. For any $2 \leq \tau \leq t$, any $(\theta^{t-1}, \theta^t_{>, \tau}) \in R^t$, define $\Theta_{\tau}(\theta^{t-1}, \theta^t_{>, \tau}) \equiv \{\theta_\tau \in \Theta_\tau : (\theta^{t-1}, \theta_\tau, \theta^t_{>, \tau}) \in R^t\}$. Define, for any $(\theta^{t-1}, \theta^t_{>, \tau}) \in R^t$ and any $e \in E$, the flow virtual surplus function $g_t(\theta^t, e) = e + \theta_t - \psi(e) - \eta(\theta_1)J^t_1(\theta^t)\psi'(e)$. For any $\theta^t = (\theta^{t-1}, \theta^t_{>, \tau}) \in R^t$, $g_t$ is Lipschitz continuous in $\theta_\tau$ and $\frac{\partial}{\partial \theta_\tau} g_t(\theta^{t-1}, \theta_\tau, \theta^t_{>, \tau}) \leq 1 + \beta L$ for all $e \in E$ and almost all $\theta_\tau \in \Theta_{\tau}(\theta^{t-1}, \theta^t_{>, \tau})$. The same sequence of inequalities as in Theorem 2 of Milgrom and Segal (2002) then implies the result. The final term on the right-hand side in (15) is no greater than $\delta(2\rho M K + m)(\theta^t_{\tau} - \theta^t_{\tau'})$. 

42
This follows because

\begin{align}
\mathbb{E}_{\theta_{t+1}^{\star} | (\theta_t^{\star}, \theta_t, \theta_t)} & [W(\tilde{t}+1, t+1)] - \mathbb{E}_{\theta_{t+1}^{\star} | (\theta_t^{\star}, \theta_t, \theta_t)} [W(\tilde{t}+1, t+1)] \\
= & \mathbb{E}_{\theta_{t+1}^{\star} | (\theta_t^{\star}, \theta_t, \theta_t)} [W((\theta_t^{\star}, \theta_t, \theta_t), t+1)] \\
& - \mathbb{E}_{\theta_{t+1}^{\star} | (\theta_t^{\star}, \theta_t, \theta_t)} [W((\theta_t^{\star}, \theta_t, \theta_t), t+1)] \\
& + \mathbb{E}_{\theta_{t+1}^{\star} | (\theta_t^{\star}, \theta_t, \theta_t)} [W((\theta_t^{\star}, \theta_t, \theta_t), t+1)] - W((\theta_t^{\star}, \theta_t, \theta_t), t+1)] \\
= & \int_{\mathcal{B}} W((\theta_t^{\star}, \theta_t, \theta_t), t+1) \left[ f_{t+1} (\theta_{t+1} | \theta_t^{\star}, \theta_t, \theta_t) - f_{t+1} (\theta_{t+1} | \theta_t^{\star}, \theta_t, \theta_t) \right] d\theta_{t+1} \\
& + \mathbb{E}_{\theta_{t+1}^{\star} | (\theta_t^{\star}, \theta_t, \theta_t)} [W((\theta_t^{\star}, \theta_t, \theta_t), t+1)] - W((\theta_t^{\star}, \theta_t, \theta_t), t+1)] \\
\leq & (2\rho MK + m) (\theta_t^{\star} - \theta_t^{\star}) ,
\end{align}

where the inequality follows from the fact that, for any \( \theta_{t+1} \in \Theta_{t+1} \), any \((\theta_t^{\star}, \theta_t, \theta_t)\), the function \( f_{t+1}(\theta_{t+1} | \theta_t^{\star}, \theta_t, \theta_t) \) is Lipschitz continuous with constant \( \rho \) together with the fact that \( |\theta_t| \leq K \) all \( t \). We conclude that

\[ TW((\theta_t^{\star}, \theta_t, \theta_t), t) - TW((\theta_t^{\star}, \theta_t, \theta_t), t) \leq (1 + \beta L + 2\delta \rho MK + \delta m) (\theta_t^{\star} - \theta_t^{\star}) = m (\theta_t^{\star} - \theta_t^{\star}) . \]

Since \((\theta_t^{\star}, \theta_t, \theta_t)\) and \((\theta_t^{\star}, \theta_t, \theta_t)\) were arbitrary, it follows that for any \( \theta_1 \in \Theta_1 \), and any \( t \), the function \( TW(\theta_1, \cdot, t) \) is Lipschitz continuous over \( \Theta(t) \) with constant \( m \), i.e. \( TW \) indeed satisfies property (ii) above. Lastly that \( TW \) satisfies property (iii) follows from the fact that the transformation \( T \) preserves the monotonicity of \( W \), as shown in the proof of Proposition 3. We thus conclude that \( TW \in \mathcal{L}(M, m) \) which verifies that \( \mathcal{L}(M, m) \) is closed under the \( T \) operator. The fact that \( \mathcal{L}(M, m) \subset \mathcal{B} \), endowed with the uniform metric, is a complete metric space, together with the fact that \( T \) is a contraction, then implies that \( W^* \in \mathcal{L}(M, m) \). Using the same argument as in (16), we then have that \( \mathbb{E}_{\theta_{t+1}^{\star} | (\theta_t^{\star}, \theta_t)} [W^*(\tilde{t}+1, \tilde{t}+1)] \) is Lipschitz continuous over \( \Theta_t(\theta_t) \) with constant \( (2\rho MK + m) \). □

The next lemma uses the result in the previous lemma to establish Part (ii) in the proposition.

**Lemma A5.** Suppose that the conditions in Lemma A4 hold. Then the result in Part (ii) in the proposition holds.

**Proof of Lemma A5.** Let \( \tilde{t} \) be as defined in Lemma A3. Take an arbitrary \( t > \tilde{t} \) and \( \theta_t^{\star} \in \mathcal{R}^{\star} \) such that \( \theta_t^{\star} \in \{ Supp[F_1(\cdot, [\theta_t^{\star}])] \} \). The continuity of \( \mathbb{E}_{\theta_{t+1}}[W^E(\tilde{t}+1, t+1)] \) in \( \theta_t^{\star} \) established in the previous lemma, implies \( \mathbb{E}_{\theta_t^{\star}}[W^E(\tilde{t}+1, t+1)] = \mathbb{E}_{\theta_t^{\star}}[W^E(\tilde{t}+1, t+1)] \). Since \( t > \tilde{t} \), by Lemma A3, it follows that \( \mathbb{E}_{\theta_t^{\star}}[W^E(\tilde{t}+1, t+1)] > \mathbb{E}_{\theta_t^{\star}}[W^E(\tilde{t}+1, t+1)] \). By Lemma A4, \( \mathbb{E}_{\theta_t^{\star}}[W^E(\tilde{t}+1, t+1)] \) is continuous in \( \theta_t^{\star} \). Since \( \theta_t^{\star} \in \{ Supp[F_1(\cdot, [\theta_t^{\star}])] \} \), there exists \( \epsilon > 0 \) such that, for all \( (\theta_t^{\star} - \epsilon, \theta_t^{\star}) \), \( \mathbb{E}_{\theta_t^{\star}}[W^E(\tilde{t}+1, t+1)] > \mathbb{E}_{\theta_t^{\star}}[W^E(\tilde{t}+1, t+1)] \). It follows that \( \theta_t^{\star} (\theta_t^{\star} - \epsilon) < \theta_t^{\star} \). □
**Proof of Proposition 6.** Because the process is autonomous, $\theta^E_t = \phi^E_t$ for all $t$. Firstly, suppose that $\theta^E < \bar{\theta}_1$. Consider the case that, for all $\theta_1 > \theta^E$, $\E_{\theta_2|\theta_1}[W^*(\theta^2, 2)] > \E[W^*(\bar{\theta}_1, 1)]$. Proposition 4, together with the monotonicity property of $W^*(\cdot, t)$ established in Proposition 3, then implies that, for any $t \geq 1$, any $\theta^t \in R^t$ such that $\theta_1, \theta_t > \theta^E$, $\E_{\theta^t+1|\theta^t}[W^*(\theta^t+1, t + 1)] > \E[W^*(\bar{\theta}_1, 1)]$. This means that, for any $t$ any $\theta^t \in R^t$ such that $\kappa^*_t(\theta^{t-1}) = 1$ and $\kappa^*_t(\theta^t) = 0$, necessarily $\kappa_t^E(\theta^t) = 0$ (except for the possibility that $\theta^t$ is such that $\theta_s = \theta^E$ for some $s \leq t$, which, however, has zero measure). That is, any manager who is fired in period $t$ under the firm’s profit-maximizing contract, is either fired in the same period or earlier under the efficient contract, which establishes Case (i) in the proposition.

Next, assume that there exists a $\theta_1 > \theta^E$ such that $\E_{\theta_2|\theta_1}[W^*(\theta^2, 2)] < \E[W^*(\bar{\theta}_1, 1)]$, which implies that $\theta^*_1 > \theta^E$. By assumption, the manager is retained with positive probability after the first period, i.e. $\theta^*_1 \in (\theta^E, \bar{\theta})$. The result in Part (a) then holds by letting $t = 2$, whereas the result in Part (b) follows from Part (i) of Proposition 5. Indeed, from that proposition, there exists a $\bar{t} > t$ such that, for all $t > \bar{t}$, if $\theta^t \in R^t$ is such that $\kappa^*_{t-1}(\theta^{t-1}) = 1$ and $\kappa^*_t(\theta^t) = 0$, then $\theta_s \leq \theta^E$ for some $s \leq t$. Hence $\kappa_t^E(\theta^t) = 0$ (once again, except for the possibility that $\theta_s = \theta^E$, which however has zero measure).

Lastly suppose that $\theta^E = \bar{\theta}_1$. Case (i) then always trivially applies. ■